

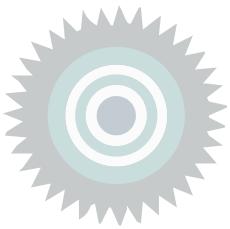
Module 7B - section 5 : La préservation des documents numériques au stade courant

Natasha Zwarich PIAF © AIAF

v1 25/07/2024

Table des matières

Objectifs	3
Introduction	5
1. Enjeux de la préservation numérique au stade courant	6
1.1. Les enjeux organisationnels	6
1.2. Les enjeux technologiques	6
1.3. Les enjeux archivistiques	7
1.4. Les enjeux liés à la nature même du document numérique	7
2. La planification de la préservation	8
2.1. La sélection des documents	8
2.2. Le registre des documents numériques	9
3. Formats de fichiers	10
3.1. Les deux grandes catégories de formats de fichiers	10
3.2. Les critères de sélection des formats de fichiers	10
4. Supports d'enregistrement V -	12
5. Stratégies de préservation	13
5.1. La migration	13
5.2. L'émulation	13
5.3. La conservation de la technologie	14
5.4. La copie sur papier (conversion)	14
Conclusion	15
Glossaire	16
Bibliographie	17



Objectifs

Alors que le numérique occupe une place très importante dans le quotidien de l'archiviste, ce module vise à porter un regard sur les opportunités et les défis que posent la gestion et la conservation des documents numériques. En effet, il traite des questions fondamentales relevant de la gestion des documents numériques ainsi que des interventions pouvant être réalisées par les archivistes. En tenant compte des aspects conceptuels, fonctionnels et techniques de la création, la description, la classification et la gestion au moyen de diverses plateformes, on y présente les principales méthodes d'analyse propres aux documents numériques et les principales normes nationales et internationales applicables. Ainsi, ce module permet de mieux comprendre les diverses facettes de la gestion des documents numériques.

L'apprenant doit être en mesure de :

- décrire les caractéristiques des documents numériques;
- évaluer différentes approches, méthodes et technologies pour assurer la gestion des documents numériques;
- prendre en compte les leviers ou les difficultés que pose l'environnement organisationnel et technique.
- avoir une connaissance des différents niveaux d'intervention des archivistes dans la gestion des documents numériques.

Il nous semble important de préciser que ce module se veut une initiation à la gestion des documents numériques. Par ailleurs, la gestion des documents numériques requiert l'intervention de divers acteurs, notamment les professionnels des technologies de l'information et les usagers. L'archiviste doit donc s'appuyer sur des compétences et des savoir-faire spécialisés qui sont complémentaires à l'archivistique. L'archiviste doit aussi disposer d'une vue globale complète et d'une bonne connaissance de tous les besoins afin d'être un acteur incontournable de la gestion des documents numériques. Les connaissances et les méthodes de gestion des documents numériques sont en constante évolution. Par conséquent, ce module vise à présenter des bonnes pratiques. Celles-ci doivent cependant être adaptées selon les contextes organisationnels, la technologie disponible et les acteurs impliqués.

Les notions abordées dans ce module peuvent être complétées avec d'autres modules de formation du PIAF. Comme les principes de l'archivistique demeurent valides avec le numérique, il est recommandé de consulter le module 5 – Gestion et traitement des archives courantes et intermédiaires. De plus, les problématiques liées à la conservation à long terme sont abordées dans la suite du module 7 (7C) – *Préservation numérique*.

Le module 9 – Section 2 : *Numériser les documents* est un complément de formation pertinent puisqu'il présente les techniques de base pour l'utilisation des moyens de numérisation et le transfert de support vers le numérique.

Le module, « *Un autre regard* » : gérer les archives courantes : approche pratique peut également offrir à l'apprenant des suggestions quant aux niveaux d'intervention à adopter dans la mise en œuvre de dispositifs de gestion des archives courantes.

Le module 14 - *Partenariat et coopération* fournit des suggestions pour mener un projet de gestion des documents ainsi que sur les façons de développer des collaborations entre professionnels.

Objectifs

Finalement, le glossaire du PIAF doit être consulté pour les définitions des termes employés dans ce module de formation.



