



Au CINES

Plateforme d'archivage, projets et perspectives

Pendant près de quinze ans, la question de la conservation à long terme de l'information sous forme numérique n'a été un sujet de préoccupation que pour de rares institutions scientifiques ou patrimoniales qui ont joué un rôle moteur dans la prise de conscience des risques encourus et dans l'émergence de normes de référence dans ce domaine. La progression exponentielle du numérique dans tous les domaines et le caractère impératif de la préservation de l'information ont accéléré cette prise de conscience jusqu'au plus haut niveau de la hiérarchie des administrations et des entreprises.

Aussi, ces quatre dernières années ont-elles vu l'émergence de projets, dans les sphères publique ou privée, au niveau national, européen et mondial, visant à développer des infrastructures dédiées à la conservation de l'information électronique. Parmi ces projets, celui du Centre informatique national de l'enseignement supérieur (CINES) vise à doter la communauté Enseignement supérieur et recherche française d'un véritable service d'archivage à long terme des données sous forme numérique, qui est maintenant opérationnel.

L'archivage pérenne au CINES

L'archivage pérenne des documents électroniques consiste à conserver le document et l'information qu'il contient dans son aspect intellectuel comme dans son aspect physique, de manière à pouvoir le rendre accessible et compréhensible sur le très long terme. Or, la plupart des fichiers informatiques de plus de dix ans sont aujourd'hui pratiquement illisibles, conséquence de plusieurs facteurs inéluctables tels que la connaissance perdue du contenu des fichiers, l'obsolescence des formats de fichier, la détérioration des supports physiques ou encore la disparition des logiciels ou matériels de lecture. Il y a donc des choix à faire concernant la définition des processus et des pratiques de préservation à mettre en œuvre pour atténuer les effets de ces risques lorsqu'ils se produiront...

Depuis 2004, le CINES travaille à la mise

en place d'un service pour l'archivage pérenne du patrimoine scientifique, qui permet à tout organisme produisant ou collectant en grande quantité des documents électroniques, dont le contenu possède une valeur patrimoniale avérée pour la communauté, d'initier un projet d'archives, dans le respect du contexte légal archivistique français.

Tout d'abord, une équipe dédiée à la plateforme d'archivage a été constituée, chargée de couvrir les aspects organisationnels (définition et expertise des processus métiers et des méthodes), et « culturels » (renforcement des collaborations entre informaticiens et archivistes), tout en s'assurant de la réalisation technique. Ensuite, les aspects fonctionnels ont été analysés, à la fois sur le plan théorique et sur le plan pratique, avec notamment les retours d'expérience d'autres organismes travaillant sur des projets similaires. La future plateforme a commencé à se dessiner, en suivant des contours donnés par les normes internationales en vigueur.

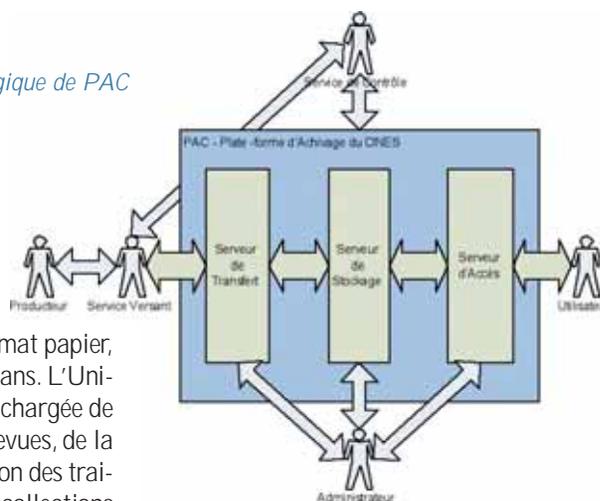
Quelques exemples : pour éviter la perte d'informations relatives au document électronique et à son contenu, des métadonnées génériques décrivant les propriétés du document (auteur, titre, résumé, mots clés, etc.) ont été utilisées afin de le situer dans son contexte et d'en préserver le sens. Il a également été décidé d'attribuer un identifiant unique et pérenne aux documents au moment de leur archivage pour permettre, notamment, de les retrouver et de les référencer. Les formats de fichiers durables ont été privilégiés afin d'éviter une obsolescence trop rapide des formats de fichiers acceptés par la plateforme. Tout document versé est contrôlé pour s'assurer que les formats de fichiers qu'il contient sont conformes. En outre, des procédures de veille technologique et de migration logique ont été élaborées pour identifier les formats émergents ou obsolètes, et migrer les fichiers vers un format offrant de meilleures garanties de pérennité lorsque c'est nécessaire. Des outils de gestion du vieillissement des supports utilisés pour stocker les documents, ainsi que des procédures de migration physique, ont été mis en place, accompagnés d'un effort de veille et d'anticipation sur les technologies émergentes en termes de médias de stockage.

