Principales exigences pour un système de préservation numérique – version 1.1 (dernière mise à jour : décembre 2022)

Principales exigences pour un système de préservation numérique

Introduction

Ce document propose un ensemble de dix exigences fonctionnelles de haut niveau pour un système de préservation numérique, défini comme ***l'ensemble des applications et des outils utilisés pour préserver des objets numériques en assurant des processus tels que la prise en charge, le stockage, la préservation et l'accès.***

Il concerne avant tout les exigences fondamentales en matière de ***préservation numérique***, plutôt que les exigences générales auxquelles doit répondre un système de gestion de l'information (par exemple en matière de sécurité ou de sauvegarde) ou le détail des exigences portant sur une instance spécifique d'un système de préservation (par exemple le nombre de copies à conserver). Ce document ne vise pas à établir un ensemble d'exigences de base valable pour tous les systèmes de préservation numérique ou pour tous les contextes, mais plutôt à identifier les éléments dont nous considérons qu'ils distinguent habituellement un système de préservation numérique des autres applications et outils utilisés pour gérer des objets numériques.

Ce document fait partie de la boîte à outils pour les achats (*Procurement Toolkit*) de la DPC, qui fournit des conseils sur les différentes approches possibles pour acquérir des systèmes et services tiers. Il est recommandé d'utiliser ce document conjointement avec les autres éléments de la boîte à outils, notamment les documents sur les leçons retenues (*Lessons Learned*) et les exigences courantes (*Common Requirements*).

Objectifs

Ce document vise à simplifier et à améliorer le processus d'acquisition d'un système de préservation numérique, aussi bien pour l'acheteur que pour les fournisseurs tiers répondant à des appels d'offres[1]. Les organismes faisant l'acquisition de systèmes de préservation numérique peuvent prendre ces exigences comme point de départ, puis chercher à identifier des fonctionnalités supplémentaires qui seraient importantes pour eux (par exemple en raison des spécificités de leur contexte organisationnel, des types d'objets à prendre en charge, de l'environnement technique déjà existant ou d'une procédure de certification).

Ce document peut aussi servir d'outil pédagogique, par exemple pour aider un professionnel de la préservation à expliquer pourquoi un système informatique ordinaire ne répond pas forcément aux besoins de la préservation à long terme.

1 Les difficultés liées à ce processus (pour tous ses acteurs) avaient au préalable été identifiées lors d'un atelier auquel participaient des membres et partenaires de la DPC. https://www.dpconline.org/events/past-events/guide-to-dp-procurement-event

Mode d'emploi

Les termes « doit », « devrait » et « peut » sont utilisés dans la formulation des exigences de ce document pour indiquer l'importance de chacune. Les professionnels de la préservation sont encouragés à modifier leur formulation et leur importance ainsi qu'à ajouter de nouvelles exigences en fonction de leurs propres besoins. Ces exigences supplémentaires peuvent concerner l'intégration à d'autres systèmes, les flux de travail dédiés à des types d'objets spécifiques ou les fonctions supplémentaires exigées du fournisseur (telles que la migration de données). Un autre document décrivant comment articuler ces exigences avec la Grille d'évaluation rapide de la DPC est disponible.

Le Manuel de préservation numérique fournit un glossaire des termes utilisés dans le domaine de la préservation numérique.

Remerciements et appel à commentaires

Ces exigences ont été élaborées en collaboration avec la *UK Nuclear Decommissioning Authority*. Nous remercions également les membres et partenaires de la DPC dont les remarques et commentaires ont enrichi ce travail. Ces exigences pour la préservation numérique continueront à évoluer en fonction des commentaires qu'elles susciteront, tout en demeurant étroitement associées à la boîte à outils pour les achats.

Informations sur la traduction française

La traduction française de ces exigences a été réalisée dans le cadre des travaux de la Cellule nationale de veille sur les formats entre mars et avril 2023. Lancée en 2019, la Cellule nationale de veille sur les formats (CNVF), sous l'égide de l'association française Aristote et de son groupe de travail sur la Pérennisation de l'information numérique (PIN), regroupe à ce jour une douzaine de partenaires. Ses principaux objectifs sont la mutualisation des activités de veille sur les formats, la sensibilisation des professionnels sur le sujet, la contribution ou l'influence sur les outils associés. Elle ambitionne d'être un interlocuteur francophone reconnu dans les travaux internationaux sur ces sujets.