Introduction

La préservation des *documents numériques*^{p.16} pose de grands défis pour l'archiviste. La mise en place de mécanismes de préservation doit débuter dès la création des documents numériques afin qu'ils demeurent intègres et accessibles durant tout le cycle de vie. Nous aborderons, dans cette section, certaines bonnes pratiques en matière de préservation des documents numériques. Les dispositifs servant à assurer la préservation et la conservation des archives numériques à long terme seront abordés plus en détail dans le module 7C « *Préservation numérique* ».

1. Enjeux de la préservation numérique au stade courant



Introduction

La *préservation numérique* ^{p.16} « réfère à la série d'activités jugées nécessaires pour garantir un accès continu aux objets numériques aussi longtemps que nécessaire » (Digital Preservation Coalition, 2021, p. 282). Ainsi, il importe d'assurer l'accessibilité aux documents numériques tout en limitant l'impact négatif que les changements technologiques peuvent avoir sur ces derniers. Afin de planifier les stratégies de préservation, l'archiviste doit connaître les défis qui peuvent se poser. Les défis rencontrés peuvent être de diverses natures.

1.1. Les enjeux organisationnels

Tout d'abord, l'archiviste peut être confronté à des enjeux d'ordre organisationnel :

- **Ressources** : La préservation numérique, même au stade courant, requiert l'investissement de ressources, notamment des ressources matérielles, technologiques et financières.
- **Parties prenantes** : Dans la mise en place de stratégies de préservation numériques, plusieurs parties prenantes peuvent être impliquées, dont la direction, les créateurs de documents, l'archiviste, les technologies de l'information, des tiers, etc. Ainsi, il faut assurer la collaboration et la coordination des divers acteurs par la définition de rôles et responsabilités. Il importe d'identifier les personnes responsables de la préservation.
- **Temps** : Il est important de planifier les initiatives de préservation numérique dès la création afin d'assurer le maintien de l'accessibilité des documents et que les mesures de préservation soient mises en place en temps opportun.
- **Législation** : Les activités de préservation doivent être mises en place dans le respect de la législation en vigueur, notamment l'accès à l'information, la protection des renseignements personnels ainsi que le droit d'auteur.

1.2. Les enjeux technologiques

Ensuite, l'archiviste sera assurément confronté à des enjeux d'ordre technologique. Ces enjeux sont multiples, mentionnons :

- **Équipement** : Les plateformes numériques évoluent rapidement et elles sont souvent dépendantes de plusieurs autres applications. Il faut également considérer la diversité des outils technologiques et leur obsolescence.
- **Formats de fichiers et supports de stockage** : Les formats de fichiers et les supports de stockage évoluent très rapidement également. En plus, plusieurs types de formats et de supports existent offrant différentes garanties de conservation à long terme.

1.3. Les enjeux archivistiques

Puis l'archiviste doit considérer les enjeux d'ordre archivistique :

- **Intégrité et authenticité** : Le maintien de l'intégrité et l'authenticité des documents numériques durant tout le cycle de vie.
- **Accessibilité** : Le maintien de l'accessibilité du contenu des documents numériques durant tout le cycle de vie.
- **Qualité et contexte** : L'importance de conserver la qualité et le contexte du document numérique par le biais de métadonnées de qualité.

1.4. Les enjeux liés à la nature même du document numérique

Enfin, l'archiviste doit considérer les caractéristiques du document numérique lui-même. **Le document numérique est dépendant d'un ordinateur, d'applications et de périphériques** pour qu'il soit accessible et fonctionnel. Le document numérique possède une présence physique, soit « une série de zéros et de un, ou de bits (chiffres binaires) » (...) qui « doivent être saisis et conservés dans le temps, sans perte ni dommage, pour assurer la survie des objets numériques » (Digital Preservation Coalition, 2021, p. 25). Le document numérique comprend également un contenu, une présentation ainsi que des fonctionnalités.

Certains documents ne peuvent être maintenus dans le temps que sous forme numérique puisqu'il n'existe pas d'équivalent analogique comme le papier qui permettrait la conservation de toutes ses composantes dont les métadonnées et les fonctionnalités.

