

Myriam Dumais-DesRosiers
8423164

*Numérisation et préservation d'une partie de la collection de la bibliothèque de Parc Canada :
analyse du projet et recommandations*

Remis à Madame
Inge Albert
Dans le cadre du cours
Préservation de l'information numérique – ISI 6722

Hiver 2016
Université d'Ottawa

Table des matières

Introduction	2
Analyse des besoins	3
Besoins relatifs à l'accès.....	3
Les besoins relatifs à la préservation numérique	4
Analyses des risques	6
Risques et catégories	6
Approche recommandée.....	8
La numérisation.....	8
La préservation et l'accès	9
Conclusion.....	10
Références	12

Introduction

À la suite de récentes réformes, la bibliothèque de Parc Canada (désormais BPC) a centralisé ses opérations et rapatrié ses collections, à l'origine réparties dans les bureaux satellites à travers le Canada. Cette concentration des ressources et des services permet à Parc Canada de mieux atteindre les objectifs liés à son mandat qui est de protéger et présenter le patrimoine naturel et culturel canadien à sa population. Située aujourd'hui dans les bureaux de Cornwall, la collection complète de la BPC contient environ 200 000 items, et est constituée de livres, de périodiques, de rapports, d'atlas, de microfiches, de microfilm et de documents audio/vidéo. Ces ressources sont classées selon les normes de la Library of Congress. Ce rapatriement ne fait toutefois pas l'unanimité de la clientèle de la BPC, employés et usagers, puisque, dans certains cas, les ressources se trouvent maintenant à des milliers de kilomètres. Une solution à ce problème serait la numérisation de la collection de la BPC, ce qui aurait le double avantage de rendre disponible la collection et, dans une optique où le travail est fait en ce sens, de préserver le contenu des collections sous format numérique.

L'organisation a donc le projet de numériser une partie de cette collection et initiera le projet avec les rapports. Cette collection contient 452 volumes dont chaque pièce peut contenir jusqu'à 3 rapports. Pouvant avoir été rédigés en anglais ou en français, ces rapports sont toutefois unilingues. Ils contiennent à la fois du texte et des images, qui devront être pris en compte dans le processus de numérisation. Ces rapports occupent des étagères fixes dans la bibliothèque et peuvent être consultés par les visiteurs sur place. Il est aussi possible d'emprunter un rapport, soit par le biais de la BPC, ou une autre participant au service de prêt entre bibliothèques.

Cette collection, comme la totalité des collections de la BPC, sont gérées par une équipe constituée d'un chef bibliothécaire, d'un bibliothécaire service client, d'un catalogueur/administrateur de système, de deux techniciens bibliothécaires et d'étudiants en stage

Co-op. Les compétences détenues par les employés de la bibliothèque sont donc axées sur la bibliothéconomie, ce qui nous permet de présager qu'aucun membre de l'équipe ne possède une expertise technologique très développée. De même, la BPC ne possède pas l'équipement nécessaire à la numérisation de sa collection ni d'entrepôt numérique dédié sur lequel elle pourrait emmagasiner sa collection numérique. Ces éléments devront être considérés dans le plan de numérisation et de préservation des documents.

Ce travail a pour but de proposer un plan de numérisation des rapports contenus dans la BPC. Ce plan devra introduire dans le processus des éléments permettant la préservation des documents à long terme.

Pour ce faire, nous analyserons, dans un premier temps, les besoins de la BPC en matière de numérisation et de préservation des rapports. Nous nous pencherons par la suite sur les risques inhérents à ces besoins, pour être en mesure, finalement, de recommander une approche axée sur les besoins tout en minimisant les risques.

Analyse des besoins

Besoins relatifs à l'accès

La collection qui fera l'objet de la numérisation, les rapports, a une valeur administrative pour les employés ainsi qu'une valeur patrimoniale pour le reste de la population canadienne. Elle doit donc être facilement accessible puisque, dans le premier cas, elle constitue un matériau de travail, et dans le second, un héritage renfermant des informations sur l'histoire du peuple canadien. La collection n'étant pas soumise à une classification de sécurité, nous n'avons pas à restreindre certaines données à la population, et par conséquent, les fichiers n'ont pas à être protégés par des codes et des mots de passe, à l'exception de l'inscription nécessaire au moteur de recherche AMICUS, qui permet d'accéder à tous les documents de la BPC, mais aussi à ceux de plusieurs autres centres de documentations à travers le Canada.

