

# @Risk North 2

*Digital Collections*  
*Collections numériques en péril*  
2018.II.09 - MONTRÉAL

## Rapport sur le Forum ouvert

*Publié en mars 2019*

### Contexte/introduction

Une présentation de l'Association des bibliothèques de recherche du Canada (ABRC) en collaboration avec le Réseau canadien de documentation pour la recherche (RCDR), Bibliothèque et Archives Canada (BAC), ainsi que Bibliothèque et Archives nationales du Québec (BAnQ), cet événement a été une réalisation clé du Groupe de travail multilatéral sur la conservation numérique (GTCN) de l'ABRC. Créé en 2017, l'objectif du GTCN consiste à évaluer la capacité, les ressources disponibles et les possibilités de financement présents dans le milieu canadien de la recherche en matière de conservation numérique; déterminer des approches, des normes, des pratiques et des technologies pertinentes et en faire la promotion; établir des relations avec les efforts internationaux en la matière afin de cultiver au Canada une connaissance appropriée du contexte plus large de la conservation numérique.

L'événement était un complément prévu à *@Risk North: Collections en péril*, une réunion semblable tenue un an plus tôt, dont le but était d'examiner la collecte de ressources imprimées partagées à l'échelle régionale et nationale. (Cet événement a puisé son inspiration de la réunion organisée par le forum sur les collections 2016 *@Risk: Stewardship: Due Diligence, and the Future of Print (Gérance : La diligence raisonnable et l'avenir de l'imprimé)* du Center for Research Libraries.)

Avant ce forum ouvert d'une journée complète, ceux qui s'occupent de la conservation numérique dans l'ensemble des établissements canadiens de mémoire avaient peu d'occasions de se rencontrer et de discuter des investissements et de la capacité de conservation numérique au sein de divers établissements de mémoire du pays. Les participants et les présentateurs à l'événement *@Risk North 2* comprenaient des représentants de bibliothèques, d'archives, de musées et d'organismes communautaires dont les exposés et les discussions ont permis d'accroître la compréhension de l'état actuel de préparation, des ressources et des initiatives de collaboration en matière de conservation numérique au Canada, en vue de

déterminer les possibilités de renforcer notre capacité collective de conservation numérique au Canada.

## La situation dans son ensemble

Avant de se pencher sur l'état actuel de la pratique au Canada, Clifford Lynch, directeur général de la Coalition for Networked Information (CNI), a invité les participants à une réflexion globale sur l'évolution des défis liés à la conservation numérique. Au cours de son exposé, M. Lynch a mis les participants au défi « d'élaborer une stratégie globale plutôt que de se concentrer sur le travail qui définit la mécanique de la conservation numérique. »

En se référant à la documentation savante, M. Lynch a affirmé « Il est de plus en plus considéré d'inclure non seulement les résultats de la recherche, mais également les données probantes sur lesquelles la recherche est fondée, ainsi que tous les autres éléments nécessaires, tels que les logiciels, la méthodologie (parfois enregistrée sous forme de vidéo), les flux de travail analytiques. Il s'agit d'un ensemble d'entités raisonnablement délimitées à conserver, même si nous essayons toujours de comprendre ce que nous devons conserver dans les ensembles de données ». Bien que le volume considérable à conserver pose des défis évidents, et bien que le contenu qui n'est pas hébergé par les grands éditeurs soit plus à risque, il a souligné que c'est le contenu culturel général qui est « beaucoup moins déterminé, plus amorphe et fluide et beaucoup moins sûr. » Il a qualifié cette situation de catastrophe qui se déroule lentement.