Conclusion

Ainsi, les activités de préservation doivent être planifiées et ce, dès la création des documents numériques. Cette planification doit nécessairement considérer l'ensemble des enjeux énoncés précédemment.



2. La planification de la préservation

Introduction

Puisque la préservation numérique se planifie dès la création des documents, nous aborderons, dans ce chapitre, la planification de la préservation, principalement la sélection des documents à préserver ainsi que la documentation des fichiers.

2.1. La sélection des documents

Dans le chapitre précédent, nous avons abordé l'évaluation des documents numériques. Parmi tous les documents produits et reçus par un organisme, certains documents ont une plus grande valeur que d'autres. Ainsi, dès le début du cycle de vie des documents numériques, il est possible d'identifier ceux qui seront à préserver pour une longue période. Plus les documents numériques sont importants et ont de la valeur pour l'organisme, plus leur préservation est nécessaire. Ça peut être le cas des procès-verbaux d'un organisme qui retracent les décisions prises ou encore les documents constitutifs. Il s'agit là de documents qui ne doivent pas être perdus. Il est donc nécessaire pour l'archiviste d'être proactif pour leur préservation.

Pour aider dans la sélection des documents, des organismes rendent disponibles des arbres de décision. Outre la valeur des documents, ces arbres de décision aident à évaluer la capacité de l'organisme à réaliser la préservation des documents numériques et vérifier s'il détient les ressources nécessaires pour mener de telles activités.

Exemple

Il est possible de consulter ici un exemple d'un arbre décisionnel produit par le Réseau canadien d'information sur le patrimoine du Gouvernement du Canada pour déterminer si un document numérique devrait être préservé : <https://www.canada.ca/fr/reseau-information-patrimoine/services/preservation-numerique/arbre-decisionnel.html>⁴

La sélection des documents numériques doit permettre la préservation de la version officielle des documents numériques, tout en la distinguant des doublons qui peuvent exister. Aussi, l'organisme doit préserver la version qui demeure intègre et qui n'a pas été altérée, volontairement ou par inadvertance.

1. <https://www.canada.ca/fr/reseau-information-patrimoine/services/preservation-numerique/arbre-decisionnel.html>

2. <https://www.canada.ca/fr/reseau-information-patrimoine/services/preservation-numerique/arbre-decisionnel.html>

3. <https://www.canada.ca/fr/reseau-information-patrimoine/services/preservation-numerique/arbre-decisionnel.html>

4. <https://www.canada.ca/fr/reseau-information-patrimoine/services/preservation-numerique/arbre-decisionnel.html>

2.2. Le registre des documents numériques

Lors de la planification des activités de préservation, il est important de connaître ce que son organisme détient réellement en termes de documents numériques et où sont enregistrés ces documents. La création d'un registre peut être très utile pour l'archiviste afin d'identifier quels sont les documents numériques à conserver, identifier les priorités et planifier les activités de préservation. Pour créer le registre, il est pertinent d'identifier l'activité documentée par les documents, la cote de classement, la date des documents, la taille des documents, le format, le support d'enregistrement, s'il s'agit de documents confidentiels. Ainsi, un tel registre permet d'identifier la nature des documents à préserver, leurs supports et leurs formats, car tous n'exigent pas les mêmes mesures de traitement. Le registre peut prendre différentes formes, mais il peut très bien s'agir d'un tableau de suivi constitué avec des logiciels de la suite bureautique, tel qu'illustré ci-dessous (Tableau 1) ou consigné dans les métadonnées.

Série documentaire	Cote de classement	Taille	Dates	Support	Format	Confidentiel	Remarques

Tableau 1. En-tête d'un registre de préservation

Lorsque les documents numériques à préserver ont été sélectionnés et documentés dans un registre, il est important, pour l'archiviste, de rapidement sécuriser ces documents. Cette sécurisation peut se faire de différentes façons, notamment par la migration des documents numériques dans un format plus pérenne. Par exemple, les systèmes d'archivage électronique permettent, pour les documents devant être conservés à long terme, de les convertir automatiquement dans un format PDF/A, pour les documents textuels, de façon à assurer l'accessibilité de ces documents et de limiter leur manipulation.