L'aspect collectif du moteur de recherche soulève un point important : la BPC, en plus de ne pas posséder son propre moteur de recherche, n'a pas une complète autorité sur le système d'exploitation de sa bibliothèque (Vubis Web). En effet, le propriétaire de la plateforme est le ministère du Patrimoine, ce qui restreint les possibilités de gestion de l'administrateur du système de la BPC. De plus, la BPC ne dispose d'aucun entrepôt numérique dédié pour entreposer ses documents électroniques. L'équipe utilise des disques durs partagés pour stocker l'information numérique, incluant les collections. Le fait que la BPC ne soit pas la principale gestionnaire de la plateforme internet, de même que l'absence d'entrepôt dédié peuvent poser problème pour assurer une diffusion et une préservation du contenu numérique adéquates. L'absence d'équipement pour la numérisation est aussi un aspect important à considérer, puisque la BPC aura besoin d'acquérir ou de louer l'équipement ou les services, ce qui ajoutera aux coûts du processus.

La numérisation en vue de la diffusion et de la préservation d'une collection demande aussi une expertise à l'interne. En effet, une personne doit être désignée comme étant responsable de la collection et de son maintien. Cette personne doit posséder les compétences nécessaires pour effectuer ce type de travail et doit donc être formée en conséquence. Cette personne pourrait ainsi s'assurer du bon déroulement du processus, de la numérisation de la collection jusqu'à la maintenance.

Enfin, nous devons aussi considérer l'ajout de métadonnées de préservation afin d'assurer aux rapports numérisés une accessibilité et une authenticité à long terme. Ces métadonnées, ajoutées à celles existantes, pourront certifier le contenant, et par le fait même le contenu, pour les années à venir.

Les besoins relatifs à la préservation numérique

Les rapports à numériser contiennent de nombreuses facettes qui devront être analysées dans un contexte de préservation et des choix devront être faits. Dans un premier temps, nous

pensons que la numérisation des rapports ne devrait pas tenir compte des volumes dans lesquels ils se trouvent. L'absence de lien entre les rapports contenus dans un même volume nous incite à penser que la filiation ne met en place aucun contexte. Les rapports devraient donc être numérisés individuellement, ce qui faciliterait la recherche. Une indication devrait toutefois être ajoutée pour permettre à l'utilisateur qui aimerait consulter la version papier de la retrouver sans difficulté. Une fois ce questionnement résolu, nous aurons à déterminer les composantes mêmes du rapport que nous jugeons essentielles et qui, par conséquent, devront être conservées.

Du point de vue physique de l'objet, nous notons que la reliure du document est peu solide (de type « collé »). Par conséquent, il sera important que les personnes manipulant les rapports en vue de les numériser le fassent avec soin pour éviter de perdre des pages et ainsi endommager le document. Cela nécessitera donc une expertise certaine pour la manipulation de ces documents. De même, une connaissance précise du nombre de rapports contenus dans les 452 volumes permettra une meilleure estimation des coûts. Enfin, il faudra un endroit fiable pour déposer les fichiers numérisés afin d'assurer une meilleure accessibilité et une meilleure préservation à long terme.

Le contenu de chacun des rapports devra aussi faire l'objet de choix quant à la préservation. Dans un premier temps devra être considéré le choix de mettre l'accent sur la préservation du texte ou des images. En effet, le choix du type de contenu à favoriser aura un impact sur le choix du fichier à privilégier. Dans ce cas-ci, nous pensons qu'il vaudrait mieux privilégier le texte aux images, puisque les rapports se veulent des comptes rendus de ces images. Ainsi, le texte est porteur d'information que viennent appuyer les images. Nous pensons aussi que le texte et les images doivent être numérisés ensemble pour conserver le contexte des informations contenues dans le rapport.

