

# La préservation des données de recherche au Canada : Livre blanc

Préparé par le Groupe d'experts sur la préservation de Portage, au nom de l'Association des bibliothèques de recherche du Canada (ABRC)

Umar Qasim, président du GEP (University of Alberta)  
Corey Davis (University of Victoria and Council of Prairie and Pacific University Libraries)  
Alex Garnett (Simon Fraser University)  
Steve Marks (University of Toronto)  
Michael Moosberger (Dalhousie University)

AVRIL 2018

Réseau Portage  
Association des bibliothèques de recherche du Canada  
[portage@carl-abrc.ca](mailto:portage@carl-abrc.ca)

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

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

**CARL ABRC**  
CANADIAN ASSOCIATION OF  
RESEARCH LIBRARIES ASSOCIATION DES BIBLIOTHÈQUES  
DE RECHERCHE DU CANADA

# Table des matières

Résumé exécutif.....	2
Introduction.....	4
Définition de la préservation numérique .....	4
Principes directeurs s’appliquant aux efforts de préservation numérique de Portage	5
Défis associés à la préservation des données de recherche .....	7
Le contexte canadien de la préservation des données de recherche.....	8
<b>Modèle pour un service national de préservation numérique distribué.....</b>	<b>10</b>
Le modèle de référence pour un système ouvert d’archivage d’information (OAIS) ....	10
Décentralisation du modèle OAIS.....	11
Rôle de Portage dans la coordination d’une approche décentralisée de la préservation des données de recherche au Canada .....	11
<i>Services de préservation.....</i>	<i>12</i>
<i>Services de dépôt.....</i>	<i>13</i>
<i>La planification et l’administration de la préservation.....</i>	<i>13</i>
<b>Prochaines étapes.....</b>	<b>15</b>
<b>Conclusion .....</b>	<b>16</b>
Annexe A. Termes traduits de la Figure 1. ....	17

## Résumé exécutif

Le présent document fournit un cadre dans lequel la préservation numérique est définie, de même qu'un ensemble de principes directeurs qui reflètent les valeurs et les engagements des organisations et des communautés qui évoluent actuellement dans ce secteur. En s'appuyant sur ces fondements, une approche fédérée à la préservation des données de recherche au Canada est proposée, qui s'appuierait sur les efforts continus de Portage pour établir des réseaux d'expertise et des communautés de pratique partout au pays.

En utilisant comme guide le modèle de référence pour un système ouvert d'archivage d'information (OAIS), il est recommandé que Portage appuie l'établissement d'un réseau d'archives coordonné et distribué, qui désigne collectivement les « fournisseurs de services de préservation » (FSP), afin de répondre aux besoins des chercheurs canadiens en matière de gestion des données. Le modèle proposé répartirait six fonctions OAIS selon trois secteurs :

Les **services de dépôt**, qui s'occuperaient des fonctions *Intégration*, *Accès* et *Gestion des données* du modèle OAIS. Cela pourrait comprendre les deux plateformes de dépôt actuellement prises en charge par Portage, le Dépôt fédéré de données de recherche (DFDR) et Dataverse, ainsi que n'importe quel nombre de dépôts spécialisés et institutionnels basés au Canada.

Les **services de préservation**, qui s'occuperaient des fonctions d'*Archivage* dans le modèle OAIS et qui pourraient être entrepris par l'entremise d'un réseau coordonné de FSP numériques.

La **planification et la surveillance**, où Portage serait responsable des fonctions d'*Administration* et où les responsabilités de *Planification de la préservation* seraient partagées entre Portage et les FSP.

Voici les prochaines étapes proposées afin de permettre au réseau Portage de réaliser cette vision d'un réseau fédéré de préservation numérique :

- Établir une compréhension commune des exigences de base en matière de préservation numérique, afin de déterminer les attributs fondamentaux d'une infrastructure durable et distribuée de préservation numérique des données de recherche au Canada.
- Cultiver et alimenter des partenariats avec les intervenants nationaux et régionaux pour harmoniser les services de préservation numérique existants et nouveaux, coordonner les activités de communication, promotion et sensibilisation, et explorer les possibilités de financement collaboratif.

- Continuer de véhiculer un message clair et unifié sur toutes les questions liées à l'établissement et au déploiement d'une infrastructure nationale de préservation des données de recherche durable.
- Définir les compétences essentielles à l'appui de la formation des personnes responsables des données de recherche et des activités de préservation numérique dans leurs institutions et organismes respectifs.