Les exigences ont été traduites en français par André Falut (Archives nationales), Marion Humbert (Conseil départemental de la Moselle), Thomas Ledoux (Bibliothèque nationale de France) et Édouard Vasseur (École nationale des chartes). Une première traduction automatique a été réalisée avec le logiciel DeepL puis a été revue et corrigée par le groupe de travail.

10 exigences pour un système de préservation numérique

1. Le système doit enregistrer de façon précise et gérer l'intégrité et l'authenticité des objets numériques et des métadonnées qu'il conserve. Il doit utiliser des empreintes pour vérifier que les objets numériques n'ont pas été modifiés accidentellement ou délibérément au fil du temps, et des métadonnées pour enregistrer toutes les actions effectuées sur ces objets.

- Finalité : Vérifier de façon transparente et contrôlable que les objets numériques restent inchangés au fil du temps et garantir que les changements prévus soient entièrement documentés.

2. Le système doit être doté d'un modèle de données exhaustif permettant de récupérer, lors de la prise en charge, la structure complexe des objets numériques et de la représenter de façon exacte tout au long de leur durée de conservation et de préservation.

- Finalité : Permettre au système de conserver tous les objets numériques sélectionnés et de maintenir les relations existantes entre les différents éléments des objets numériques complexes pour permettre leur préservation sans pertes.

3. Le système doit fournir une stratégie claire d'export vers d'autres systèmes, indépendante des fournisseurs, de façon à garantir que la migration inévitable vers un nouveau système de préservation numérique à l'avenir sera possible. Il devrait également minimiser les efforts et les risques qu'implique une telle migration.

- Finalité : Garantir que les objets numériques et toutes leurs métadonnées associées puissent être extraits du système de préservation numérique au besoin.

4. Le système doit permettre la prise en charge d'objets numériques et de métadonnées authentiques.

- Finalité : Garantir que les objets et métadonnées puissent être pris en charge et décrits sans pertes ni altérations.

5. Le système doit avoir la possibilité d'évaluer les caractéristiques des objets numériques pris en charge et d'enrichir les métadonnées associées en conséquence.

- Finalité : La gestion des objets numériques en vue de leur pérennisation a plus de chances d'être possible et réussie si le système dispose d'informations sur la nature de ces objets.

6. Le système doit prendre en charge la réplication et la gestion du stockage. Le système doit avoir la capacité de stocker des copies multiples des objets numériques pris en charge dans différents systèmes de stockage et à des emplacements géographiques différents.

- Finalité : Une gestion efficace du stockage limite les risques d'altération ou de perte des objets.

7. Le système devrait prendre en charge la planification de la préservation, y compris l'évaluation des risques, la conception et l'expérimentation de plans pour la préservation des objets numériques.

- Finalité : Créer des plans fiables de protection des objets numériques sur le long terme.

8. Le système devrait prendre en charge des actions de préservation correspondant à des plans de préservation destinés à réduire les risques identifiés pour la préservation.

- Finalité : Permettre les actions nécessaires pour protéger les objets numériques et garantir la possibilité d'y accéder à l'avenir.

9. Le système doit prendre en charge la gestion d'objets numériques et de métadonnées au fil du temps.

- Finalité : La gestion et la préservation d'objets numériques sur de longues périodes constitue un processus actif.

10. Le système doit permettre de rechercher et trouver de façon contrôlée les objets numériques et leurs métadonnées et d'y accéder.

- Finalité : Les objets numériques sont préservés afin de pouvoir être recherchés, trouvés et utilisés par d'autres usagers.

Détail des exigences

1. Le système doit enregistrer de façon précise et gérer l'intégrité et l'authenticité des objets numériques et des métadonnées qu'il conserve. Il doit utiliser des empreintes pour vérifier que les objets numériques n'ont pas été modifiés accidentellement ou délibérément au fil du temps, et des métadonnées pour enregistrer toutes les actions effectuées sur ces objets.

