

La conservation numérique et les archives en accès ouvert. Un accès permanent aux fonds numériques en accès ouvert.

Les dépôts en accès ouvert assurent une très large diffusion des productions scientifiques et académiques. Les chercheurs et les enseignants publient librement en ligne leurs articles pour faire connaître leurs activités et partager les résultats de leurs recherches avec d'autres chercheurs. En particulier les universités, les centres de recherches, les bibliothèques et, pour des ensembles limités de leurs collections, les musées, les archives administratives et autres organismes culturels promeuvent l'accès ouvert.

Dans l'avenir, une part importante de la mémoire des institutions académiques, universitaires et culturelles sera constituée de patrimoines originellement numériques, conservés dans des dépôts en accès ouvert. Leurs collections numériques auront une importance croissante dans la constitution de l'héritage scientifique et informationnel des générations à venir. Afin de s'assurer que ces objets survivront et continueront d'être cités, les communautés universitaires et académiques devraient s'engager dans la conservation à long terme de leurs dépôts.

Contexte et objectifs

D'après le JISC la conservation numérique est : « les ensembles d'interventions et d'actions nécessaires pour assurer un accès continu et fiable aux objets numériques authentiques aussi longtemps qu'ils seront considérés comme ayant de la valeur ». Trois concepts clés peuvent être soulignés dans le contexte des archives en accès ouvert :

- Les dépôts en accès ouvert emmagasinent des objets numériques qui sont couramment des productions « originellement numériques » sans analogue équivalent et originel.
- La disponibilité et la fiabilité des fonds en accès ouvert sont pertinentes pour leurs auteurs et pour les communautés scientifiques et académiques.
- Les patrimoines sont en accès ouvert et permanent : ceci concerne l'utilisation en continue d'un patrimoine numérique, la possibilité de citer, de découvrir, de transmettre et d'utiliser cette ressource après sa création et son dépôt dans un fonds d'archives.

L'accessibilité doit être assurée aussi longtemps qu'il est nécessaire en prévenant les défauts de supports ou de locaux et en maîtrisant les effets des changements technologiques. Les procédures de préservation doivent s'appuyer sur des stratégies, des organisations et des politiques parfaitement bien définies.

Modèles

Le modèle de référence OAIS définit un système d'information d'archives ouvertes (OAIS) comme « une institution d'archive, se composant d'une organisation de personnes et de méthodes, qui assume la responsabilité de conserver l'information et de la rendre disponible pour une communauté visée ». Ce modèle fournit toutes les fonctions et toutes les spécifications propres à un dépôt et « décrit les responsabilités obligatoires qu'une organisation doit remplir afin de gérer une archive OAIS ».

L'application des principes d' OAIS, et en particulier l'implémentation d'archives en accès ouvert selon les modèles fonctionnels et d'information d' OAIS, est un préalable majeur à la mise en œuvre de dépôts habilités et à la garantie d'une préservation à long terme de leurs patrimoines numériques.

Propriétés

La conservation numérique doit donner confiance aux auteurs et aux utilisateurs en ce qui concerne les dépôts numériques en accès ouvert. Un dépôt certifié doit assurer au minimum les propriétés suivantes :

- Authenticité. C'est la certitude qu'une source numérique a été créée par la personne qui le proclame. L'authenticité fournit la certitude que l'auteur d'une source numérique ne peut pas nier en être le créateur. Les signatures numériques et le tatouage électronique sont des techniques qui garantissent l'authenticité des objets numériques.

Sources et informations complémentaires

CrossRef est une application dédiée à l'identification rémanente des contenus académiques et au nommage cross éditorial reliant le plein texte et les ressources concernées en utilisant le DOI
www.crossref.org/02publishers/16openurl.html

Citebase Search est un index de citation semi autonome pour la recherche ouverte de textes en ligne. Il touche les archives conformes au pré et post format OAI-PMH.
www.citebase.org

OAI-ORE : Open Archives Initiative Object Reuse and Exchange
www.openarchives.org/ore

L'Open Archival Information System (OAIS) Reference Model (Standard ISO 14721)
ssdoo.gsfc.nasa.gov/nost/isoas
 Digital Preservation Europe (DPE) collection de ressources en ligne pertinentes et importantes pour les archives et les travaux sur le record management.
www.digitalpreservationeurope.eu/resources

Digital Preservation Coalition (DPC) a démarré en 2001. Elle encourage l'action collective afin de relever les défis pour sécuriser rapidement la préservation des ressources numériques dans le Royaume Uni et afin d'œuvrer à l'échelle internationale pour sécuriser notre mémoire numérique globale et notre base de connaissance.
www.dpconline.org