L'une des raisons qu'il a invoquées, c'est que, alors que les établissements de mémoire pouvaient autrefois acheter et recueillir le contenu culturel qui allait éventuellement devenir la base de futures bourses d'études, les conditions de service des services de diffusion en continu sur le Web nuisent beaucoup aux établissements de mémoire. Que ce soit pour la musique, le cinéma ou les livres électroniques, lorsque les personnes ne possèdent plus de collections, ils ne peuvent plus les léguer aux archives. M. Lynch est également préoccupé par la conservabilité des sites de nouvelles, des médias sociaux et du Web en général. Alors que les sites Web étaient autrefois relativement statiques, ils comprennent désormais presque tous une personnalisation omniprésente, de sorte qu'il est impossible de saisir autre chose qu'une « méthode de conservation de l'échantillonnage ». M. Lynch a postulé qu'il faudra élaborer de nouvelles stratégies pour réagir aux évolutions du marché. Les bibliothèques nationales pourraient être appelées à jouer un rôle plus important; des mesures législatives pourraient être nécessaires; nous devrions sensibiliser les auteurs, les éditeurs et les créateurs à la nécessité de travailler ensemble pour répondre aux besoins actuels et futurs en matière de conservation.

## Un aperçu canadien

Un élément clé de la journée a été l'exposé de Grant Hurley (Scholars Portal) des résultats de *l'Enquête sur la capacité et les besoins en matière de conservation numérique des établissements canadiens de la mémoire* menée par le Groupe de travail sur la conservation numérique de l'ABRC. Les résultats ont été fondés sur des questionnaires remplis par 26 bibliothèques de recherche (établissements membres de l'ABRC) et 25 autres établissements de mémoire (un mélange de petites bibliothèques universitaires, d'organismes communautaires et sans but lucratif et d'organismes gouvernementaux).

M. Hurley a brossé le portrait d'une petite collectivité, mais active, qui s'occupe de nombreux types et formats de collections à grande échelle, mais qui manque de ressources suffisantes pour assurer la maturité et la viabilité des programmes. Les résultats indiquent que l'ensemble des organisations affichent un engagement général à l'égard de la conservation numérique, bien que la plupart aient mentionné que cet engagement ne s'est pas traduit par des ressources pour l'élaboration de programmes. Une tendance semblable a été constatée en ce qui concerne les politiques et les procédures. Bien que 59 % aient déclaré que l'élaboration de procédures documentées est en cours, 75 % ont dit qu'ils manquaient de temps et de ressources pour exécuter ce travail; 47 % ont également répondu que les procédures actuelles sont ponctuelles ou propres à un projet. Dans l'ensemble, l'adoption d'outils de traitement liés à la conservation numérique et à l'expertise judiciaire en informatique était faible. Les plateformes d'accès au contenu numérique constituent un domaine assez bien développé, particulièrement parmi les établissements de l'ABRC, bien qu'il puisse être nécessaire de consolider ou d'intégrer davantage les outils. Les répondants ont indiqué qu'il y a des problèmes de protection de la vie privée et de propriété intellectuelle, et que l'extensibilité des outils existants constitue un obstacle.

Les répondants conservent leurs actifs numériques dans une variété de systèmes et de supports d'entreposage et la transition vers le nuage ou vers d'autres réseaux d'entreposage reproduits est lente, surtout en dehors des grandes bibliothèques universitaires. Une proportion relativement importante de 37 % des actifs est encore conservée sur des supports externes vulnérables. C'est particulièrement le cas pour les plus petites organisations, dont certaines dépendent de disques durs ou d'autres supports amovibles pour entreposer leurs actifs.

Peu de personnel se consacre à la conservation numérique dans l'ensemble, et pour la plupart de ces employés, la conservation ne représente qu'une petite partie de leurs responsabilités. Alors que l'ETP moyen consacré à la conservation numérique dans l'ensemble des établissements était de 1,11, une minorité (8 %) d'établissements ont déclaré un effectif beaucoup plus important (5 employés à temps plein ou plus

chacun). Au total, 65 % des répondants avaient moins d'un ETP pour exécuter le travail de conservation numérique dans tous les rôles applicables énumérés. Bon nombre de répondants étaient optimistes quant à l'augmentation du nombre d'employés dédié à ce domaine, plusieurs mentionnant qu'ils avaient l'intention de réaffecter du personnel actuel à cette fin.