3. Formats de fichiers

1. Introduction

Le choix d'un format de fichier est très important dans la gestion des documents numériques et leur préservation. Les formats de fichiers sont habituellement organisés par type de contenu et développés selon leurs fonctionnalités. « Les principaux types de contenu sont les images, la vidéo, l'audio et le texte ; cependant, un nombre croissant de formats sont structurés pour répondre aux exigences des nouveaux médias, comme les formats pour les modèles 3D ou l'archivage web » (Digital Preservation Coalition, 2021, p. 212). Les formats de fichiers sont très variables en termes de complexité. Ce chapitre vise à identifier certains critères de sélection des formats de fichiers et non pas de fournir une liste exhaustive de l'ensemble des formats de fichiers actuels. Les formats n'offrent pas tous les mêmes garanties de conservation à long terme. Certains seront mieux adaptés que d'autres.

3.1. Les deux grandes catégories de formats de fichiers

Rappelons qu'il existe deux grandes catégories de formats :

- **Format propriétaire** ^{p.16} : Il s'agit d'un format qui est défini par une organisation ou un propriétaire privé qui détient les droits de propriété intellectuelle et qui en précise les usages autorisés. Par exemple, les formats de la suite bureautique de Microsoft. Les caractéristiques d'un format et ses usages sont habituellement consignés dans une documentation. Ainsi, dans le cas des formats propriétaires, il est possible qu'un format soit publié, c'est-à-dire que le propriétaire ait défini les usages autorisés et que la documentation soit accessible. Il est également possible que pour un format propriétaire, la documentation ne soit pas disponible.
- **Format ouvert** ^{p.16} : Il s'agit d'un format publié et libre de droit, sans restriction d'usage et de mise en œuvre. Par exemple, les formats définis par le World Wide Web Consortium (W3C) responsable de formats de publication Web, tels XML et HTML.

3.2. Les critères de sélection des formats de fichiers

L'archiviste doit considérer plusieurs critères lors de la sélection de formats de fichiers aux fins de gestion et de préservation, notamment :

- **Ouverture.** Il est préférable qu'un format soit publié et ouvert. Cependant, il est possible aussi qu'un format propriétaire largement diffusé et utilisé soit préférable à un format ouvert qui serait très peu utilisé. En effet, plus un format est utilisé, plus il y aura d'outils qui seront développés pour exploiter ce format. Il faut vraiment étudier les caractéristiques des formats à la lumière des types de documents numériques détenus par l'organisme.
- **Indépendance.** Pour être pérenne, un format doit être indépendant c'est-à-dire ne pas dépendre d'autres formats. Certains formats seront ouverts, mais une composante permettant l'affichage du contenu peut être propriétaire, par exemple pour les jeux de caractères. Les formats devraient être indépendants des systèmes d'exploitation ou encore d'une composante matérielle tel un support d'enregistrement.

- **Limiter le nombre de formats utilisés.** À des fins de préservation, et ce même au stade courant, il est important de limiter le nombre de formats utilisés, puisque plus il y a de formats, plus d'interventions seront nécessaires pour maintenir l'intégrité et l'accessibilité des documents numériques.
- **Vérifier l'existence de plusieurs logiciels de lecture.** Il faut se rappeler qu'il est important que le contenu des documents numériques demeure accessible. Par conséquent, l'archiviste doit s'assurer de l'existence d'appareils de lecture.
- **Volume et compression.** L'archiviste doit également vérifier le volume du format et si celui-ci compresse l'image afin de ne pas altérer l'intégrité des documents. En effet, la compression peut occasion des pertes ou des modifications de l'image. À des fins de préservation, il est important de privilégier des formats sans perte.
- **Usage prévu des formats.** L'archiviste doit vérifier quel est l'usage prévu des formats. Par exemple, certains formats d'image sont utilisés pour la diffusion puisque les fichiers sont compressés, moins volumineux et permettent un téléchargement rapidement (ex. : JPEG). Tandis que certains formats se prêtent davantage à la préservation puisqu'ils ne compressent pas l'image (ex. : TIFF).

Il s'agit là de certains critères qui peuvent guider le choix de l'archiviste pour la gestion et la préservation des documents numériques. Le choix d'un format peut également se faire selon les normes applicables et les recommandations formulées par les institutions nationales. Il importe de choisir une stratégie qui permettra la mise en place de solutions de préservation simples, mais efficaces.