## Introduction

Le Groupe d'experts sur la préservation (GEP) a été créé afin de conseiller Portage au sujet de l'élaboration de l'infrastructure de gestion des données de recherche (GDR) et des pratiques exemplaires en matière de préservation des données et métadonnées de recherche.<sup>1</sup> Les membres du GEP ont rédigé le présent livre blanc comme document de référence pour décrire le contexte actuel de préservation numérique, souligner une partie du travail de préservation numérique déjà amorcé au Canada, et cerner les défis que Portage et d'autres intervenants doivent relever pour développer et améliorer la capacité et l'infrastructure de GDR partout au pays. Le présent livre blanc vise à stimuler la discussion sur les orientations et les initiatives de préservation numérique au Canada et sur la façon dont Portage, en collaboration avec d'autres intervenants, peut participer à définir et à faire avancer ces initiatives.

## Définition de la préservation numérique

« La préservation numérique est la gestion active du contenu numérique au fil du temps pour assurer l'accès continu. »<sup>2</sup> [TRADUCTION] La préservation numérique comprend une série d'activités, comme la sélection du contenu à préserver, la préparation et le maintien du contenu dans un environnement qui permet l'accès, et la mise en place de stratégies pour assurer la disponibilité de ce contenu au fil du temps.

Les pratiques exemplaires de préservation numérique sont constamment élaborées et améliorées au fur et à mesure que les contextes technologiques, économiques, politiques et sociaux changent. L'un des cadres les plus achevés ayant guidé efficacement de nombreuses organisations et initiatives dans leurs efforts de préservation numérique est le modèle de référence pour un système ouvert d'archivage d'information (OAIS), initialement élaboré par la communauté des données spatiales et désormais adopté à grande échelle en tant que cadre de pratiques exemplaires. Le modèle OAIS décrit « une archive, constituée de personnes et de systèmes, qui a accepté la responsabilité de préserver l'information et de la rendre accessible à une communauté désignée »<sup>3</sup> [TRADUCTION]. Ce modèle fournit non seulement un cadre pour les pratiques de préservation numérique, mais il fournit aussi à la communauté de la préservation numérique un vocabulaire commun pour parler de ses activités et initiatives.

---

<sup>1</sup> <https://portagenetwork.ca/fr/collaborer-avec-portage/reseau-dexperts/groupe-dexperts-sur-la-preservation/groupe-dexperts-sur-la-preservation-du-reseau-portage-propos-de-nous/>

<sup>2</sup> <http://www.digitalpreservation.gov/about/>

<sup>3</sup> <https://www.iso.org/fr/standard/57284.html>

# Principes directeurs s'appliquant aux efforts de préservation numérique de Portage

Les principes directeurs reflètent les valeurs d'une organisation ou d'une communauté et son niveau d'engagement pour préserver le contenu numérique dont elle est responsable. Ces principes fondamentaux, qui font partie intégrante de tout effort de préservation numérique, peuvent varier d'une organisation à l'autre ou d'une communauté à l'autre. Ils servent de fondement à l'élaboration de politiques qui orientent les mesures prises pour relever les défis actuels et futurs posés par la préservation numérique.

Les principes directeurs suivants sous-tendent l'analyse et les recommandations contenues dans le présent livre blanc. La liste qui suit n'est pas destinée à être normative ni exhaustive. Elle reflète plutôt l'expérience des membres du GEP dans l'élaboration et la mise en œuvre de systèmes à l'appui de la préservation numérique, l'état actuel du domaine et les objectifs de Portage de façon plus générale.

## **Principe 1 : Les processus et les outils doivent être communautaires, transparents et ouverts**

La préservation numérique pose un ensemble de défis qu'il est impossible de relever dans un délai court ou à l'aide de solutions simples. La préservation à long terme englobe un éventail de dimensions techniques, juridiques et éthiques complexes. Par conséquent, il devra toujours y avoir une grande variété de voix qui éclairent le processus, et cela ne peut être fait que dans un environnement où la prise de décisions est ouverte et soumise à l'examen et à l'orientation des communautés d'intervenants. Une volonté de mobiliser un large éventail d'intervenants et de collaborer avec eux, de même qu'un engagement à l'égard de la transparence et de l'ouverture, sont essentiels au développement continu et à la pérennité des efforts de préservation numérique au Canada.