La planification du processus de numérisation et de préservation devra aussi être élaborée en tentant de minimiser les coûts le plus possible. En effet, Parc Canada doit diviser son budget en

plusieurs catégories pour arriver à remplir son mandat et par conséquent les sommes allouées pour le projet de numérisation ne devraient pas être très importantes. Toutefois, le cadre légal n’opposera pas de restriction, puisqu’aucun droit d’auteur ne devra être versé, les auteurs des rapports étant des employés de Parc Canada (ou du gouvernement canadien), donc les droits sont déjà détenus par les organisations fédérales.

Les besoins de la BPC sont donc constitués d’environ 8 composantes. Dans un premier temps, la planification devra tenir compte de l’accès à la collection, de la technologie nécessaire au processus de numérisation et de préservation, de l’expertise humaine et des métadonnées. Enfin, le processus de numérisation devra considérer le support fragile des rapports, favoriser la numérisation du texte, bien que les images doivent être préservées avec le texte pour que soit conservé le contexte. Un endroit sécuritaire devra être prévu pour déposer les fichiers numérisés. Le tout devra se dérouler sous le sceau des contraintes budgétaires de l’organisation pour que le projet soit réalisable.

Analyses des risques

Risques et catégories

Les besoins liés à la numérisation et à la préservation d’une partie de la collection de la BPC ne viennent pas sans risques. Nous retrouvons dans cette section les risques associés aux processus de numérisation et de préservation pour la BPC sous forme de tableau :

Catégories	Risques
Risques humains	Aucune personne responsable de l’aspect numérique de la collection
	Manque d’expertise du personnel
Risques technologiques	Inadéquation de la ressource selon le besoin (ex. format incompatible avec le système utilisé)
	Absence de plan de migration (logiciel, format de fichier)
	Désuétude de l’équipement à long terme (matériel informatique ordinateur, numériseur, etc.)

Risques des métadonnées	Insuffisance ou inexactitudes des métadonnées compromettant l'accès ou l'authenticité du document
Risques de l'objet à numériser	Risques de bris de l'objet
	Risques de pertes d'information : pas d'entrepôt dédié, aucun endroit où déposer les fichiers à l'exception des disques durs partagés.
Risques financiers	Fonds insuffisants pour mener à bien le projet numérisation
	Augmentation des frais de maintenance
	Incapacité à assurer la préservation financièrement sur plusieurs années

Les risques à la numérisation sont donc nombreux. Ils n'ont toutefois pas la même importance dans le contexte qui nous intéresse. Le tableau plus bas permet de les adresser hiérarchiquement selon l'impact qu'ils auraient sur le projet, de même que sur la possibilité qu'ils surviennent lors du processus (DRAMBORA).

Risques	P	G	Commentaires
Aucune personne responsable de l'aspect numérique de la collection	2	4	6 : Réduction du risque nécessaire. Une personne doit être nommée responsable sous peine de voir le projet échoué.
Manque d'expertise du personnel	4	4	8 : Réduction du risque nécessaire. L'absence de compétence du personnel peut aboutir à des erreurs aussi graves que la destruction de la collection.
Inadéquation de la ressource selon le besoin (ex. format incompatible avec le système utilisé)	2	2	4 : Réduction du risque non nécessaire. En axant nos choix sur l'interopérabilité, on assure une accessibilité très grande aux collections.
Absence de plan de migration (logiciel, format de fichier)	3	3	6 : Réduction du risque nécessaire. L'absence de plan de migration expose nos données à des dangers tels que la corruption, l'impossibilité d'accès, etc.
Désuétude de l'équipement à long terme (matériel informatique ordinateur, numériseur, etc.)	3	1	4 : Réduction du risque non nécessaire. Dans la mesure où les formats de fichiers sont choisis en fonction de leur interopérabilité.
Insuffisance ou inexactitudes des métadonnées compromettant l'accès ou l'authenticité du document	3	4	7 : Réduction du risque nécessaire. Les métadonnées sont un aspect crucial de la préservation pour assurer l'authenticité et la fixité du contenu. Les membres de l'équipe, au fait des métadonnées nécessaires à la gestion de la collection, le sont peut-être moins concernant celles liées à la préservation.
Risques de bris de l'objet	2	4	6 : Réduction du risque nécessaire. Diminuer, voire éradiquer si on détient l'expertise nécessaire.
Risques de pertes d'information	4	4	8 : Réduction du risque nécessaire. L'absence d'entrepôt numérique dédié nuit non seulement à la