Après trois années de conception et de développement, une première version du système PAC [1] a été mise en service, au printemps 2007, avec, comme axe initial, une intégration avec l'application STAR [2] (développée sous la responsabilité de l'ABES) pour le dépôt, la diffusion, le référencement et l'archivage des thèses électroniques.

Toutefois, l'infrastructure matérielle initiale ne permettant pas de gérer le volume de données prévu pour de nouveaux projets d'archives, il a été décidé de procéder à un appel d'offres pour l'acquisition au printemps 2008 d'une plateforme capable de gérer de larges volumes (plus de 40 Teraoctets). C'est donc une deuxième version du système PAC qui est actuellement exploitée, privilégiant toujours la même approche générique, évitant ainsi le traitement des projets au cas par cas, et permettant de mutualiser la plateforme pour tous les projets d'archives.

L'architecture de la plateforme d'archivage

Le système PAC a été conçu comme un ensemble de trois serveurs logiques, s'inspirant du modèle proposé par la norme ISO 14721 (OAIS) [3], mettant à disposition des différents acteurs impliqués les principales fonctionnalités du processus d'archivage. La plateforme se compose d'un serveur de transfert auquel le service versant, qui collecte les documents auprès de sa communauté de producteurs, pourra transmettre ses archives, d'un serveur de stockage, où sont conservés les documents sous la responsabilité du service d'archives qui en assure l'administration, et d'un serveur d'accès, où le service versant et éventuellement les utilisateurs autorisés à consulter ses archives pourront rechercher et obtenir une copie des documents archivés. L'ensemble est supervisé par un service de contrôle, qui s'assure que les échanges respectent la codification en vigueur (Code du patrimoine en matière de communicabilité des archives, législation relative à la propriété intellectuelle, au droit des auteurs, au droit de reproduction et au droit de représentation des œuvres de l'esprit).



L'implémentation des processus d'échanges s'appuie très largement sur le **standard d'échanges de données pour l'archivage** [4] défini par la DGME [5] et la DAF [6], qui est en cours de normalisation au niveau européen, et qui est utilisé pour l'implémentation du projet Pil@e des Archives de France.

Les projets d'archives au CINES et les perspectives

Actuellement, deux projets sont en cours d'exploitation.

Le premier concerne l'archivage pérenne de thèses collectées par l'ABES – Agence bibliographique de l'enseignement supérieur – par l'intermédiaire de l'outil STAR, initié à la suite de l'arrêté du 7 août 2006 relatif aux modalités de dépôt, de signalement, de reproduction, de diffusion et de conservation des thèses ou des travaux présentés en soutenance en vue d'un doctorat [7]. Cette disposition prévoit qu'un doctorant, qui doit déposer son mémoire de thèse dans la bibliothèque de l'université dans laquelle il va effectuer sa soutenance, peut le faire soit sous forme papier, soit sous forme électronique. Dans ce deuxième cas, la bibliothèque va déposer les documents numériques constituant le mémoire de thèse à l'ABES, grâce à l'application STAR, puis ajouter des informations qualifiant la thèse, sous forme de métadonnées. Après plusieurs phases de validation, la thèse sera archivée au CINES. L'interfaçage entre la plateforme PAC et STAR a été réalisé, et le projet est maintenant en exploitation depuis près d'un an. Après avoir démontré la faisabilité, et validé les choix retenus pour la première version de la plateforme, il est en phase de montée en charge, qui se fait au rythme du déploiement de l'application STAR dans les universités.