Le programme JISC, pour les dépôts et la préservation, vise à développer l'information de contexte facilitant les dépôts numériques et la préservation, y compris les solutions de recherches croisées entre dépôts. Fonds et bourses à destination des organismes pour le développement de masses critiques de contenus, de solutions de préservation et de conseils pour le développement des dépôts légaux.
www.jisc.ac.uk/whatwedo/programmes/programme_rep_pres.aspx

International Research on Permanent Authentic Records in Electronic System (INTERPARES) : Deux projets (inter pares 1, 1999-2001 et inter pares 2, 2002-2006) qui ont permis le développement d'une connaissance théorique et méthodologique essentielle à la préservation de longue durée des documents authentiques créés ou maintenus sous la forme numérique.
www.interpares.org

PREMIS (Métadonnées de préservation : stratégies d'implémentation) est une initiative collective de OCLC et RLG visant à développer un ensemble de base, de métadonnées de préservation, appuyé sur un dictionnaire de données, avec une large applicabilité à travers la communauté de la préservation numérique ; ainsi que pour l'identification et l'évaluation de stratégies alternatives pour l'encodage, le stockage et la gestion des métadonnées de préservation dans les systèmes de préservation numérique.
www.oclc.org/research/projects/pmwg/resources.htm

- L'intégrité. C'est la capacité d'assurer l'exactitude et la complétude des données en évitant les altérations accidentelles ou pernicieuses (ex : corruption de données). La constitution et la sauvegarde d'une somme de contrôle de bits/ octets, tel le mode MD5, est une technique de base pour détecter si une quelconque modification a touché un objet numérique après qu'il ait été déposé dans une archive (au sens OAIS).
- Fiabilité et disponibilité. La fiabilité se réfère à la capacité des composants matériels et logiciels de s'exécuter selon les spécifications, sans erreur ni bogue (totalement en théorie, mais avec seulement un très fort pourcentage en pratique). La disponibilité est le pourcentage de temps pendant lequel un système d'exploitation, une application ou un composant est régulièrement en fonctionnement eu égard au temps total qu'il est supposé fonctionner. Sauvegardes, logiciels antivirus, pare-feu, améliorations de systèmes d'exploitation, nouvelles versions de logiciels d'applications, composants matériels redondants et tolérants aux pannes sont quelques unes des techniques les plus communément utilisées pour garantir des pourcentages élevés de fiabilité et de disponibilité.
- Capacité à la réutilisation. C'est la possibilité d'accéder à une ressource numérique aussi longtemps que le dépôt/ l'institution s'engage à la supporter. Les sources numériques scientifiques ou académiques avec une valeur rémanente doivent pouvoir être extraites et réutilisées, même sur une très longue période (ex par des identifiants rémanents et par la veille technologique concernant les supports et les formats).

La conservation numérique inclut toute l'infrastructure des ensembles de données d'un dépôt numérique en accès ouvert : non seulement les objets numériques mais aussi leurs métadonnées, leurs identifiants et les interpréteurs de métadata.

Stratégies et politiques.

- Un système de citations fiable basé sur un identifiant unique et persistant : confiance dans l'intégrité et l'authenticité des ressources numériques identifiées, garanties par des stratégies de conservation à long terme, pour éviter les liens discontinus (ex le trop connu « 404 not found » message d'erreur) par la mise en œuvre de solutions techniques et organisationnelles qui vont au-delà des limites des citations basées sur des URLs.
- Les formulations de politique organisationnelle doivent non seulement déterminer les ressources numériques à préserver et assurer les supports financiers et organisationnels pour ce faire, mais aussi décrire qui est en charge de la préservation numérique. Les rôles et les responsabilités doivent être définis avec précision, les compétences doivent être soigneusement entretenues et les formations doivent être sans cesse encouragées.
- La préservation numérique de la connaissance pluridisciplinaire doit être basée sur l'étroite coopération des experts et des spécialistes appartenant aux différents domaines.
- Il faut accroître la sensibilité des auteurs à bien suivre les stratégies et les politiques qui garantissent la continuité d'accès à leurs sources numériques.
- Les métadonnées permettent un accès continu aux ressources numériques ainsi que leur disponibilité : la planification de la conservation numérique doit aussi inclure les stratégies et les procédures pour la conservation de ces métadonnées.
- Des efforts constants de contrôles et d'évaluation sont nécessaires pour prendre en compte les évolutions des composantes organisationnelles, politiques, juridiques et techniques.