			diffusion de l'information, mais aussi à la préservation.
Fonds insuffisants pour mener à bien le projet numérisation	4	4	8 : Réduction du risque nécessaire. Les choix devront être faits en conséquence du budget limité pour conclure le projet de façon efficace.
Augmentation des frais de maintenance	4	2	6 : Réduction du risque nécessaire. Cependant, si cet aspect est considéré dans les fonds alloués au projet.
Incapacité à assurer la préservation financièrement sur plusieurs années	4	4	8 : Réduction du risque nécessaire. Le financement doit impérativement tenir compte des coûts que le projet va engendrer à long terme si on veut être en mesure d'accéder à nos données pour de nombreuses années.

P : probabilité
G : gravité

Un premier constat s'impose : si l'on veut mener à bien le projet de numérisation des rapports, il faudra gérer de nombreux risques dont l'importance pourrait avoir un impact très important. En effet, des 11 risques répertoriés et analysés, seulement 2 peuvent faire l'économie d'une action concrète, et ce, uniquement si les autres risques sont gérés adéquatement. Ce haut niveau de risque sera pris en compte dans les recommandations.

Approche recommandée

La numérisation

En considérant les contraintes budgétaires et le manque d'expertise à l'intérieur de l'organisation, nous recommandons à la BPC de recourir à une compagnie spécialisée pour procéder à la numérisation de leur collection de rapports. En effet, l'acquisition d'un tel équipement coûte extrêmement cher, auquel on doit ajouter les coûts de maintenance. En louant les services d'une compagnie, elle épargne des frais importants qu'elle peut allouer à une autre ressource. De même, elle évite les erreurs humaines liées à un manque de compétence, puisqu'elle recourt à des experts. N'étant pas responsable des problèmes pouvant survenir lors de la numérisation, elle évite la création de documents numérique de piètre qualité, le risque reposant sur la compagnie, qui doit livrer un produit fini (transfert du risque).

La préservation et l'accès

Toutefois, il importe de conserver en tête que la numérisation et la préservation doivent aller de pairs, et que la préservation commence dès la création du document numérique. Par conséquent, nous recommandons d'employer un format de fichier ouvert tel que le format PDF/A qui est un format standard ouvert et libre de droits (BANQ, 2010, p. 28), et plus précisément le PDF/A-3 (norme ISO 32000-1), qui permet d'encapsuler l'information concernant le PDF à l'intérieur même du fichier (Library of Congress, 2014). Ainsi, les données telles que les métadonnées et le support nécessaire pour lire le fichier sont enregistrés à même le fichier, comme le recommande le modèle OAIS, norme en matière de préservation des documents numériques.

Les métadonnées de type « préservation » à rattacher au document devraient être extrait du modèle PREMIS – dictionnaire de données nécessaires à la préservation d'un document numérique – qui suit la norme OAIS. Recommandé par Groenewald et Breytenbach (2010), ce modèle ne s'intéresse qu'aux constituants numériques du document (l'objet, les événements, les agents et les droits [Premis, 2015, p. 6]) et ne tient pas compte du contenu intellectuel. Son principal avantage, outre sa spécialisation, est d'être doublement accessible : 1. pour les spécialistes de l'information qui reconnaîtront le langage employé ; 2. pour les organisations à petits budgets, puisqu'il est accessible gratuitement sur internet.

Ces documents numérisés, une fois créés et bien identifiés, devront être entreposés dans un endroit fiable pour éviter la perte de l'information et assurer une diffusion et une préservation à long terme. Bien que la conservation d'une copie sur le disque dur permette la préservation, elle n'est pas optimale pour la diffusion, qui est le but premier de la numérisation (Conseil canadien des archives, 2002, p. 1). Nous recommandons donc fortement d'investir dans un entrepôt dédié pour le stockage des collections. Le fait que la BPC ne soit pas la principale gestionnaire de sa plateforme de gestion des documents n'est pas étranger à cette recommandation. En étant

propriétaire de son propre entrepôt, elle s'assure d'offrir une accessibilité accrue à sa collection tout en conservant le contrôle nécessaire sur ses fichiers pour en assurer une préservation adéquate et de longue haleine. Pour constituer un entrepôt fiable, il serait pertinent de consulter le modèle TRAC, qui octroie la certification de dépôt numérique fiable. Sans répondre à tous les critères de TRAC dès sa création, ce qui serait ambitieux, constituer l'entrepôt selon ce modèle – inspiré du modèle OAIS – instituerait dès le départ des pratiques exemplaires contribuant à la mise en place d'un entrepôt sécuritaire et efficace. Cette norme est aussi disponible gratuitement sur internet, ce qui ajoute à la valeur du modèle.

