

L'ARCHIVAGE PÉRENNE

Marion Massol

A.D.B.S. | « I2D - Information, données & documents »

2015/1 Volume 52 | pages 62 à 63

ISSN 2428-2111

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-1-page-62.htm>

Distribution électronique Cairn.info pour A.D.B.S..

© A.D.B.S.. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

3 ENJEUX ET PERSPECTIVES

L'archivage pérenne



Ingénieur de recherche, **Marion MASSOL** travaille sur les problématiques liées à l'archivage pérenne du patrimoine numérique. Elle dirige le département Archivage et diffusion du Centre informatique national de l'enseignement supérieur (Cines), qui archive divers types de fonds : thèses de doctorat, littérature grise, ouvrages numérisés, données scientifiques, etc.

massol@cines.fr

[enjeu] Conserver les documents pour pouvoir les réutiliser à tout moment dans le futur, tel est le défi relevé par le Cines pour les documents issus de l'enseignement supérieur et de la recherche.

De nos jours, l'information sous forme numérique est omniprésente, en grande quantité, sous des multiples formes et constitue notre patrimoine. À ce titre, la question de sa conservation se pose, mais avec une acuité accrue compte tenu des masses considérables de données concernées.

Finalités

L'archivage pérenne d'un document numérique ambitionne de le conserver, le rendre accessible et en préserver l'intelligibilité. Ce triple objectif vise à permettre sa réexploitation au fil des décennies et/ou des générations. Par conséquent, il ne s'agit ni d'une sauvegarde qui se borne à assurer une préservation du train de bits

d'un fichier, ni d'un service HSM (*hierarchical storage manager*) qui migre les fichiers sur bande pour faire de la place sur disque ; et ce n'est pas non plus l'ultime étape du stockage des données avant l'oubli ou la perte définitive.

Enjeux

Les enjeux de l'archivage pérenne portent tout à la fois sur :

- l'identification unique et pérenne de l'information ;
- l'existence de métadonnées ;
- la qualité du support de stockage de l'information (dégradation ou vieillissement du média : effets chimiques, démagnétisation, etc.) ;
- l'obsolescence matérielle (absence d'outil pour lire le média) ;
- l'impasse de la disparition des formats de fichiers et des systèmes d'exploitation permettant l'exécution des logiciels associés à leur lecture/exploitation ;
- ainsi que sur la préservation de l'authenticité, de l'intégrité et de la confidentialité de l'information.

Le rôle du Cines

Le Centre informatique national de l'enseignement supérieur (Cines)¹ propose des solutions d'archivage de haut niveau à l'ensemble de la communauté scientifique. Elles complètent et fonctionnent en synergie avec le calcul numérique intensif qui constitue la seconde mission statutaire du centre depuis 1980. En effet, le Cines héberge quelques-uns des plus puissants supercalculateurs mondiaux (financés par Genci²) et les données archivées alimentent les calculs qui, eux-mêmes, génèrent de grandes quantités de données dont une partie doit être conservée.

Périmètre

Tout type de donnée numérique - qu'elle soit scientifique (observations ou calculs), patrimoniale (revues, manuscrits, littérature grise, etc.) ou administrative - peut bénéficier d'un archivage pérenne sur les plateformes mutualisées, sécurisées, performantes et personnalisables du Cines. Ces documents doivent toutefois :

- être en lien avec le monde de l'enseignement supérieur et de la recherche ;
- ne plus avoir d'utilité courante ;
- bénéficier d'un plan de classement et être intelligemment sélectionnés : pertinence de l'archivage du document ; élimination de données redondantes ; choix de documents « bruts ». En effet, il n'est pas nécessaire d'archiver des documents qui peuvent être reconstitués à partir de données déjà archivées (exemple : les OCR sans valeur ajoutée humaine). Le segment BSN³ travaille à produire des recommandations sur cette question.
- être documentés *via* un protocole de versement avec des métadonnées explicites ;
- pouvoir être convertis dans un des formats d'archivage acceptés par le Cines afin de faciliter les futures conversions de formats.

Et la littérature grise ?

Parmi l'ensemble des documents et données archivés au Cines depuis 2006, une large

partie relève de la littérature grise. Les fonds conservés proviennent de sources diverses. Tout d'abord, dans le cadre du dispositif national Star, toutes les thèses électroniques de doctorat soutenues en France sont transmises au Cines pour archivage pérenne. Au-delà de ce fonds réglementaire, le centre conserve les documents déposés dans l'archive ouverte HAL ainsi que les collections numériques de bibliothèques universitaires (Cujas, BUPMC, BIU Santé, Sainte-Geneviève, etc.). Les typologies documentaires sont variées : thèses, mémoires, rapports, actes de colloques, articles, photos, ouvrages patrimoniaux numérisés, revues scientifiques, etc. En complément de ces fonds, le Cines préserve aussi les données scientifiques⁴ d'institutions de recherche (Inserm, Irstea, Cerfacs, etc.).

D'autres programmes

Cette problématique est également traitée *via* les programmes Spar⁵ (système de préservation numérique de la BnF), Vitam⁶ (interministériel), Sipad-NG⁷ (du Cnes qui pérennise des données depuis plusieurs décennies), et les expériences de quelques collectivités territoriales (Marine, e-Megalis, e-bourgogne, ArchiLand, etc.). ■

1. Un établissement public national du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche www.cines.fr
2. Grand équipement national de calcul intensif, www.genci.fr/fr
3. www.bibliothequescientifiqnumerique.fr
4. Le nœud français de la grille d'archivage européenne Eudat (<https://www.cines.fr/archivage/le-projet-europeen-eudat>) ; le projet national Isaac (<https://www.cines.fr/archivage/archivage-intermediaire-des-donnees-scientifiques/le-projet-isaac>)
5. www.bnf.fr/fr/professionnels/spar_systeme_preservation_numerique.html
6. Valeurs immatérielles transmises aux archives pour mémoire www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/plates-formes
7. Système d'information, de préservation et d'accès aux données http://vds.cnes.fr/sipad_ng/index.html