La moitié des répondants ont déclaré que le financement pour la conservation numérique est insuffisant, et plusieurs ont indiqué que la dépendance à l'égard des fonds à court terme pour les activités à long terme était préoccupante.

Un groupe d'experts a commenté ces résultats, et Lisa Goddard (University of Victoria) a fait remarquer que, dans l'ensemble, l'étude montre un bon élan et une bonne coordination à l'échelle nationale et régionale. Le renforcement des capacités à l'échelle des établissements locaux semble être une préoccupation, mais pourrait être une occasion de collaboration, peut-être en s'inspirant des réseaux d'expertise de Portage. Geoff Harder (University of Alberta) a laissé entendre qu'en plus de cette enquête, on peut apprendre beaucoup en documentant nos « échecs retentissants » – des leçons retenues qui pourraient profiter au reste de la collectivité. Il a raconté un cas au cours duquel l'Université de l'Alberta a perdu 11 thèses électroniques en raison d'une série d'événements malheureux. Elle a été en mesure de récupérer les documents, mais seulement en retournant voir les créateurs et en admettant son erreur. Le problème leur a permis d'en apprendre beaucoup et a mené à plusieurs améliorations opérationnelles. M. Harder a laissé entendre que nous devons partager plus ouvertement ce qui fonctionne et ce qui ne fonctionne pas et être prêts à apprendre de nos erreurs. L'enquête laisse entendre que nous faisons bien de collaborer puisque les ressources sont rares, ce qui ne fait qu'accroître le risque. Les outils, les plateformes et les services de collaboration font partie de la solution pour veiller à ce que des mesures de protection soient en place à l'intérieur et à l'extérieur de nos établissements. Mireille Laforce (Bibliothèque et Archives nationales du Québec) convient que la collectivité fait des progrès, mais le manque de ressources financières, de personnes, d'expertise et d'outils demeure frappant. Elle a laissé entendre qu'il serait intéressant de mener cette enquête de nouveau périodiquement, peut-être dans cinq ans.

## **Efforts à grande échelle**

Une partie considérable des exposés de la journée portait sur les efforts « à grande échelle ». La discussion a été présentée par Carole Urbain qui a parlé brièvement du rôle de la Stratégie de numérisation du patrimoine documentaire. Au niveau institutionnel, John Richan a décrit le processus de création d'un nouveau programme de conservation numérique à l'Université Concordia, et Steve Marks a discuté de la possibilité de redéfinir un programme établi à l'Université de Toronto. À l'échelle régionale ensuite, Corey Davis a décrit le système ouvert de partage en consortium

WestVault fondé sur le projet LOCKSS du Council of Prairie and Pacific University Libraries (COPPUL), tandis que Kate Davis (OCUL Scholars Portal) a discuté de certains des défis liés au renforcement de la plateforme du Scholars Portal à l'échelle nationale et de la prestation du nouveau service d'entreposage Permafrost. Pascale Montmartin, de BANQ, et Faye Lemay, de Bibliothèque et Archives Canada, ont parlé du point de vue des bibliothèques et des archives nationales, qui doivent élaborer leurs propres pratiques dans le contexte des normes et façons de faire internationales, tout en jouant également un rôle de chef de file pour les établissements de mémoire du pays et en entreprenant des collaborations avec ces groupes. Karin MacLeod (également de Bibliothèque et Archives Canada) a décrit plus en détail les efforts déployés par BAC pour simplifier les tâches de conservation numérique en acquérant du contenu prêt à la conservation sous forme numérique dans la mesure du possible.