## **Principe 2 : L'accès est l'un des principaux objectifs de la préservation numérique**

La préservation numérique n'est pas une fin en soi. L'objectif de préserver tout contenu est de permettre un accès approprié à ce contenu à l'avenir. Le contenu numérique doit être fourni dans des formats qui répondent aux besoins des groupes d'utilisateurs désignés et qui sont utilisables et compréhensibles à un niveau approprié.

## **Principe 3 : Les données ne peuvent ou ne doivent pas toutes être préservées**

Tout préserver coûterait extrêmement cher et prendrait beaucoup de temps. La mise en place de critères raisonnables et justifiables pour la sélection des données qui seront

préservées à long terme constitue un aspect important de la planification de la préservation et est essentielle à la réussite de tout effort dans ce secteur.

#### **Principe 4 : La préservation numérique est un exercice continu de gestion des risques**

La préservation numérique n'est pas une série prescrite de mesures statiques à réaliser, après quoi l'on peut affirmer que quelque chose est « préservé ». Il s'agit plutôt d'un processus continu qui comprend l'évaluation minutieuse des risques posés à un ensemble précis de contenu numérique. Les risques peuvent provenir de la nature des données elles-mêmes, de leurs formats ou des supports sur lesquels elles sont stockées, de la ou des plateformes dans lesquelles elles se trouvent et de tout autre facteur économique et politique d'importance. Par conséquent, la surveillance constante des données préservées, des changements technologiques, des besoins des groupes d'utilisateurs et du contexte sociétal élargi est essentielle.

#### **Principe 5 : Les métadonnées sont cruciales**

Les métadonnées sont une composante essentielle des stratégies de préservation numérique.<sup>4</sup> Les organisations doivent faciliter la création et la mise-à-jour de métadonnées solides afin d'assurer la fiabilité, l'authenticité et la convivialité des objets numériques qui leur sont confiés, conformément aux normes et aux pratiques exemplaires.<sup>5</sup> Les métadonnées qui démontrent les chaînes de possession et d'authenticité revêtent une importance particulière pour les données de recherche requises pour valider les résultats et autres conclusions publiées.

---

<sup>4</sup> [https://en.wikipedia.org/wiki/Preservation\\_metadata](https://en.wikipedia.org/wiki/Preservation_metadata)

<sup>5</sup> P. ex., voir : <http://www.dpconline.org/docs/technology-watch-reports/894-dpctw13-03/>

# Défis associés à la préservation des données de recherche

## Risques économiques

La plus grande menace à l'accessibilité à long terme de l'information numérique est possiblement d'ordre économique. La préservation numérique coûte cher, et la taille et la complexité croissantes des données de recherche posent d'énormes défis en matière de coûts. Selon une étude menée en 2014, les résultats de recherche à l'échelle mondiale doublent tous les neuf ans.<sup>6</sup> Qui assume le coût de l'infrastructure d'intégration, de stockage et d'accès aux données ainsi que de l'expertise nécessaire pour prendre en charge cette masse de données de recherche?

## Hétérogénéité des données de recherche

La structure et le contenu hautement variables des données de recherche posent également des défis uniques. Les métadonnées contextuelles détaillées doivent être préservées avec les données de recherche de façon à ce qu'elles soient compréhensibles et réutilisables indépendamment par les futurs chercheurs. Cela pose une série de défis pour les intendants des données et signifie que l'expertise en métadonnées est mieux déployée vers le début du cycle de vie de la recherche. Cela est différent pour d'autres types de résultats de recherche, comme les articles de revues électroniques et les livres électroniques, qui ont tendance à faire partie d'une stratégie de préservation vers la fin du cycle de vie de la recherche.

## Changements technologiques

Les changements technologiques présentent également toute une gamme de défis, comme le manque de compatibilité rétrospective et la dépendance à des environnements matériels et logiciels spécifiques, souvent exclusifs. Ces problèmes sont aggravés par les nombreux environnements matériels et logiciels hautement spécialisés associés à de multiples domaines d'études. Alors qu'un grand corpus d'articles de revues électroniques ou d'autres résultats publiés représentant des dizaines de disciplines ne peut être stocké que dans une poignée de formats de fichiers, ce n'est pas le cas des données de recherche, pour lesquelles un grand nombre de types de données et de formats de fichiers doivent être gérés au fil du temps. Trouver des façons de produire et de préserver des copies archivistiques d'ensembles de données

---

<sup>6</sup> <http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html>



complexes sans égard au matériel et aux logiciels sera essentiel pour permettre l'accès à long terme.

