

prévenir les coupures de courant : elle est garantie par deux lignes électriques d'ERDF indépendantes et par des groupes électrogènes ; en outre, divers dispositifs de protection sont installés. Les conditions climatiques (température, humidité) des salles hébergeant les machines sont contrôlées. L'accès au site est sécurisé (filtrage des visiteurs, accès restreint à certaines zones), de même que les connexions *via* le réseau (grâce à des pare-feux, des protocoles sécurisés, des antivirus). En outre, toutes les données informatiques du centre (archives et systèmes) sont sauvegardées quotidiennement. Cette sauvegarde s'effectue de façon incrémentaire, c'est-à-dire qu'on n'enregistre que les changements.

mentale, c'est-à-dire qu'on n'enregistre que les changements.

La plate-forme d'archivage du CINES, conforme aux normes internationales du domaine, est fondée sur trois « serveurs logiques », respectivement dits de transfert, de stockage et d'accès. Le premier reçoit les documents à archiver, le deuxième les conserve et le troisième permet leur consultation. Décrivons-les plus en détail.

Un paquet d'information à archiver contient l'objet à conserver et un ensemble de métadonnées (des données permettant de comprendre d'autres données). Ces dernières sont élaborées en partie par l'organisme

dit versant, qui fournit les documents à archiver (c'est en général l'organisme qui les produit) et en partie par l'archivier. Elles décrivent le contexte (organisme producteur, organisme versant), le document lui-même (titre, auteur, format), les paramètres de gestion (localisation physique, dispositif de stockage, etc.). Un autre type de métadonnées, qui ne sont pas associées à un document précis mais à la plate-forme d'archivage, décrit les informations dites de représentations, essentielles à la lecture de l'archive : les caractéristiques des formats, les structures des métadonnées, la description de la plate-forme d'archivage et des normes utilisées...

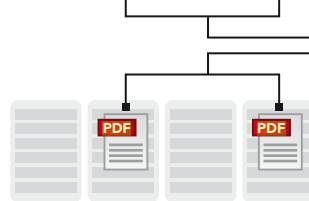
PLUSIEURS COPIES SUR DES SUPPORTS VARIÉS

Les supports physiques sont vulnérables et les estimations de leur durée de vie varient selon les études. Pour parer cette menace, cinq copies du document à archiver sont réalisées sur des types de supports différents : deux sur des bandes magnétiques et trois sur des disques durs (dont une située à plus de 300 kilomètres des autres, afin de préserver le document d'une catastrophe majeure, tel un séisme). Chaque document est accompagné d'un ensemble de métadonnées qui permettent de l'identifier, de le localiser et de faire évoluer son format.

Bandes magnétiques



Disques durs



PDF

PDF

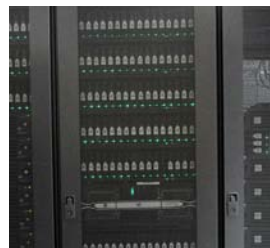
Disque dur dans un centre de données éloigné de plus de 300 km

Métadonnées :

Titre : L'archivage électronique
Auteur : Daumas, Francis
Identifiant : ark://85579/123456789
Date d'archivage : 30-09-2013
Format : PDF
Version du format : 1.7
...



Les robotiques gèrent de façon automatique un ensemble de bandes magnétiques (ci-dessus, une robotique vue de l'extérieur et, ci-contre, à gauche, vue de l'intérieur). Deux copies du document y sont enregistrées. Lorsqu'une archive est demandée, un bras robot préleve la bande où elle est stockée et la place dans un dispositif de lecture.



Les baies de stockage renferment de multiples disques durs. Deux copies du document y sont stockées, dans des salles distinctes.



Le serveur de transfert contrôle les documents reçus, en vérifiant notamment la conformité des métadonnées et des formats de fichiers. Il peut aussi effectuer diverses transformations, telle la conversion d'un format non libre (par exemple *Word*) en un autre (tel *Open office*). Il enrichit ensuite les métadonnées par des paramètres d'archivage, tel un identifiant international unique (une succession de lettres et de chiffres caractéristique du document). Enfin, il transmet le document au serveur de stockage.