Ces présentateurs ont formulé de sages réflexions et un bon nombre de questions ouvertes : Kate Davis a laissé entendre que les questions de gouvernance plutôt que les questions techniques constituent le véritable défi lorsque l'on décide d'étendre un service à l'échelle nationale. John Richan a demandé « Quelle quantité de traitement suffit? » En expliquant que l'Université de Toronto était passé d'un dépôt et point d'accès monolithique tout en un à de multiples plateformes maintenant axées sur le suivi, Steve Marks a mis en garde contre la « conservation performative » c'est-à-dire la conservation comme une activité ponctuelle plutôt que continue. Corey Davis a fait écho aux commentaires formulés plus tôt par Clifford Lynch au sujet de l'importance d'un système communautaire qui permet de gérer de façon indépendante du contenu décorélé et administré de façon indépendante à plusieurs endroits différents tout en étant en mesure d'assurer l'authenticité et la sécurité de la documentation savante.

Plus tard au cours de la journée, un participant a laissé entendre qu'à la lumière du nombre de documents numériques sans cesse grandissant, la seule façon possible de conserver l'ensemble des actifs culturels numériques du pays sera de s'informer mutuellement et régulièrement du travail que nous exécutons afin d'éviter le chevauchement des efforts à tout prix, de partager les pratiques exemplaires et d'avoir confiance au travail accompli par chacun.

## **Types de contenu précis**

La journée comprenait également des séances décrivant la conservation de certains types de médias ou de contenu et des approches novatrices. Annie Murray (University of Calgary) a parlé de la tâche herculéenne de conserver les 40 000 enregistrements audiovisuels qui composent la collection d'EMI hébergée à l'University of Calgary, tandis que Mireille Nappert (Centre Canadien d'Architecture) a parlé de la conservation des logiciels nécessaires pour rendre accessibles et utilisables les fichiers CAD au moyen de l'émulation. Umar Qasim (University of Alberta) au nom du Groupe

d'experts sur la préservation de Portage, a parlé du sujet opportun de la conservation des données de recherche. Également, Lisa Goddard (University of Victoria) a décrit les efforts déployés dans le cadre du projet Endings<sup>1</sup> visant à examiner la conservation de projets en humanités numériques (un domaine jugé comme étant peu examiné jusqu'à maintenant) au moyen d'une enquête auprès de spécialistes en humanités numériques qui permettra de archiver certains des sites définitifs des projets ainsi que le contenu la base de données qui lui est lié. Cependant, elle a mentionné que la question initiale qu'ils ont posée aux spécialistes en humanités numériques interviewés s'est avérée être révélatrice : En posant la question « Quand et comment les projets en humanités numériques se terminent-ils? », ils ont découvert que souvent, ils ne se terminent pas, ce qui ajoute une complexité inattendue à la question de leur conservation.

## Solutions touchant la programmation

Deux projets ont permis de décrire l'utilisation de la programmation interne pour simplifier les tâches répétitives et chronophages. Jess Whyte (University of Toronto) a cherché à minimiser les erreurs et à simplifier une tâche fréquente lorsqu'elle a créé « Floppycapture.py », un script qui réduit au minimum les dactylographies répétitives lors du traitement des disquettes. Pour sa part, Tim Walsh, de l'Université Concordia, a entrepris de résoudre un problème beaucoup plus vaste. Inspiré par un système utilisé par la police pour trouver des renseignements personnels dans des dossiers volumineux, il est en train de créer un système de « Bulk Reviewer » (actuellement en version alpha), qui permet de trouver, d'examiner et de retirer des fichiers sensibles des archives.

## Échecs, risques et confiance

Au cours des discussions sur les logiciels et les pratiques, trois thèmes connexes récurrents ont été abordés, soit la nécessité d'admettre et d'enregistrer les échecs, la tâche difficile d'établir et de conserver la confiance de ceux dont le travail est conservé et la nécessité de reconnaître et d'atténuer les risques.

Tout comme Geoff Harder nous a exhortés à admettre les échecs plus tôt dans le programme, Clifford Lynch est revenu sur ce thème dans ses commentaires de clôture, ajoutant qu'il faut documenter particulièrement les histoires qui se terminent mal, car elles sont essentielles pour faire valoir le bien-fondé de la conservation.