## Le contexte canadien de la préservation des données de recherche

*« Pour accroître notre capacité de préserver l'information numérique, il nous faut un cadre pour renforcer, coordonner et mieux communiquer nos efforts collectifs. »<sup>7</sup> [TRADUCTION]*

Au Canada, on reconnaît de plus en plus la nécessité d'adopter une stratégie nationale coordonnée pour la préservation numérique.<sup>8</sup> Cela est reconnu dans un large éventail de disciplines, des sciences humaines numériques<sup>9</sup> jusqu'aux sciences, surtout lorsque les projets de recherche sur les « mégadonnées » nécessitent une approche coordonnée du stockage, de la gestion et de la préservation à long terme de grands ensembles de données<sup>10</sup>.

Et bien qu'on soit de plus en plus conscient de l'importance d'une approche coordonnée de la préservation numérique au Canada :

*« [...] le Canada ne dispose pas d'un « plan directeur » visant à guider les milieux scientifique, culturel et éducatif, les entreprises et la société civile, dans le choix de méthodes de production, d'utilisation, de partage et de préservation de ce patrimoine numérique en pleine expansion. »<sup>11</sup> [TRADUCTION]*

De nombreuses organisations canadiennes travaillent à la gestion des données de recherche et à la préservation numérique. Dans l'ensemble, mais souvent dans un isolement relatif, elles ont investi de façon importante dans l'acquisition de l'expertise, le renforcement de la capacité organisationnelle et le déploiement de l'infrastructure de recherche numérique pour assurer la viabilité à long terme d'importantes données de

---

<sup>7</sup> <http://www.lac-bac.gc.ca/obj/012033/f2/012033-1000-f.pdf> (p. 7)

<sup>8</sup> <http://www.lac-bac.gc.ca/obj/012033/f2/012033-1000-f.pdf> (p. 4)

<sup>9</sup> [http://www.cwrc.ca/cwrc\\_news/lasting-change-sustaining-digital-scholarship-and-culture-in-canada/](http://www.cwrc.ca/cwrc_news/lasting-change-sustaining-digital-scholarship-and-culture-in-canada/)

<sup>10</sup> [http://sciencepourlepublic.ca/uploads/fr/assessments%20and%20publications%20and%20On%20releases/memory/CofCA\\_14-377\\_MemoryInstitutions\\_WEB\\_F.PDF](http://sciencepourlepublic.ca/uploads/fr/assessments%20and%20publications%20and%20On%20releases/memory/CofCA_14-377_MemoryInstitutions_WEB_F.PDF) (p. 34)

<sup>11</sup> <http://www.lac-bac.gc.ca/obj/012033/f2/012033-1000-f.pdf> (p. 7)

recherche et d'autres documents numériques. À ce jour, cependant, il n'existe pas de mécanisme national qui permette de mettre en œuvre l'infrastructure et l'expertise de préservation de manière à répondre aux besoins du milieu de la recherche de façon coordonnée et durable.

Un certain nombre d'organisations et d'intervenants clés au Canada se concentrent sur la prestation d'une infrastructure de recherche numérique pour la préservation à long terme. Ces organisations possèdent une expertise et une capacité de préservation numérique considérables, et leur engagement sera essentiel au succès des efforts de Portage.

<p><b>Organisations nationales</b></p> <ul style="list-style-type: none"> <li>• Association des bibliothèques de recherche du Canada</li> <li>• Réseau canadien de documentation pour la recherche</li> <li>• Calcul Canada</li> <li>• CANARIE et Données de recherche Canada</li> <li>• Bibliothèque et Archives Canada</li> <li>• Conseil du leadership sur l'infrastructure de recherche numérique</li> </ul>	<p><b>Organisations régionales</b></p> <ul style="list-style-type: none"> <li>• Consortiums régionaux de bibliothèques universitaires (COPPUL, CBUO, BCI, CAUL-CBUA)</li> <li>• Scholars Portal du CBUO</li> </ul> <p><b>Établissements</b></p> <ul style="list-style-type: none"> <li>• Un certain nombre de bibliothèques de recherche au Canada offrent des services de préservation numérique de grande capacité et évolutifs à leurs campus.</li> </ul>
--	--

# Modèle pour un service national de préservation numérique distribué

Sachant que de nombreuses organisations au Canada participent actuellement à des activités de préservation numérique, le GEP recommande que Portage pilote les efforts visant à établir un réseau de préservation de type OAIS distribué afin de fournir aux chercheurs canadiens un moyen facile et fiable de déposer, de trouver, de partager et de préserver les données de recherche. Des renseignements généraux sur le modèle de référence OAIS sont fournis afin de bien encadrer cette recommandation.

