

Enjeux clés ayant une incidence sur la préservation numérique au Canada – 2021

Un mémoire du Groupe de travail sur la préservation numérique de l'ABRC

Ce document a été préparé par Steve Marks, Corey Davis, Grant Hurley, Rebecca Ross, Roslynn Ross et Dale Askey.

Novembre 2021

www.carl-abrc.ca

Introduction

Le défi de la préservation numérique est un sujet récurrent dans les bibliothèques depuis deux décennies. Par conséquent, le paysage comprend maintenant un certain nombre d'entités des secteurs public et privé qui offrent une gamme d'options d'infrastructure. Un sous-groupe du groupe de travail sur la préservation numérique (GTPN) de l'ABRC a choisi d'approfondir la question de l'infrastructure. Nous avons d'abord pensé à une analyse environnementale; or les principaux acteurs sont déjà bien connus. Nous avons également discuté de la préparation d'une recommandation ou d'une feuille de route pour la création d'une infrastructure nationale canadienne de stockage pour la préservation numérique, mais cet effort semble prématuré sans une consultation plus générale sur certains des défis et des variables.

Le document qui suit positionne la préservation numérique autour de quatre catalyseurs que nous voyons en ce moment :

- L'émergence de l'Alliance de recherche numérique du Canada;
- L'avenir des dépôts numériques fiables (DNF);
- L'état de l'infrastructure de stockage pour la préservation numérique au Canada;
- L'évolution des acteurs sur le terrain, notamment l'ajout des entités commerciales.

Ce document présente l'état actuel, les opportunités et les risques inhérents à chacun de ces quatre domaines. Lorsque vous examinez ces points, nous vous recommandons de vous poser les questions suivantes afin d'orienter vos recherches et vos actions.

- Que voulez-vous que nous examinions de plus près? Où faut-il agir?
- Que devons-nous savoir sur l'état de la préservation numérique au Canada?
- Quel paysage voulez-vous voir?
- Qui contrôle ce paysage ?

Alliance de recherche numérique du Canada

État actuel

- Après des années de consultation et de planification, le gouvernement fédéral a inauguré l'[Alliance de recherche numérique du Canada](#) (anciennement la NOIRN), une organisation qui [reprend les mandats précédents de Calcul Canada, de la FCI, du projet Portage de l'ABRC et de certains aspects de CANARIE](#) (le diagramme au milieu de cette page Web donne certains éclaircissements). Ce

groupe aura une grande visibilité dans l'espace de la gestion de données de recherche, ou GDR, (y compris les aspects de préservation des données) en raison de son financement initial substantiel et d'une augmentation de la notoriété de la [Politique des trois organismes sur la GDR récemment publiée](#).

- L'Alliance a hérité, entre autres, du groupe d'experts de la préservation de Portage (GEP) et souhaite promouvoir les services qu'elle soutient, notamment [Dataverse](#) et le [DFDR](#), non seulement en tant que dépôts d'accès, mais aussi en tant que services de préservation des données.
 - Vous vous souvenez peut-être que le GEP a coécrit (avec le GTPN de l'ABRC) un [exposé de position](#) (en tant que soumission à la consultation de la NOIRN) soulignant les priorités et les acteurs pertinents dans la préservation des données de recherche.
 - Un groupe de travail du GEP a également produit un [rapport avec des recommandations en matière de préservation](#) pour le [portail Dataverse de Scholars Portal](#) impliquant l'association des dépôts institutionnels Dataverse abonnés à un stockage à long terme robuste et répliqué des ensembles de données dans l'[Ontario Library Research Cloud](#) (OLRC), assurant un niveau minimum de stockage fiable des dépôts sous-jacents à Dataverse pour soutenir la «préservation en place».

Opportunités

- Nous prévoyons un engagement accru autour de la gestion des données, y compris la préservation des données. Les bibliothèques ont joué un rôle important dans le développement de l'Alliance en raison de l'initiative Portage de l'ABRC. Les intervenants locaux pourraient donc souhaiter entendre ce que les bibliothèques ont à dire dans les discussions sur la GDR sur les campus.

Risques

- Les risques sont principalement liés au financement et à la stabilité.
 - En tant que nouvelle organisation, la force du mandat de l'Alliance n'est pas encore vraiment connue. Ses recommandations en matière de gestion et de préservation des données seront-elles rigoureusement appliquées par les organismes subventionnaires fédéraux et d'autres instances?
 - Également lié à son nouveau statut, on ne sait pas comment le financement et le mandat de cette organisation résisteront aux gouvernements fédéraux successifs. Les personnes qui dépendent des services de préservation numérique soutenus entièrement ou partiellement par l'Alliance pourraient courir un risque supplémentaire jusqu'à ce que l'on en sache plus.
 - L'Alliance reprend les mandats de plusieurs organismes précédents. Le financement de la gestion des données et des activités de préservation qui

s'y rattachent sera-t-il en mesure de se tailler une place face aux demandes de financement du calcul à haute performance ?