Dans son exposé sur le travail exécuté dans le cadre du projet « Indigitization »<sup>2</sup>, Sarah Dupont (University of British Columbia) a souligné l'importance d'établir la confiance lorsqu'on travaille à des projets axés sur la conservation des histoires orales

---

<sup>1</sup> <https://onlineacademiccommunity.uvic.ca/endingsproject/>

<sup>2</sup> <http://www.indigitization.ca/>

autochtones. C'est pour cette raison que les subventions d'Indigitization aux groupes communautaires n'exigent pas que le contenu numérisé soit accessible au public, mais laissent plutôt les décisions stratégiques d'accès à chaque groupe.

Cliff Lynch est revenu au thème du risque à la fin de la journée, en insistant sur l'importance de faire preuve d'une grande prudence lors de l'adoption du traitement automatique (p. ex. l'application Bulk Reviewer) pour réduire les exigences en matière de traitement humain. Il a admis qu'en l'absence de tels systèmes automatisés, il serait probablement impossible de conserver le monde numérique à grande échelle, mais a laissé entendre qu'avant de s'engager dans cette direction, nous « devons d'abord bien comprendre où les humains ont tendance à échouer et où les machines ont tendance à échouer. » Nous devons ensuite nous efforcer de définir explicitement ce qui constitue un risque acceptable et d'établir « des pratiques exemplaires en matière de risque pour différentes catégories de collections ».

## Prochaines étapes

En plus des exposés, les participants ont participé à des discussions en table ronde au cours desquelles on leur a demandé comment aller de l'avant pour tirer parti des réussites à ce jour et combler les lacunes actuelles, et examiner le rôle des organisations dans ce travail.

Voici les principales suggestions formulées par les participants :

- Entretenir une communication plus fréquente sur le travail à tous les niveaux (en fonction des collections mais également des pratiques – et ne pas attendre que le travail soit terminé pour partager l'information sur les projets avec d'autres établissements) afin d'éviter le chevauchement des efforts et d'accroître les occasions d'apprendre les uns des autres;
- Mener des activités de suivi de cet événement pour maintenir l'élan; plus d'occasions de se réunir en tant que collectivité, même s'il s'agit simplement de rassemblements virtuels réguliers;
- Accroître les efforts de promotion afin de mieux faire connaître les enjeux, les défis et la nécessité de la conservation numérique auprès des membres de l'administration de haut niveau afin que davantage de ressources puissent être consacrées à cette tâche et que le risque puisse être réduit au minimum;
- Il faut entamer davantage de discussions et déployer plus d'efforts sur la façon d'assurer une représentation diversifiée du contenu qui est conservé;
- Élaborer davantage de stratégies pour le contenu numérique d'origine et une complémentarité plus officielle entre les bibliothèques nationales pour conserver les documents culturels publiés et ceux des bibliothèques de recherche;
- Élaborer une stratégie pour répondre aux besoins de formation en ressources humaines sur les campus définis dans l'enquête du GTCN sur l'état de la

capacité de conservation numérique (p. ex. demande de subvention Connexion du Conseil de recherche en sciences humaines (CRSH) pour offrir de la formation dans diverses régions);

- Continuer d'inclure des organisations de différents types et tailles plutôt que de se limiter à un type d'association ou d'établissement;
- Assurer la coordination de tout ce qui précède.

L'ABRC assurera le suivi de ces suggestions ainsi que des lacunes relevées dans le rapport final de l'enquête sur la capacité de conservation numérique. Cela se reflétera dans la prochaine itération du plan de travail du Groupe de travail sur la conservation numérique et pourrait inclure un éventuel @RiskNorth 3, dont la portée et le calendrier seront examinés attentivement et pour lequel d'autres commentaires de la collectivité seront sollicités.