## Le modèle de référence pour un système ouvert d'archivage d'information (OAIS)

Le modèle de référence OAIS définit et décrit un ensemble fondamental de fonctions qu'une organisation utilise pour remplir sa mission principale, qui consiste à préserver l'information à long terme pour sa communauté.<sup>12</sup>

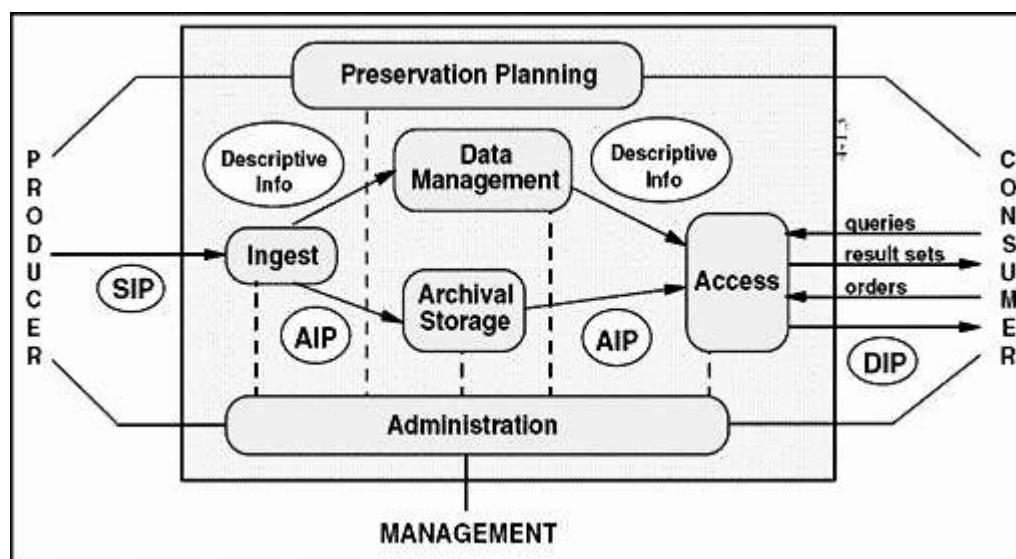


Figure 1 : Le modèle de référence OAIS.<sup>13</sup> Voir l'annexe A pour tous les termes traduits.

1. La fonction **d'intégration** (Ingest) accepte l'information des producteurs de contenu, la valide et prépare son inclusion dans les archives. La création de

<sup>12</sup>« Communauté désignée : Un groupe déterminé de consommateurs potentiels qui doivent être en mesure de comprendre un ensemble de renseignements donné. La communauté désignée peut être composée de multiples groupes d'utilisateurs. Une communauté désignée est définie par les Archives et cette définition peut changer au fil du temps. »

<https://public.ccsds.org/pubs/650x0m2.pdf> (p. 21)

<sup>13</sup> <https://commons.wikimedia.org/wiki/File:OAIS-.gif>

métadonnées descriptives fait également partie de cette fonction. La fonction d'intégration donne lieu à la création de paquets d'information archivé (AIP)<sup>14</sup>, qui sont transférées dans un entrepôt d'archives pour être conservées à long terme.

2. La fonction de **gestion des données** (Data Management) maintient les métadonnées descriptives, gère les données administratives et appuie la recherche et la découverte du contenu archivé.
3. La fonction d'**archivage** (Archival Storage) gère le stockage et le maintien à long terme du contenu numérique dans les archives.
4. La fonction d'**accès** (Access) diffuse le contenu des archives.
5. La fonction de **planification de la préservation** (Preservation Planning) surveille continuellement l'état des archives et utilise des stratégies de préservation appropriées pour maintenir l'accessibilité du contenu.
6. La fonction d'**administration** (Administration) gère les opérations des archives et coordonne les activités des cinq autres fonctions.<sup>15</sup>

## Décentralisation du modèle OAIS

En règle générale, une seule organisation est responsable des personnes et de l'infrastructure nécessaires pour soutenir un système compatible avec OAIS. Cependant, il est possible d'adopter une approche décentralisée pour la mise en œuvre de diverses fonctions OAIS. Grâce à cette approche, même une seule fonction OAIS peut être mise en œuvre dans de multiples organisations, chaque organisation fournissant un service unique portant sur différents segments du modèle fonctionnel. À l'aide d'un modèle de service décentralisé, les fonctions essentielles d'un système d'archivage basé sur OAIS peuvent être attribuées à différentes organisations au sein d'un réseau, surtout dans les situations où une seule organisation est incapable d'accomplir toutes les fonctions essentielles de façon isolée en raison de contraintes de ressources ou d'autres limitations.