- Quelle est la vision à long terme de l'Alliance pour la préservation des données au Canada ? Cette question n'a pas encore été clairement formulée.

Avenir des dépôts numériques fiables (DNF)

État actuel

- Des changements se préparent pour les deux DNF certifiés du Canada!
 - En 2018, le Groupe de travail sur un DNF du Réseau canadien de documentation pour la recherche (RCDR) a fait une recommandation au conseil d'administration du RCDR pour adopter le DNF de Scholars Portal comme service national de DNF pour les ressources de recherche sous licence des membres du RCDR ; recommandation qui a ensuite été approuvée par les membres à l'AGA 2018 du RCDR. Le RCDR continue de travailler en étroite collaboration avec l'Université de Toronto et Scholars Portal à l'élaboration d'un protocole d'entente qui mettrait en œuvre les recommandations du GTDNF et guiderait les prochaines étapes de cette initiative. Lors de la conférence 2020 du RCDR, les membres du RCDR ont reconfirmé leur engagement et leur intérêt pour ce projet.
 - Le RCDR a récemment mis sur pied un [groupe d'audit du DNF de Canadiana](#) afin d'entreprendre un [audit du DNF de Canadiana](#). De plus, un groupe d'audit communautaire composé de membres du RCDR et d'intervenants accroît encore la capacité canadienne en ce qui concerne les exigences de certification continue des DNF. Depuis la fusion entre Canadiana et le RCDR en 2018, le DNF Canadiana est entièrement intégré en tant que service du RCDR.
- En partie grâce aux [initiatives de Portage/l'Alliance](#), la [certification Core Trust Seal](#) gagne en visibilité et en notoriété auprès de la communauté comme alternative aux DNF traditionnels basés sur la norme ISO16363.
 - Bien que la certification CTS soit issue du monde des dépôts de données (en tant que successeur des programmes de certification Data Seal of Approval et World Data System), il repose sur les mêmes fondements philosophiques que les certifications de DNF basées sur TRAC/ISO16363.

Opportunités

- Le débat sur la certification des DNF semble passer d'un jugement d'une autorité centrale à un consensus de la communauté sur la fiabilité d'un dépôt.

- Une communauté canadienne pour la certification du DNF offre-t-elle une occasion de promouvoir des normes et des pratiques particulières?
- Peut-il y avoir une approche normalisée/harmonisée pour des choses comme la planification des dépôts en cas d'urgence?

Risques

- La plupart des dépôts actuellement actifs dans la certification sont des dépôts partagés. Les intervenants doivent veiller à ce que les conversations qui ont lieu dans ces groupes reflètent les besoins réels de leurs établissements.
- La certification de DNF est une activité qui demande beaucoup de temps et de ressources. L'accent mis sur la certification se fait-il au détriment du travail réel de préservation?

Infrastructure de stockage pour la préservation numérique

État actuel

- Au Canada, il y a actuellement deux principaux fournisseurs de stockage consorcial : [Ontario Library Research Cloud](#) (OLRC) de Scholars Portal et [WestVault](#) de COPPUL.
- OLRC de Scholars Portal est actuellement dans une phase de développement via un projet de renouvellement de l'infrastructure et met en place des nœuds de test dans des établissements canadiens en dehors de l'Ontario. Scholars Portal travaille également à la révision des accords et à la mise en place de nouveaux tarifs pour les universités, les consortiums et les autres établissements de mémoire qui ne font pas partie d'OCUL.

Opportunités

- Existe-t-il des possibilités pour OCUL d'harmoniser les offres de services entre WestVault et d'autres fournisseurs de stockage universitaires? Des services différents pourraient-ils exister pour différentes communautés, mais utiliser des mécanismes techniques communs?
- Si OCUL devient un service plus largement utilisé, peut-il s'intégrer à d'autres plateformes où les données sont absorbées ou traitées?
 - Le service [Permafrost](#) de Scholars Portal est un exemple de la manière dont cela pourrait fonctionner. Scholars Portal a également travaillé sur des intégrations semblables avec son service Dataverse.

- Comme c'est le cas avec les infrastructures pour les produits de base, les collaborations plus importantes sur le plan régional ou national peuvent représenter des économies de coûts substantiels grâce aux économies d'échelle.

