

Pour en savoir plus sur la préservation et l'accès pérennes des documents numériques

Boîte à outils

La bibliothèque de l'université de Cornell a développé un didacticiel sur la préservation des collections numériques. Glossaires, méthodologie, contexte, modèle OAIS, ce didacticiel offre une approche complète des problématiques à l'œuvre.

Disponible en français

<http://www.library.cornell.edu/iris/tutorial/dpm-french/index.html>

Bibliographie

The KB e-Dépôt : building and managing a safe place for E-Journals

Marcel Ras, congrès LIBER, Istanbul, 2008

<http://www.ku.edu.tr/ku/images/LIBER/ras.ppt>



The digital challenge. News perspectives for the Metamorfoze programme

Dennis Schouten, Liber Quaterly, vol.17, 2007, n° 3 et 4

Voir aussi la présentation donnée lors du congrès LIBER, en 2007, à Varsovie

<http://www.enssib.fr/bibliotheque-numerique/notice-1292>

Accès via la bibliothèque numérique de l'ENSSIB

<http://www.enssib.fr/bibliotheque-numerique/document-1292>



How much does it cost? The LIFE Project - Costing models for digital curation and preservation

Richard Davies, Paul Ayris, Rory McLeod, Helen Shenton, Paul Wheatley, in Liber Quaterly, 2007, vol.17, n° 3 et 4



L'archivage pérenne des documents numériques
Michel Auffret, *Arabesques*, n° 39, 2005



Guide des bonnes pratiques, version 1.3, groupe de travail Minerva n° 6, 3 mars 2004

 RECOMMANDATION, La conservation électronique des documents, Le Forum des droits sur l'internet, 1^{er} décembre 2005

 <http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/conservation.pdf>

À la BNF

La préservation des documents numériques

La préservation des documents numériques à long terme est un enjeu fondamental pour les bibliothèques et particulièrement pour les bibliothèques nationales.

Dans beaucoup de cas, la conservation des documents patrimoniaux passe par le numérique : ouvrages sur papier acide numérisés, mais aussi documents nés numériques – audiovisuels, web...

La Bibliothèque nationale de France (BNF) tire trois conclusions de cette situation :

- 1 toute numérisation doit être considérée comme une numérisation de sauvegarde ;
- 2 il est nécessaire de construire un magasin numérique de préservation à long terme ;
- 3 il est important d'offrir un entrepôt fiable pour les établissements qui ont des objets numériques à conserver et n'ont pas les moyens de bâtir leur propre infrastructure.

Toute numérisation est une numérisation de sauvegarde

La politique de numérisation s'inscrit désormais dans une logique de conservation. D'une part, la reproduction de sauvegarde des documents bascule progressivement de l'argentique au numérique, d'autre part, l'étude des données et des formats numériques montre qu'ils sont déjà soumis à de nombreux facteurs d'obsolescence matérielle et logicielle. Le défi à relever pour assurer la pérennité des collections est donc double : maîtriser la forme que prennent les données et garantir la fiabilité des formats.

Selon les spécificités des modes d'entrée de documents numériques, la BNF recense huit filières :

- la numérisation de conservation des imprimés, images fixes et manuscrits ;
- la numérisation de conservation de l'audiovisuel ;
- la numérisation de reproduction ;
- le dépôt légal automatique d'Internet ;
- le dépôt légal négocié ;
- les productions administratives et techniques ;
- le dépôt ou tiers archivage ;
- les acquisitions et dons.

Chacune de ces filières doit faire l'objet de négociations particulières entre les acteurs de la préservation, pour que soient prises en compte les exigences en matière de versement, de stockage et de diffusion.

La BNF a défini une typologie des formats qui conditionne la qualité de préservation qu'elle s'engage à mettre en œuvre (veille, migration, émulation...). Cette typologie fait apparaître quatre degrés de maîtrise croissante, depuis les formats seulement stockés, pour lesquels on n'assure que la conservation du train de bits, jusqu'aux formats identifiés, connus et maîtrisés.