## Rôle de Portage dans la coordination d'une approche décentralisée de la préservation des données de recherche au Canada

Le GEP recommande que Portage dirige les efforts visant à établir un réseau de préservation de type OAIS distribué pour les données de recherche au Canada. Certaines fonctions du réseau sont déjà en place et d'autres sont en cours d'élaboration

---

<sup>14</sup> Une « AIP » telle que définie par le modèle de référence OAIS est une trousse d'information qui sert à transmettre des objets d'archives dans un système d'archivage numérique, à stocker les objets dans le système et à transmettre les objets à partir du système. Une « AIP » contient à la fois des métadonnées qui décrivent la structure et le contenu, et le contenu lui-même.

<sup>15</sup> <http://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file>

en partenariat avec plusieurs organisations. Dans le cadre de ce modèle, Portage jouerait principalement un rôle de coordination, en aidant à harmoniser les activités de diverses organisations pour veiller à ce que toutes les fonctions nécessaires définies dans le modèle OAIS soient en place pour appuyer la préservation à long terme des données de recherche au Canada.

Les fonctions basées sur OAIS d'un réseau distribué seraient réparties selon trois secteurs :

1. Les **services de préservation**, surtout l'archivage
2. Les **services de dépôt** pour l'intégration, l'accès et la gestion des données
3. La **planification et l'administration** de la préservation

## Services de préservation

Des services de préservation spécialisés dans le modèle de service distribué appuieraient la fonction d'archivage du système. Les fournisseurs de services d'archivage et de services de préservation connexes (fournisseurs de services de préservation numérique, ou FSP) travailleraient en collaboration avec Portage pour déterminer la portée de la fonction de planification de la préservation dans son ensemble, tandis que Portage jouerait un rôle de coordination en aidant à intégrer plusieurs FSP au processus de planification de la préservation dans son ensemble.

Les FSP seraient des organisations établies au Canada qui ont déjà une capacité de préservation numérique. Il pourrait s'agir, par exemple, de bibliothèques de recherche qui offrent déjà des services de préservation numérique aux communautés de leurs campus, surtout en ce qui a trait à l'intégration des données et à l'archivage permanent, ou de dépôts spécialisés qui répondent aux besoins d'un milieu universitaire ou de recherche en particulier. Il pourrait s'agir de consortiums régionaux de bibliothèques universitaires, comme le Scholars Portal du CBUO et COPPUL, qui offrent tous deux à leurs membres une vaste gamme de services liés à la préservation. Il pourrait également s'agir d'organisations centrées sur des réseaux de recherche à très haute vitesse et des installations informatiques de haute performance, comme Calcul Canada, qui, en partenariat avec Portage, offre le traitement des données à des fins de préservation dans le cadre de la plateforme du DFDR.

## Archivage

L'archivage est la fonction du réseau qui gère le stockage et le maintien à long terme du contenu. Les activités régulières de maintenance, comme la migration des formats, l'actualisation des supports, la vérification des erreurs et la planification de la reprise après sinistre constituent une partie importante de ce service. L'archivage n'offre aucun accès direct aux utilisateurs finaux; il interagit plutôt avec d'autres fonctions du réseau de préservation.

Dans le modèle de service de réseau de préservation proposé, la fonction d'archivage serait assumée par des organisations possédant les capacités et l'expertise pour remplir ce rôle efficacement. En s'associant avec un ensemble diversifié de FSP, le réseau pourrait atténuer les risques liés au fait de travailler avec une plateforme technique unique. Parallèlement, Portage travaillerait avec les FSP afin d'élaborer un ensemble d'exigences communes pour l'archivage et les services connexes, tandis que l'exploitation et le maintien de l'infrastructure associée à chaque FSP seraient laissés à leur organisation hôte.