Risques

- La structure des coûts des services de stockage administrés localement est souvent différente de celle des fournisseurs commerciaux de services infonuagiques. La méconnaissance de la part de la communauté pourrait conduire à une prise de décision qui expose les établissements de mémoire à de mauvaises surprises telles que des frais de récupération de données lorsqu'elles doivent retirer des données des fournisseurs commerciaux.
- L'absence de prise de décision ferme par les groupes de gouvernance d'infrastructures pourrait faire en sorte que les établissements canadiens manquent des occasions d'adoption et pourrait renforcer la position des fournisseurs de services commerciaux et non universitaires.

Des acteurs en mutation

État actuel

- Internet Archive fait à nouveau des efforts pour s'implanter plus généralement au Canada. Cet acteur est en train de mettre en place un centre de données canadien (sur la côte ouest). Il est en pourparlers avec l'Université de Victoria pour un accord de colocalisation qui verrait les nœuds d'Internet Archive établis dans le centre de données de l'Université de Victoria. Il s'est également montré ouvert à la colocalisation de nœuds dans d'autres établissements bien qu'il ne semble pas y avoir de projets précis à ce sujet. Un protocole d'accord avec l'Alliance est en cours qui prévoit qu'IA agisse en tant que fournisseur de services de préservation pour le DFDR.
- Le RCDR en est aux premières étapes de l'exploration des services de préservation numérique pour les établissements canadiens également.
- Les membres du groupe de travail sur la préservation numérique de l'ABRC constatent une augmentation des discussions et de l'activité des fournisseurs commerciaux dans cet espace, comme l'offre gratuite «[Starter](#)» de Preservica et le [service Perpetua](#) d'EBSCO via Arkivum. Une partie de cette activité découle de l'expansion de leurs offres de services dans le cycle de vie de la recherche, une autre partie est simplement une quête générale.

Opportunités

- Les nouveaux services sont souvent présentés comme des offres «à perte». Si les établissements peuvent utiliser ces services tant que les coûts restent bas,

tout en évitant les pièges comme le verrouillage par le fournisseur, il peut y avoir des possibilités de minimiser les coûts au moins temporairement.

- (Par rapport à Internet Archive) Une présence accrue d'IA au Canada pourrait conduire à de nouvelles opportunités d'intégration avec des services couramment utilisés comme Archive-IT et les services de numérisation d'IA.

Risques

- L'évolution des offres de services et des structures tarifaires des organisations extérieures peut être imprévisible. Cette imprévisibilité peut présenter des dynamiques défavorables à la préservation numérique.
 - Il faut être très conscient des problèmes possibles en matière de compatibilité, car le verrouillage de la part des fournisseurs et les formats protégés sont des tactiques bien établies.
 - Les services commerciaux représentent une opportunité d'offrir l'expertise de préservation en sous-traitance, or l'expertise obtenue par les bibliothèques de cette manière revient rarement.
- (Par rapport à Internet Archive) Certains membres de la communauté de l'ABRC s'inquiètent du [procès en cours contre IA au sujet de son initiative de bibliothèque d'urgence nationale pendant la pandémie de COVID-19](#), car il touche la durabilité des activités d'IA.

Conclusion

Ce document d'information n'a pas pour but de dresser un portrait exhaustif ou une analyse complète de l'état de la préservation numérique au Canada, mais plutôt — comme le titre l'indique — de mettre en évidence certaines questions clés pour lesquelles les bibliothèques doivent prendre des décisions individuellement et collectivement au cours des prochaines années. Avec l'augmentation des besoins de préservation numérique, il faudra observer attentivement la croissance du marché commercial et l'influence d'organisations telles que l'Alliance afin d'éviter la perte de contrôle sur nos biens numériques, l'augmentation excessive des coûts ou l'échec de la préservation numérique. Naturellement, nous devons être conscients que nos stratégies pour certains biens numériques peuvent ne pas correspondre aux besoins et aux souhaits des communautés autochtones. Ces notions clés et d'autres soulevées dans ce document nous ramènent aux questions posées au début de ce document :

- Que voulez-vous que nous examinions de plus près ? Où faut-il agir ?
- Que devons-nous savoir sur l'état de la préservation numérique au Canada ?
- Quel paysage voulez-vous voir ?
- Qui contrôle ce paysage ?

En formulant des réponses claires à ces questions, les bibliothèques et leurs directions pourraient simplifier le travail dans le domaine de la préservation numérique en réduisant une partie du bruit et en concentrant les efforts sur l'élaboration de solutions pour le Canada. Nous avons une réelle opportunité de créer une infrastructure, des services et une expertise nationale robuste pour la préservation numérique et en lien avec celle-ci si nous arrivons à établir un programme clair et travailler collectivement pour le réaliser.