4. Supports d'enregistrement

Les supports de stockage, tout comme les formats de fichiers, possèdent différentes caractéristiques qui peuvent influer sur leur capacité à assurer la préservation des documents numériques. Les besoins en termes de stockage sont très variables d'un organisme à l'autre. L'archiviste doit analyser les besoins au regard du volume des documents à préserver et du contexte organisationnel.

Les pratiques en termes d'enregistrement ont beaucoup évolué et tendent à varier selon les milieux. Lorsque l'infrastructure technologique le permet, plusieurs milieux utilisent l'espace d'enregistrement disponible sur un serveur compte tenu du volume important de documents à conserver et de la possibilité de redondance du serveur. Par ailleurs, certains milieux, selon les ressources disponibles, vont privilégier des supports autonomes à des fins de préservation. Le choix du support de stockage s'effectue habituellement dès le stade courant de la gestion des documents numériques.

L'archiviste doit considérer plusieurs critères afin de faire un choix judicieux pour le support de stockage à des fins de préservation. Voici certains critères à considérer :

- *Redondance et copies.* Capacité à faire des copies des documents numériques qui seront conservées dans différents endroits. Les documents numériques doivent pouvoir être accessibles en ligne et hors ligne, si nécessaire.
- *Stabilité.* La stabilité du support (inertie physique et chimique) et la robustesse (résistance aux agressions).
- *Normes et fiabilité.* Caractère plus ou moins répandu du support et le fait qu'il repose sur des normes reconnues.
- *Coût.* Le coût du support est à considérer.
- *Protection.* L'existence de protections contre l'effacement accidentel ou volontaire, contre les défaillances et les erreurs. (Digital Preservation Coalition, 2021)

De plus, l'archiviste devra aussi considérer les documents numériques enregistrés sur d'anciens supports de stockage, tels que les supports magnétiques et optiques. Selon le support d'enregistrement choisi, l'archiviste devra analyser, en amont, les risques de perte de documents numériques et les stratégies de préservation pour remédier à la dégradation et à l'obsolescence.



5. Stratégies de préservation

Introduction

Dans ce chapitre, nous aborderons brièvement différentes stratégies de préservation. L'archiviste peut avoir recours à l'une ou l'autre de ces stratégies avant même que les documents numériques aient atteint le stade de conservation permanente.

5.1. La migration

La migration consiste à transférer périodiquement les documents numériques conservés d'une génération technologique à une génération suivante en utilisant de nouvelles applications. Il s'agit de l'une des stratégies les plus utilisées. La migration permet ainsi de maintenir l'accessibilité aux documents numériques.



Attention

Cependant, il peut y avoir certains risques de perte d'informations puisque « toute intervention susceptible de modifier la structure et le contenu des données, peut introduire des erreurs et des pertes d'informations » (Digital Preservation Coalition, 2021, p. 182). Il faut donc prévoir une assurance qualité afin d'évaluer le taux d'erreur.

Par ailleurs, **il est important, pour l'archiviste, d'établir des priorités de migration selon les ressources disponibles.** En effet, selon les besoins, il n'est pas toujours possible de migrer l'ensemble des documents numériques. L'archiviste doit également s'interroger sur le meilleur moment pour initier la migration. Une migration prématuée entraînera des coûts tandis qu'une migration tardive peut entraîner une perte irréversible des documents numériques. Ainsi, l'archiviste peut faire des contrôles périodiques des formats et des supports de façon à identifier les documents numériques à risque ou encore de planifier les migrations.



Conseil

Parmi les bonnes pratiques, il est important de noter toutes les migrations effectuées, leurs résultats, les pertes d'informations détectées afin de conserver l'intégrité des fichiers.

5.2. L'émulation

L'émulation consiste imiter des programmes devenus désuets afin de lire des fichiers conservés dans leur format d'origine.



Définition

En effet, un émulateur « est un programme qui fonctionne sur une architecture informatique courante mais qui offre les mêmes possibilités et le même comportement qu'une architecture ancienne » (Digital Preservation Coalition, 2021, p. 183).