## Services de dépôt

En général, les dépôts de données fournissent des fonctions d'intégration, d'accès et de gestion des données pour le contenu déposé. Dans un modèle distribué, tout fournisseur canadien de services d'archivage de données capable de répondre à certaines modalités et exigences techniques de base, établies par la communauté, pourrait être inclus dans le réseau de préservation.

Dans sa vision d'un cadre national commun de services de données au Canada, Portage prend en charge actuellement deux technologies de dépôt, soit le Dépôt fédéré de données de recherche (DFDR) et Dataverse. Le DFDR, créé en partenariat avec Calcul Canada, vise à combler les lacunes connues dans le contexte des dépôts; notamment la capacité de manipuler de grands ensembles de données (« mégadonnées »), la prestation d'une interface d'exploration nationale et l'élaboration d'une filière de préservation grâce à l'intégration d'Archivematica.<sup>16</sup> Parallèlement, le groupe de travail Dataverse North de Portage est une communauté de pratique entre les bibliothèques universitaires du Canada qui a pour but de renforcer la capacité et le soutien de Dataverse.<sup>17</sup> Dataverse North examine les options envisageables pour la création d'une instance nationale de Dataverse qui serait mise à la disposition de tous les chercheurs canadiens. Nous croyons que ces deux technologies de dépôt conviennent pour remplir les fonctions d'intégration, d'accès et de gestion des données du modèle OAIS.

## La planification et l'administration de la préservation

Portage serait responsable de l'ensemble de la fonction d'*administration* du réseau de préservation, dont la portée serait déterminée en consultation avec les FSP, et partagerait la responsabilité de la *planification de la préservation* avec ces mêmes organisations et initiatives.

---

<sup>16</sup> <http://frdr.thedev.ca/francais/>

<sup>17</sup> <https://dataverse.org/>

## **Planification de la préservation**

La planification de la préservation consiste à documenter les stratégies et les pratiques appropriées de préservation pour maintenir l'accessibilité du contenu numérique au fil du temps. Dans ce contexte, une organisation doit se tenir au courant des changements apportés à l'environnement externe, afin de mettre en œuvre de nouvelles stratégies d'atténuation des risques au besoin.

Dans le modèle de service de réseau de préservation proposé, Portage travaillerait avec les FSP pour exécuter cette fonction. Étant donné que les FSP conservent d'importantes quantités de données à leur propre compte, ils sont des partenaires naturels pour s'engager dans la fonction de planification de la préservation du réseau. Les FSP collaboreraient avec Portage pour veiller à ce que les exigences en matière de préservation et les stratégies d'atténuation des risques utilisées localement soient compatibles avec celles du réseau de préservation dans son ensemble.

En fonction de l'évolution des besoins collectifs et des environnements technologiques, Portage recommanderait aux FSP des stratégies et des pratiques afin d'empêcher que le contenu préservé ne devienne inaccessible. De même, les FSP pourraient cerner de nouvelles pratiques ou de nouveaux risques nécessitant une intervention de la part de Portage ou d'autres FSP. Portage collaborerait également avec les FSP pour déterminer les spécifications pour la création de paquets d'information archivé (AIP). De plus, Portage tiendrait un registre du contenu préservé lié à Portage chez chaque FSP. Un registre central des formats de fichiers recommandés pour la préservation à long terme devrait également être élaboré par Portage, en collaboration avec d'autres initiatives de préservation nationales et internationales.

## **Administration**

La fonction d'administration est responsable des activités quotidiennes du réseau de préservation, y compris la coordination des activités entre les diverses fonctions du réseau et entre les organisations participantes. Dans le modèle proposé, Portage serait responsable de la maintenance du réseau d'un point de vue logistique et stratégique. Le travail stratégique entraînerait vraisemblablement la négociation d'ententes de service réciproques avec les FSP, et la supervision de la transition des documents d'un FSP à un autre, au besoin. Le travail stratégique pourrait également comprendre l'élaboration de lignes directrices et de procédures pour la certification des FSP en tant que dépôts dignes de confiance, de critères d'évaluation du contenu intégré et de modalités pour les licences de dépôt. De façon plus générale, Portage s'assurerait également à ce que tous les participants au réseau demeurent au fait des pratiques courantes en matière de préservation numérique, et que la planification de la préservation dans son ensemble soit cohérente et compatible entre les FSP.

## Prochaines étapes

Le Canada a besoin d'un programme de préservation coordonné comprenant les principales activités qui appuient la préservation des données de recherche à l'échelle nationale. La création d'un réseau de préservation distribué compatible avec OAIS en partenariat avec les FSP est essentielle à la préservation à long terme des données de recherche au Canada. Comme l'indique Richard Whitt, de Google : « Le principal... défi n'est pas de supplanter l'activité qui se déroule aujourd'hui, mais de contribuer à coordonner, à élargir et à approfondir ces efforts »<sup>18</sup>.

En tenant compte du contexte actuel lié à la GDR, à l'infrastructure numérique et à la préservation numérique au Canada, et en tenant particulièrement compte des efforts déjà déployés, le GEP recommande qu'au cours des deux prochaines années, Portage se concentre sur les secteurs suivants, de façon à promouvoir un réseau de préservation distribué pour les données de recherche au Canada.

- **Favoriser une compréhension commune des exigences de base en matière de préservation numérique :** Former un groupe de travail sous la gouverne du réseau Portage, dont font partie un large éventail d'intervenants, afin de déterminer les attributs fondamentaux d'un réseau de préservation numérique durable et distribué pour les données de recherche au Canada, y compris les options de financement de l'archivage à long terme.
- **Mettre l'accent sur les partenariats :** Conclure des ententes avec les intervenants nationaux et régionaux afin d'harmoniser l'infrastructure de préservation numérique existante et nouvelle, de coordonner les activités de communication et de sensibilisation auprès des intervenants, et d'explorer les possibilités de financement collaboratif.
- **Unifier les messages :** Aligner les efforts de plaidoyer et de sensibilisation parmi les groupes d'intervenants, avec la nécessité d'adopter une approche coordonnée pour le déploiement et la gestion de l'infrastructure de préservation des données de recherche à l'échelle nationale. Préconiser une rationalisation des mandats parmi plusieurs organisations afin de permettre l'établissement d'importantes ressources de stockage permanentes en parallèle avec des infrastructures et des installations de réseaux informatiques de haute performance ou de recherche à très haute vitesse.
- **Définir les compétences essentielles :** Définir les compétences nécessaires aux activités de préservation des données de recherche au Canada. Il peut s'agir

---

<sup>18</sup> <http://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=1609&context=chtlj> (p. 79)



d'offrir des possibilités de formation aux personnes intéressées à faire partie de ce domaine, et de faciliter les discussions et le partage des connaissances entre les experts partout au pays.

## Conclusion

Le GEP entrevoit un avenir où le Canada deviendra un chef de file mondial de la recherche et de l'innovation, en gérant efficacement les actifs de données de recherche à grand volume. L'une des clés de cette vision est l'établissement d'une stratégie nationale efficace pour relever les nombreux défis associés à l'accès significatif aux données de recherche à long terme. En pilotant l'élaboration d'un réseau de préservation numérique distribué pour les données et métadonnées de recherche au Canada, et en s'engageant à collaborer et à faire preuve de transparence et d'ouverture, nous croyons que Portage peut faire une réelle différence dans la vie des Canadiens. Lorsque les données sont préservées efficacement, elles peuvent être partagées et réutilisées plus facilement, ce qui permet aux chercheurs canadiens de s'appuyer sur le travail des autres, de stimuler de nouvelles découvertes et d'accroître la transparence et la responsabilisation au sein de l'entreprise de recherche. Pour reprendre les mots de l'ancien directeur de Portage, Chuck Humphrey : « les nations les plus innovatrices à l'avenir seront celles qui gèrent le mieux leurs données de recherche aujourd'hui »<sup>19</sup>.

---

<sup>19</sup> <https://preservingresearchdataincanada.net/category/rdmi-1/>

## Annexe A. Termes traduits de la Figure 1.

Anglais	Français
Access	Accès
Administration	<i>Administration</i>
Archival Information Package (AIP)	<i>Paquet d'information archivé</i>
Archival Storage	<i>Archivage</i>
Consumer	<i>Consommateur</i>
Data management	<i>Gestion des données</i>
Descriptive info	<i>Information descriptive</i>
Dissemination Information Package (DIP)	<i>Paquet d'information diffusé</i>
Ingest	<i>Intégration</i>
Management	<i>Gestion</i>
Orders	<i>Demandes</i>
Preservation planning	<i>Planification de la préservation</i>
Producer	<i>Producteur</i>
Queries	<i>Requêtes</i>
Result sets	<i>Ensembles de résultats</i>
Submission Information Package (SIP)	<i>Paquet d'information à verser</i>