1.1 Le système doit enregistrer une empreinte pour chaque fichier.

1.2 Le système doit être capable de comparer ces empreintes à celles fournies avec les objets.

1.3 Le système doit permettre la réalisation de contrôles d'intégrité réguliers, en signalant tout fichier endommagé ou manquant.

1.4 Le système doit mettre en œuvre les étapes nécessaires pour réparer les fichiers endommagés ou les remplacer par des copies répliquées et fournir un rapport des actions effectuées.

1.5 Le système doit être capable de générer un journal de contrôle et enregistrer des métadonnées associées à chaque événement survenu (telles que celles exigées par PREMIS[2]) décrivant toutes les actions effectuées sur les objets numériques.

2. Le système doit être doté d'un modèle de données exhaustif permettant de récupérer, lors de la prise en charge, la structure complexe des objets numériques et de la représenter de façon exacte tout au long de leur durée de conservation et de préservation.

2.1 Le modèle de données doit être en mesure de refléter et de représenter les objets numériques composés de plusieurs éléments hiérarchisés, tels que fichiers, brouillons, versions publiées ou copies créées ultérieurement à des fins de préservation ou d'accès.

2.2 Les objets numériques devraient accepter des identifiants uniques et pérennes.

2 https://www.loc.gov/standards/premis/v3/index.html

3. Le système doit fournir une stratégie claire d'export vers d'autres systèmes, indépendante des fournisseurs, de façon à garantir que la migration inévitable vers un nouveau système de préservation numérique à l'avenir sera possible. Il devrait également minimiser les efforts et les risques qu'implique une telle migration.

3.1 La structure des objets numériques stockés peut être compréhensible/interprétable indépendamment du système de préservation lui-même.

3.2 Le système doit avoir la capacité d'exporter les objets numériques en masse avec toutes les métadonnées qui leur sont associées dans un format/une structure pouvant être pris en charge par un autre système.

3.3 Les codes sources du système de préservation peuvent être déposés auprès d'un tiers de confiance, ce qui fournit une forme d'assurance en cas de défaut du fournisseur.

4. Le système doit permettre la prise en charge d'objets numériques et de métadonnées authentiques.

4.1 Le système doit permettre la prise en charge en masse d'objets numériques et des métadonnées qui leur sont associées.

4.2 Les objets à prendre en charge devraient faire l'objet d'une analyse antivirus, associée à des moyens de mise en quarantaine.

4.3 Le système doit être capable de conserver non seulement le train de bits d'origine, mais aussi les autres caractéristiques nécessaires à la préservation et à l'accès aux objets, telles que la structure des dossiers d'origine, les informations sur les fichiers comme les horodatages, la documentation les accompagnant et les métadonnées qui leur sont associées.

4.4 Le système doit être capable de fournir des rapports de gestion sur le succès ou l'échec des actions de prise en charge et devrait également être en mesure de signaler tout problème potentiel des objets pris en charge (par exemple les formats de fichiers obsolètes ou non reconnus, les métadonnées manquantes, les fichiers protégés par des mots de passe ou par un chiffrement, etc.).

5. Le système doit avoir la possibilité d'évaluer les caractéristiques des objets numériques pris en charge et d'enrichir les métadonnées associées en conséquence.

5.1 Le système doit identifier les formats de fichiers jusqu'au niveau de la version exacte du format et renvoyer aux registres appropriés, tels que PRONOM et/ou Wikidata, pour des informations plus détaillées.

5.2 Le système doit extraire les caractéristiques techniques (telles que taille, dimensions d'image, codec vidéo, durée de piste audio, application qui a servi à créer le fichier).

5.3 Le système devrait identifier les objets ne pouvant pas être représentés, tels que les objets corrompus, mal structurés ou chiffrés.

5.4 Le système devrait valider les formats de fichiers d'après les spécifications de ces formats ou des profils adaptés.

5.5 Le système devrait identifier les dépendances externes, c'est à dire les situations où des contenus (ou logiciels) ne faisant pas partie de l'objet numérique sont essentiels à sa représentation ou à son utilisation (par exemple, des polices de caractères non embarquées, des médias non embarqués tels que des vidéos YouTube, ou des bibliothèques logicielles).