Cette stratégie vise ainsi à conserver les fonctionnalités, les caractéristiques d'origine ainsi que le contenu.



Cependant, un émulateur est une composante technologique. Par conséquent, il risque de devenir obsolète avec les développements technologiques.

Cette stratégie de préservation requiert souvent la collaboration d'experts et l'investissement de ressources importantes.

5.3. La conservation de la technologie

Cette stratégie de préservation consiste à conserver les ordinateurs et les logiciels qui ont permis la création des documents numériques. Cette stratégie nécessite beaucoup de maintenance afin que toutes ces composantes demeurent fonctionnelles. Cela suppose également qu'en cas de bris, des experts seront en mesure de faire les réparations nécessaires et que les pièces seront disponibles. Elle présente néanmoins l'avantage de préserver l'intégrité, les fonctionnalités et l'entièreté du document puisque l'environnement technologique est conservé.



Cependant, le fait de conserver la technologie est peu réaliste et comporte d'importantes exigences en termes de coûts et de ressources.

5.4. La copie sur papier (conversion)

Finalement, l'impression sur support analogique peut être, dans une certaine mesure, utilisée pour préserver des documents numériques puisqu'elle permet de conserver le contenu.



Cependant, cette approche occasionne une perte du caractère numérique (devient analogique) et la perte des fonctionnalités (ex.: documents multimédias ou courriels).

Ces différentes stratégies peuvent être approfondies dans le module 7C « *Préservation numérique* ».



Conclusion

En somme, la préservation numérique est complexe et requiert que certaines actions soient entreprises dès la création des documents. En effet, il faut bien identifier les enjeux et planifier la stratégie de préservation numérique. L'archiviste doit également identifier les formats et les supports appropriés pour la préservation des documents numériques ayant une valeur pour l'organisme. Il importe aussi de faire une réévaluation périodique des documents numériques afin qu'ils demeurent accessibles et d'effectuer une veille technologique.

Plusieurs autres composantes doivent être considérées dans une stratégie de préservation numérique à long terme, dont l'importance des normes (ex. : norme OAIS), les métadonnées de préservation (ex. : PREMIS) et des logiciels dédiés à la préservation dont le dépôt numérique fiable. Ces composantes seront abordées dans le module de formation 7C « *Préservation numérique* ».



Glossaire

Document numérique

Terme générique désignant tout type de document sous forme d'enregistrement électronique.

Il s'agit donc d'un document constitué sous forme numérique, soit de façon native/initiale, soit par numérisation de document non numérique.

Format numérique ouvert

Il s'agit d'un format publié et libre de droit, sans restriction d'usage et de mise en œuvre.

Format numérique propriétaire

Il s'agit d'un format qui est défini par une organisation ou un propriétaire privé qui détient les droits de propriété intellectuelle et qui en précise les usages autorisés.

Préservation numérique

La préservation numérique « réfère à la série d'activités jugées nécessaires pour garantir un accès continu aux objets numériques aussi longtemps que nécessaire » (Digital Preservation Coalition, 2021, p. 282).

Bibliographie



- [01] DIGITAL PRESERVATION COALITION. (2021). Manuel de préservation numérique. Disponible en ligne : <https://www.dpconline.org/docs/digital-preservation-handbook2/2519-handbook-2021-fr/file>
- [02] GOUVERNEMENT DU CANADA. Réseau canadien d'information sur le patrimoine. (2021). Modèle d'arbre décisionnel pour la préservation numérique servant à déterminer si les ressources numériques doivent être préservées. En ligne : <https://www.canada.ca/fr/reseau-information-patrimoine/services/preservation-numerique/arbre-decisionnel.html>
- [03] ISO 14721:2012 Systèmes de transfert des informations et données spatiales — Système ouvert d'archivage d'information (OAIS) — Modèle de référence.
- [04] McGOVERN, Nancy. Y. (2017). Lignes directrices sur l'élaboration d'un cadre stratégique pour la préservation numérique – Version 2.1. Réseau canadien d'information sur le patrimoine (RCIP). En ligne : <https://www.canada.ca/fr/reseau-information-patrimoine/services/preservation-numerique/lignes-directrices-elaboration-cadre-strategique.html>