Enfin, nous ne saurions insister sur l'importance qu'un employé soit chargé de la collection numérique, de la création au maintien de la collection numérique. Il doit recevoir une formation qui lui permettra de mener à bien le projet sur de nombreuses années et être remplacé lors de son départ. Nous recommandons donc que la responsabilité soit associée à la fonction plutôt qu'à la personne pour assurer une continuité logique du travail.

Conclusion

Dans le cadre du projet de numérisation d'une collection de la BPC, nous avons considéré les besoins spécifiques à l'organisation et les risques y étant associés. Nous avons conclu que les besoins relevaient des catégories financière, humaines et technologiques. En effet, nous avons besoin d'un projet nécessitant peu de frais, effectué par une main d'œuvre de qualité et effectué au moyen de procédés technologiques concordant avec les objectifs de numérisation. Les risques découlant de ces besoins concernent plus particulièrement les procédées de numérisation (bris du matériel, erreur dans le choix des formats de fichiers et supports de stockage), du processus de préservation (métadonnées erronées ou absentes, aucun lieu fiable pour entreposer les fichiers), de même que la gestion humaine (manque d'expertise et absence d'une personne dédiée au projet). À cela s'ajoutent les risques financiers, qui menacent le projet à chacune des étapes.

Nos recommandations se basent en majorité sur le modèle OAIS, qui est une norme en matière de préservation numérique. Nous avons suggéré des outils dérivés de ce modèle pour chacune des étapes du processus (choix de format, de plateforme de stockage, métadonnées, etc.) Employés à bon escient, ces outils permettront à la BPC de constituer une collection numérique basée sur des pratiques exemplaires dont l'accessibilité, l'authenticité et la fixité seraient protégées pour de nombreuses années. Ces outils, en grande partie accessibles sur internet, et parfois même gratuits, viendraient pallier le manque d'expérience technologique actuelle de l'équipe de la BPC, sans pour autant faire augmenter les dépenses. Peu importe les choix que fera la BPC, nous lui recommandons surtout de nommer un responsable, formé pour la tâche, à la tête du projet et d'allouer un budget suffisant à la maintenance des documents numériques pour les années à venir.

Références

- Bibliothèque et Archives nationales du Québec (2010). La numérisation des documents administratifs : méthodes et recommandations. Repéré à http://www.banq.qc.ca/documents/services/archivistique_ged/Numerisation_des_documents.pdf
- Conseil canadien des archives. 2002. *Numérisation et archives*. Repéré à http://www.cdncouncilarchives.ca/digitization_fr.pdf
- Digital Curation Centre and Digital Preservation Europe (Février 2007). *DCC and DPE Digital Repository Audit Method Based on Risk Assessment*, v1.0. Repéré à <http://www.repositoryaudit.eu/download>
- Groenewald, R., and A. Breytenbach. (2011). The Use of Metadata and Preservation Methods for Continuous Access to Digital Data. *The Electronic Library* 29, no. 2, pp. 236-248.
- Library of Congress. (2014). Sustainability of Digital Formats Planning for Library of Congress Collections : PDF/A-3, PDF for Long Term Preservation, Use of ISO 32000-1, with Embedded Files. Repéré à <http://www.digitalpreservation.gov/formats/fdd/fdd000360.shtml>
- Preservation Metadata : Implementation Strategies (PREMIS). (2015). Data Dictionary for Preservation Metadata. Repéré à <http://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>
- The consultative committee for space data system. (2012). *Reference Model for an Open Archival Information System (OAIS) : Magenta Book*.
- US Center for Research Libraries. (2007). *Trusted Repositories Audit & Certification (TRAC)*. Repéré à http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf