

Francis Daumas et Marion Massol

Au cœur d'un centre d'archivage

Pour stocker le déluge récent de données numériques, notamment scientifiques, de grands centres d'archivage ont été mis en place. Comment fonctionnent-ils ? Exemple avec le CINES, un centre national à Montpellier.

La plupart du temps, quand nous souhaitons conserver un document numérique, nous le copions sur un support externe, que nous stockons dans un placard. Mais sa récupération longtemps après peut se révéler compliquée. D'abord, on note rarement, ou de façon imprécise, la localisation du support et les informations pertinentes sur le contenu. Si l'on parvient à retrouver le support, encore faut-il pouvoir le lire : imaginez une photo de mariage stockée sur une vieille disquette des années 1980 ! Les équipements de lecture de ce support ont parfois disparu ou sont difficiles à trouver après tant d'années. Supposons que vous ayez trouvé le bon équipement : vous devez aussi disposer du logiciel capable de reconnaître le format du

fichier et de restituer le contenu sous une forme intelligible. Or les formats évoluent et les logiciels peuvent disparaître ou créer des fichiers qui ne seront pas lisibles par leurs futures versions. Enfin, le fichier initial ne doit pas comporter trop d'erreurs et son support doit avoir été bien conservé.

Cet exemple illustre les difficultés posées par la préservation à long terme des données : l'obsolescence et la détérioration du support, la disparition de l'outil logiciel de lecture, les évolutions du format du fichier et la perte de la localisation et de la signification de son contenu. Pour répondre à ces difficultés, des techniques d'archivage ont été développées. Nous les présenterons à travers l'exemple d'un des principaux sites d'archivage français : celui du CINES (Centre

informatique national de l'enseignement supérieur), à Montpellier, un établissement public placé sous la tutelle du ministère de l'Enseignement supérieur et de la Recherche.

Des données variées

Aujourd'hui, l'information numérique est omniprésente et protéiforme. Les données à conserver sur le long terme sont de natures variées : administratives, culturelles, scientifiques (issues d'observations ou de calculs), etc. Les livres des bibliothèques, les thèses, les articles scientifiques, les encyclopédies sont numérisés (voir le dossier « *Le savoir numérisé* » dans ce numéro). En résultent une multitude de documents à préserver, sous divers formats : textes, sons, images, vidéos...