Ce dernier en réalise cinq copies. Deux d'entre elles sont enregistrées dans deux baies de stockage distinctes (des armoires comprenant un certain nombre de disques durs), situées dans deux salles distantes. Deux autres sont inscrites sur des bandes magnétiques gérées par des dispositifs automatisés nommés robothèques. Ces robothèques, de deux pétaoctets chacune, sont aussi dans deux salles différentes. Une dernière copie est réalisée à plus de 300 kilomètres du site, dans un autre centre informatique, pour garantir les données contre les risques de catastrophe naturelle majeure.

Enfin, le serveur d'accès reçoit les demandes de consultation des archives, vérifie les autorisations de leurs émetteurs et leur transmet ou non les documents. Il peut également effectuer diverses modifications, comme flouter des visages pour préserver l'anonymat des personnes lors de l'envoi d'une photographie.

Des métadonnées bien renseignées répondent aux problèmes de localisation et de référencement des documents soulevés au début de cet article. Qu'en est-il de l'obsolescence et de la détérioration des supports et des formats ? Deux options se présentent : conserver en état de marche tous les dispositifs matériels et logiciels permettant la lecture de l'archive, afin de les réutiliser plus tard ; effectuer une migration des supports et des formats à mesure que les techniques évoluent (là encore, les métadonnées sont essentielles). Le CINES a choisi la seconde option.

Ainsi, les documents sont transférés vers les types de supports les plus récents – par exemple, aucune donnée n'est conservée sur disquette aujourd'hui. En outre, la durée de vie des supports est prise en compte pour assurer un renouvellement adapté. Les documents stockés sur des disques durs sont ainsi transférés tous les cinq ans, et ceux stockés sur bande magnétique tous les dix ans.

Les migrations concernent aussi les formats (PDF, JPEG...). La plate-forme du CINES accepte une quinzaine de formats, couvrant les différents types de données. Le choix est fait en privilégiant ceux dont les caractéristiques sont publiées et disponibles, et dont on pense que les versions successives seront compatibles avec les documents anciens. Une procédure de migration se déclenche lorsqu'un format est obsolète. Elle consiste en général à adapter un document aux versions successives d'un même format, mais il arrive aussi qu'un format disparaisse (tel celui des fichiers créés par le logiciel de bureautique *Claris*, abandonné au début des années 2000) et qu'il faille en changer. Elle se déroule en étroite collaboration avec l'organisme versant des données archivées, qui valide la nouvelle forme du document.

ENVIRON 40 RISQUES SONT IDENTIFIÉS aujourd'hui, tels qu'un incendie ou un format de fichier qui se périmé.

La gestion des archives à long terme nécessite également une « démarche qualité » – un ensemble de bonnes pratiques techniques et organisationnelles. Le respect des normes (assurant notamment l'interopérabilité avec d'autres systèmes d'archivage ou de collecte de données), la vérification du bon encodage du document selon le format choisi, les copies multiples et le contrôle de leur état participent à cette démarche. Au niveau organisationnel, tous les processus mis en œuvre (par exemple, l'ensemble des actions à réaliser pour faire migrer un document d'un format à un autre) sont répertoriés et conservés par écrit. Cela permet de transmettre aux générations suivantes le savoir-faire et l'historique du service, mais aussi de mieux analyser et d'améliorer les procédures. En outre, les risques potentiels et les actions à mener pour y parer sont actualisés tous les six mois. Environ 40 risques sont identifiés, tels qu'un incendie ou un format de fichier qui se périmé.

Offrir un service d'archivage numérique ne se résume donc pas à acquérir des dispositifs de stockage et à y copier des données. La pérennisation de documents nécessite une gestion active et une surveillance régulière, aussi bien des logiciels que du matériel. ■

■ LES AUTEURS



Francis DAUMAS est directeur du CINES [Centre informatique national de l'enseignement supérieur], à Montpellier.



Marion MASSOL est responsable du Département d'archivage du CINES.

■ SUR LE WEB

Groupe PIN (pérennisation de l'information numérique) de l'association Aristote : <http://pin.association-aristote.fr>

■ BIBLIOGRAPHIE

Collectif, *Archivage numérique pérenne*, *La Gazette du CINES*, février 2013 (www.cines.fr).

J.-C. Hourcade et al., *Longévité de l'information numérique : Les données que nous voulons garder vont-elles s'effacer ?*, EDP Sciences, 2010.

Norme iso 14721, ou OAIS [*Open archival information system*, soit *Système ouvert d'archivage d'information*]. www.iso.org