Canadian Research Knowledge Network

Réseau canadien de documentation pour la recherche

Cadre stratégique pour la préservation du DNF du RCDR

Approuvé par le Comité de préservation et d'accès : 26 juillet 2021

À propos de ce document

Le présent document décrit l'approche adoptée par le RCDR pour assurer la préservation à long terme des contenus numériques stockés dans le Dépôt numérique fiable (DNF) et l'accès à ces contenus. Il vise à orienter l'élaboration, la révision et la mise en œuvre des politiques et des pratiques relatives à la préservation numérique.

Ce cadre est basé sur les premiers travaux de Canadiana.org dans l'élaboration du DNF et évoluera au fil du temps pour s'adapter aux transformations du contexte de la préservation numérique.

Cadre stratégique pour la préservation

La stratégie du RCDR pour la préservation à long terme des contenus numériques au sein du Dépôt numérique fiable de Canadiana (« le Dépôt ») repose sur cinq approches fondamentales.

1. Intégrité du processus

La préservation des contenus numériques exige que les processus du Dépôt fonctionnent de manière prévisible et cohérente. Pour ce faire, le RCDR :

- Maintient des pratiques, des processus, des procédures, des spécifications et des normes documentés pour faire en sorte que les opérateurs du Dépôt (tout le personnel impliqué dans l'administration ou l'exploitation du Dépôt) s'acquittent de leurs tâches de manière cohérente et prévisible, et permettent l'examen et l'audit;
- Procède régulièrement à un examen interne et externe et à une mise à jour de toutes les procédures et de leurs systèmes associés pour s'assurer que les processus sont pertinents, à jour et respectés. Ce processus d'examen comprend un audit interne au moins une fois tous les deux ans et des audits externes si le Comité de préservation et d'accès le juge approprié.

2. Intégrité des données

La préservation des contenus numériques exige que l'intégrité des données numériques sous-jacentes soit sauvegardée et vérifiée. Pour ce faire, le RCDR :

- Maintient un système de préservation distribué, qui garantit qu'une copie de chaque objet numérique est stockée dans au moins trois lieux distincts, afin de s'assurer que les dommages ou la perte d'un seul nœud n'entraîneront pas une perte de données permanente;
- Effectue un contrôle automatique continu de la stabilité pour détecter les données manquantes ou corrompues, conjugué à la réplication et à la réparation automatiques des informations perdues ou corrompues afin de garantir que les défaillances cumulatives non détectées n'entraînent pas de perte ni de corruption;
- Examine et vérifie les métadonnées structurelles des objets numériques pour s'assurer qu'elles peuvent être identifiées et utilisées en dehors du contexte de leur AIP;
- Préserve la structure des objets numériques grâce à des métadonnées qui peuvent être liées sans ambiguïté à l'objet;

- Utilise pour les objets numériques et les métadonnées des formats de fichiers ouverts qui sont bien pris en charge au moment de leur adoption, qui devraient le rester pendant une longue période et qui devraient disposer d'une voie de migration viable lorsqu'ils sont remplacés par des formats plus récents;
- Valide et normalise les formats et le contenu des fichiers à l'ingestion afin de s'assurer que le Dépôt héberge uniquement des objets de types connus et pris en charge;
- Effectue des analyses environnementales régulières du niveau de prise en charge de tous les formats de fichiers et de données préservés, afin de détecter les obsolescences imminentes et d'alerter les déposants des formats à risque;
- Fait migrer les formats en fin de vie vers des formats plus récents, lorsque cela est possible, dans le but de préserver autant que possible la structure et le contenu d'origine, en conservant le contenu migré aux côtés des originaux, de même que la documentation du processus utilisé pour créer le premier à partir du second.

3. Intégrité de l'infrastructure

La préservation des contenus numériques exige que les systèmes sur lesquels les contenus sont hébergés soient fiables, sûrs et prévisibles dans leur fonctionnement. Pour ce faire, le RCDR :

- Rafraîchit les systèmes de façon continue au moyen d'un processus de modification progressive, afin que l'infrastructure soit toujours moderne et fiable.
- Normalise les logiciels et le matériel à un nombre limité de configurations, ce qui permet de maintenir la complexité et la variabilité globales de l'infrastructure à un faible niveau, simplifiant ainsi la capacité à comprendre et à entretenir l'infrastructure;
- Normalise et automatise le déploiement et la gestion des logiciels et des systèmes, pour faire en sorte que le Dépôt soit configuré et fonctionne de manière cohérente et prévisible;
- Privilégie les composants standards de l'industrie et les logiciels libres afin de réduire le risque de dépendance à l'égard de systèmes et d'outils non pris en charge, sans issue ou mal compris;
- Assure une formation polyvalente du personnel impliqué dans les opérations liées au Dépôt afin de garantir la redondance de toutes les opérations clés et des domaines de connaissances, offrant ainsi une certaine résilience en cas de perte ou de rotation du personnel et permettant une vérification croisée des tâches;
- Gère et surveille l'accès aux infrastructures physiques afin de réduire le risque de perte ou de dommage, que ce soit de façon intentionnelle ou accidentelle.

4. Origine des objets

La préservation des contenus numériques et la fourniture d'informations utiles et fiables aux parties prenantes exigent que les objets puissent être retracés jusqu'à leurs sources. Pour ce faire, le RCDR :

- Documente et vérifie les processus de traitement des données et d'enregistrement des opérations effectuées sur les données afin de fournir une piste d'audit pour examiner, tester et reproduire les modifications apportées aux données et la création de dérivés;
- Associe chaque objet d'archivage (AIP = *Archival Information Package*) à un déposant en particulier et tient à jour les ententes avec les déposants afin de garantir que les objets sont liés à leurs propriétaires;
- Crée et maintient des métadonnées décrivant l'emplacement des objets originaux à partir desquels les versions numérisées sont créées, afin de faciliter la localisation des documents originaux à des fins de comparaison et de vérification;
- Maintient des processus documentés, testés et reproductibles pour la création de DIP (*Dissemination Information Packages*) et des journaux de création de DIP garantissant que leur authenticité peut être établie;
- Enregistre toutes les transformations effectuées sur les objets de préservation, jumelées à la conservation de copies des formats précédents, afin de permettre la validation et la réplication des activités de migration.

5. Accès à l'information

L'accès permanent à des informations utiles et fiables est l'objectif ultime de la préservation numérique. Pour ce faire, le RCDR :

- Privilégie les normes et les formats ouverts largement utilisés, ainsi que la validation structurelle des formats de fichiers lors de l'ingestion afin de réduire la probabilité d'incompatibilité avec les outils actuels et futurs et de maximiser la probabilité que les formats puissent être transférés en cas de besoin;
- Utilise un nombre limité de formats et de schémas pour réduire le nombre de cas d'utilisation et de scénarios qui doivent être pris en compte et testés pour garantir un accès continu;
- Fournit un accès public aux contenus du Dépôt conjugué à une rétroaction et une communication continues avec les utilisateurs et les parties prenantes, afin de détecter les incapacités du Dépôt à répondre aux besoins d'information des parties prenantes;
- Effectue des tests continus de l'accès et de la récupération, pour vérifier que les systèmes d'accès fonctionnent comme prévu.

Annexe A : Glossaire

Archival Information Package (AIP)	Une structure constituée de tous les fichiers composant un objet numérique préservé et de ses métadonnées associées et autres données de contrôle.
Dissemination Information Package (DIP)	Un objet ou un ensemble d'objets dérivés d'un AIP pour la diffusion à un système ou à un utilisateur final.

Annexe B : Historique des modifications

2021-07-26	Première version