Construire un magasin numérique de préservation à long terme

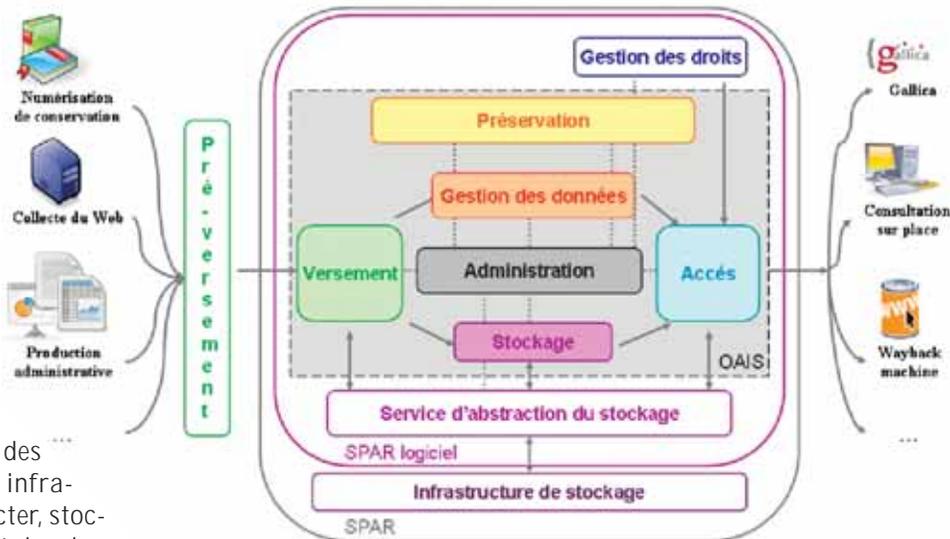
Afin de mettre en œuvre l'archivage pérenne des documents numériques, la BNF crée une infrastructure sécurisée qui sera capable de collecter, stocker, préserver et diffuser en grande quantité des documents numériques de nature diverse, numérisés ou nés numériques. Cette infrastructure, véritable magasin numérique, est le Système de préservation et d'archivage réparti (SPAR).

L'architecture de SPAR est conforme au modèle OAIS (Open Archival Information System – ISO 14721), dont les entités définissent le périmètre de la préservation du document numérique. Elles ne se résument pas à la seule entité Stockage : s'y ajoutent l'entité Versement, qui gère les entrées, l'entité Gestion des données, qui indexe les métadonnées des documents et permet leur exploitation, et l'entité Accès, qui gère les communications. S'ajoutent des entités qui ont un rôle transversal comme l'Administration, pour la gestion courante de l'ensemble du système, et la Planification de la préservation, qui définit les stratégies de préservation (émulation, migration) puis planifie leur mise en œuvre.

Lors de la conception du système SPAR, la BNF a choisi de faire correspondre **un module du système à chaque entité OAIS**. Ce choix d'implémentation assure la souplesse et l'évolutivité du système : chaque module étant indépendant, il est possible de remplacer un module sans nécessairement remplacer l'ensemble. Pour le développement du système, la BNF a fait le choix de l'**open source**, plus approprié à un système conçu pour le très long terme qu'un système propriétaire.

Les fonctionnalités définies par la BNF l'ont amenée à développer des modules supplémentaires par rapport aux entités de la norme OAIS : en amont du versement, le module Préversement se charge de rendre les documents livrés conformes à ce que l'on souhaite faire entrer dans SPAR (format du document, format de métadonnées) ; au moment de l'accès, le module Gestion des droits filtre les documents numériques qui sortent de SPAR selon les droits qui y sont attachés et le contexte dans lequel on les demande. Enfin, le Système d'abstraction de stockage permet de rendre le système indépendant de l'infrastructure matérielle sur laquelle il repose. Au-delà, le système SPAR s'intègre à un ensemble d'applications plus vaste, le système d'information numérique de la BNF ; par exemple, Gallica prend en charge la remise des objets numériques aux utilisateurs finaux.

Ce système repose sur une infrastructure matérielle, qui combine un stockage sur bandes, fiable et peu coûteux mais lent, et un stockage sur disques, plus rapide, destiné aux flux de données, lors du versement d'un document ou de



sa communication. Dans un souci de gestion des risques, les documents archivés dans SPAR, ainsi que le système lui-même, sont enregistrés sur deux sites, ce qui permet de prévenir une éventuelle perte de données ou la défaillance de l'un des sites.

Un entrepôt de confiance pour les autres établissements

Établissement d'envergure nationale, la BNF coopère avec les grandes institutions françaises impliquées dans la préservation du numérique. Ainsi, les équipes du CINES (Centre informatique national de l'enseignement supérieur) et de la BNF travaillent déjà ensemble pour partager leur savoir-faire et leur expérience. On peut aussi penser qu'à terme, la BNF et le CINES auront des systèmes interopérables qui deviendront « miroirs » l'un de l'autre pour accroître la sécurité et la disponibilité des données. Dans cette optique de mutualisation, la BNF mène actuellement une étude, avec la Caisse des dépôts et consignations, afin de proposer un service d'archivage à des institutions tierces n'ayant pas les moyens humains ou financiers de relever seules le défi de la préservation de leur patrimoine numérique.

L. Fauduet et S. Peyrard

louise.fauduet@bnf.fr
sebastien.peyrard@bnf.fr

BNF – Direction des services et des réseaux
Christian Lupovici, directeur du département
Information bibliographique et numérique
christian.lupovici@bnf.fr
☎ 01 53 79 59 00 📠 81 50
Louise Fauduet - département de la conservation
Sébastien Peyrard - département de l'information
bibliographique et numérique
📮 Quai François-Mauriac 75706 PARIS CEDEX 13