6. Le système doit prendre en charge la réplication et la gestion du stockage. Le système doit avoir la capacité de stocker des copies multiples des objets numériques pris en charge dans différents systèmes de stockage et à des emplacements géographiques différents.

6.1 Le système doit gérer automatiquement la réplication des objets numériques à plusieurs emplacements de stockage (potentiellement à différents emplacements géographiques).

6.2 Le système devrait régulièrement effectuer des sauvegardes système.

6.3 Le système devrait être capable de tester régulièrement ses capacités de sauvegarde et de restauration et de fournir des rapports sur ces tests.

6.4 Le système devrait créer et conserver des rapports de gestion sur ses actions de réplication, de gestion du stockage, de sauvegarde et de restauration.

7. Le système devrait prendre en charge la planification de la préservation, y compris l'évaluation des risques, la conception et l'expérimentation de plans pour la préservation des objets numériques.

7.1 Le système devrait prendre en charge l'identification, la gestion et l'analyse des risques associés à la préservation (par exemple dans les cas où le format de fichier d'un objet numérique n'est plus pris en charge).

7.2 Le système devrait permettre la conception, le développement et la gestion de plans de réduction des risques associés à la préservation identifiés.

7.3 Le système devrait fournir des rapports permettant une gestion efficace des objets numériques. Cela doit comprendre une large variété de rapports fonctionnels, opérationnels et statistiques ainsi que d'analyses.

8. Le système devrait prendre en charge des actions de préservation correspondant à des plans de préservation destinés à réduire les risques identifiés pour la préservation.

8.1 Le système devrait permettre la migration de fichiers d'un format vers un autre.

8.2 Le système devrait permettre la représentation d'objets numériques via une émulation et/ou d'autres outils spécialisés.

8.3 Le système devrait assurer un contrôle de la qualité des résultats de toute action de préservation.

8.4 Le système doit enrichir les métadonnées associées aux objets à la suite des actions de préservation (et de leurs résultats).

9. Le système doit prendre en charge la gestion d'objets numériques et de métadonnées au fil du temps.

9.1 Le système doit fournir des moyens de contrôle permettant de minimiser les risques de suppression ou d'altération accidentelle ou délibérée des objets.

9.2 Le système doit être capable de prendre en charge et de gérer toutes les métadonnées nécessaires, y compris les métadonnées adaptées à des types de contenus spécifiques (données géographiques ou audiovisuelles, par exemple).

9.3 Le système doit faciliter l'enrichissement des métadonnées et de la documentation tout au long de la vie des objets numériques.

9.4 Le système doit permettre la mise en œuvre contrôlée du sort final des objets (élimination ou transfert) et permettre la conservation des métadonnées alors même que les objets numériques auxquels elles étaient associées ont été retirés du système.

9.5 Le système doit restreindre toutes les actions de gestion des objets numériques en fonction de rôles utilisateurs paramétrables.

10. Le système doit permettre de rechercher et trouver de façon contrôlée les objets numériques et leurs métadonnées et d'y accéder.

10.1 Le système doit garantir que tous les objets numériques et les métadonnées qui leur sont associées ne soient requêtables et accessibles que par les utilisateurs autorisés.

10.2 Le système peut avertir les utilisateurs s'ils tentent d'accéder à des contenus pouvant présenter des problèmes (par exemple des objets numériques conservés dans un format de fichier non reconnu, dont les métadonnées sont incomplètes, nécessitant des autorisations supplémentaires, etc.).

10.3 Le système peut fournir des fonctionnalités d'émulation, des visionneuses spécialisées, une option de migration à la demande ou des applications auxiliaires pour permettre l'utilisation d'objets numériques dont les formats sont obsolètes ou non reconnus.

10.4 Le système doit fournir aux utilisateurs et/ou aux API dotées des capacités nécessaires une interface d'accès permettant son intégration à d'autres systèmes d'accès ou outils de recherche.

10.5 Lors de l'accès à des objets numériques, le système devrait afficher les informations dont il dispose à leur sujet, notamment leurs métadonnées de préservation et empreintes, pour faire la preuve de la chaîne d'authenticité.

10.6 Les métadonnées et/ou objets numériques devraient pouvoir être moissonnés.