Le deuxième projet concerne l'archivage de revues en sciences humaines et sociales (SHS), numérisées dans le cadre du programme national Persée. Cette initiative vise à valoriser et à préserver des collections

rétrospectives originales, au format papier, dont certaines ont plus de cent ans. L'Université Lumière (Lyon-II) y est chargée de la digitalisation de masse des revues, de la centralisation et de la robotisation des traitements, de la description des collections et de leur mise en ligne via le portail Persée. La chaîne de numérisation est maintenant intégrée à la plateforme PAC et les documents électroniques créés sont déposés au CINES, pour une préservation à long terme. D'autres projets d'archives sont en cours, notamment l'archivage des documents versés dans la plateforme d'archives ouvertes HAL [8] ou l'archivage de corpus sonores pour le CRDO [9], dans le cadre d'un projet pilote pour le TGE-Adonis [10], ou encore l'archivage des vidéos produites et diffusées par Canal-U.

En parallèle à ces projets, le CINES est membre de plusieurs initiatives ou groupes de travail au plan national ou européen : il participe depuis 2004 au groupe PIN (Préservation de l'information numérique) de l'association ARISTOTE, dont les principaux acteurs français du domaine de l'archivage public (BNF, CNES, DAF) font également partie. Au niveau européen, le CINES est membre, depuis 2007, de l'Alliance pour la préservation de l'information scientifique. Cette initiative a pour objectif la coordination de la mise en place d'une infrastructure européenne pour l'archivage pérenne des données scientifiques. À cette même date, le CINES a été référencé comme centre de formation sur l'archivage par la DPE (Digital Preservation Europe). Le CINES est maintenant un acteur reconnu du domaine de la préservation à long terme des documents numériques. Fort de la mission nationale que lui a confiée le ministère de l'enseignement supérieur et de la recherche (MESR), il est appelé à jouer un rôle clé dans la réussite, au niveau national, d'une stratégie pour l'archivage pérenne des documents électroniques produits par la communauté Université-Recherche. Il reçoit aujourd'hui de nombreuses sollici-

tations émanant d'universités, de bibliothèques ou de laboratoires, pour divers services dans ce domaine, parmi lesquels l'aide et le conseil à la construction de projets d'archivage à long terme, les retours d'expérience ou encore des propositions de mise en place en partenariat de nouveaux projets d'archives. L'année 2009 verra la poursuite et le développement de tous ces projets, mais elle sera également une année de consolidation et de professionnalisation qui devrait se concrétiser par l'aboutissement d'une démarche de certification du service d'archives à la fin 2009.

Olivier Rouchon
 ✉ rouchon@cines.fr

- [1] PAC - Plateforme d'archivage du CINES <http://www.cines.fr/>
- [2] STAR - Signalement des thèses, archivage et recherche - Agence bibliographique de l'enseignement supérieur (ABES) <http://www.abes.fr/>
- [3] OAIS - Open Archival Information System – Modèle de référence pour un système ouvert d'archivage d'information http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf
- [4] https://www.ateliers.modernisation.gouv.fr/ministeres/projets_adele/a103_archivage_elect/public/standard_d_echange_d_archives_echanges_v0/downloadFile/file/archives_echanges_v0-1_description_standard_v1-0.pdf?nocache=1141748589.19
- [5] DGME - Direction générale de la modernisation de l'État <http://www.modernisation.gouv.fr/>
- [6] DAF - Direction des Archives de France <http://www.archivesdefrance.culture.gouv.fr/>
- [7] http://www.legifrance.gouv.fr/jo_pdf.do?cidTexte=JORFTEXT000000635069
- [8] HAL - Hyper article en ligne - Archives ouvertes <http://hal.archives-ouvertes.fr/>
- [9] CRDO - Centre de ressources pour la description de l'oral <http://crdo.risc.cnrs.fr/exist/crdo/>
- [10] <http://www.tge-adonis.fr/>

Francis Dumas, directeur du Centre informatique national de l'enseignement supérieur CINES ✉ www.cines.fr
 Olivier Rouchon ☎ 04 67 14 14 67 📠 52 37 63
 📧 950 rue de Saint-Priest 34097 MONTPELLIER CEDEX 5

