



Jardin des Plantes
© MNHN - François Grandin.

« Pour mener à bien cette numérisation, nous avons besoin d'un système de stockage sécurisé, économique et évolutif : c'est Active Circle »

Henri Michiels, Directeur Informatique



MUSÉUM NATIONAL D'HISTOIRE NATURELLE

STOCKAGE DES IMAGES DU GRAND HERBIER

Le Muséum National d'Histoire Naturelle

Créé en 1635 par un décret royal, ce qui était au départ le jardin des plantes du roi devient le Muséum National d'Histoire Naturelle en 1793. Fort d'une expertise reconnue, ses missions couvrent la conservation et l'enrichissement de collections exceptionnelles, la recherche, l'enseignement et la diffusion des savoirs auprès du public. Deux mille personnes travaillent au Muséum, dont 500 chercheurs, aidés dans leurs tâches par la direction des services informatiques (DSI) dirigée par Henri Michiels. La DSI gère plus de deux mille postes de travail, 120 serveurs physiques ou virtuels répartis dans trois salles informatiques situées sur le Jardin des Plantes et interconnectées par fibre optique, ainsi que les moyens de stockage.

La numérisation de l'herbier bouscule les besoins

Les besoins de stockage évoluent de façon importante depuis 2004, quand la numérisation systématique des spécimens de collections commence, générant de nombreuses images de taille allant de quelques Mo à près de 200 Mo. Pour répondre à ce besoin, le choix s'est porté sur un réseau SAN à base de fibres optiques supportant le protocole Fibre Channel, qui interconnecte les trois salles informatiques situées au Jardin des Plantes. Sur ce réseau sont connectés des baies de stockage SAN, ainsi que les serveurs. Un système de sauvegarde sécurise l'ensemble des données. Mais le projet de rénovation du Grand Herbier initié en 2009 bouscule cette architecture. Le Grand Herbier, qui remonte au 17^{ème} siècle est le plus grand du monde avec celui de Londres, avec près de 10 millions de planches au format A3. Le projet de rénovation comprend le reconditionnement des planches, l'intégration des spécimens en attente, la réorganisation des collections et bien sûr la numérisation, indispensable à la diffusion de la connaissance. La projection du volume à stocker avoisine les 500 teraoctets !

L'extension du SAN n'est pas possible

Le stockage est tout d'abord envisagé sur le SAN, mais une étude de coût montre que cette solution n'est pas envisageable économiquement compte tenu des budgets et des volumes prévus pour l'herbier. « Nous avons rapidement écarté la solution à base de SAN, car le coût au Go est trop élevé et il faut le doubler d'une sauvegarde, et donc ajouter des coûts supplémentaires de logiciel et de matériel de stockage » déclare Henri Michiels, directeur informatique du Muséum National d'Histoire Naturelle.

Recherche très grand stockage à faible coût

Le besoin de stockage est remis à plat par l'équipe informatique : les données doivent être accessibles en permanence, mais un temps de latence est acceptable. Elles doivent être protégées sans faire appel au système de sauvegarde hebdomadaire classique, qui n'est pas adapté à ces données qui ne sont pas modifiées. Mais sans sacrifier la sécurité, car la perte de ces données originales n'est pas envisageable, il n'existe pas de copie ! Et côté infrastructure, le système doit permettre la conservation à très long terme, et avoir un coût minimal au Go à la fois en acquisition et en fonctionnement, ce qui implique de consommer peu d'énergie électrique. En bref, un système de stockage pérenne, ouvert, sécurisé et très peu coûteux... cette équation semble difficile à résoudre.



L'herbier avant numérisation
© MNHN - Bernard Faye

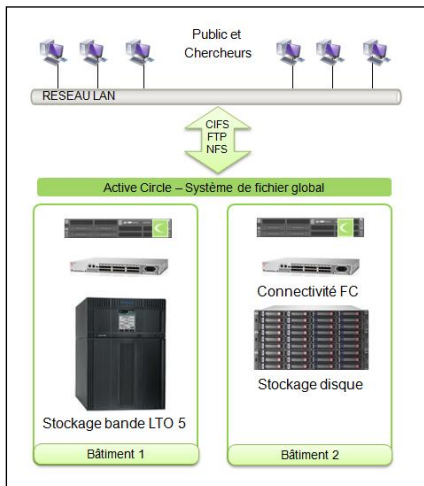


Vue de la galerie de Phanérogamie
© MNHN - Laurent Bessol

« L'un des intérêts de la solution choisie est de sécuriser les données sans faire de sauvegarde.

Chaque donnée nouvelle est dupliquée et stockée physiquement en deux endroits. Cette duplication automatique au fil de l'eau est la bonne solution. »

Henri Pham - administrateur de la solution



Architecture de stockage



Planches de l'herbier de Rousseau
© MNHN - Bernard Foye

Une sélection impitoyable

Suite au cahier des charges publié en 2011, le MNHN reçoit des réponses variées mais beaucoup ne correspondent pas au besoin. Les solutions des constructeurs sont trop coûteuses et trop propriétaires, manquent d'ouverture et d'évolutivité. La solution Active Circle se démarque car ouverte et non propriétaire, elle résout l'équation de la capacité de stockage au meilleur coût tout en garantissant la protection des données.

La solution Active Circle : un HSM ouvert et distribué

La solution proposée par Active Circle est deux fois moins chère que les autres solutions ; il s'agit d'un logiciel fonctionnant sur matériel banalisé, qui propose un environnement de stockage de fichiers hiérarchisé de type HSM sur disque et bande LTO, avec gestion de la protection des données en continu et garantie de pérennité des données grâce au format TAR. La solution est portée par l'intégrateur CEFI, partenaire certifié Active Circle, qui possède une forte expérience de déploiement de solutions Active Circle. Le matériel Dell choisi par l'intégrateur CEFI se compose d'une librairie de bande LTO 5 et d'une baie de disque. Les deux éléments de stockage constituent grâce à Active Circle un espace de stockage global. Ils sont placés dans deux bâtiments distincts de manière à protéger les données contre un sinistre.

Déploiement et numérisation en parallèle

Le déploiement de la solution et la numérisation s'effectuent en parallèle. La numérisation démarre véritablement de manière massive en 2010. Les planches doivent être numérisées au format TIFF, chaque fichier pesant environ 50 Mo. Elles sont ensuite compressées en JPEG de 5 Mo qui sera le format utilisé pour la conservation et la consultation. Le format de numérisation Tiff est transitoire, les fichiers sont stockés temporairement sur disques durs externes. La phase de numérisation dure deux ans et se termine en décembre 2012. Une phase de test se déroule au printemps : la solution complète est conforme aux attentes et entre en production. A la fin de l'été les premières données numérisées sont accueillies par le système Active Circle.

Les données sont accessibles

Le public et les chercheurs accèdent aux données de l'herbier au moyen d'une application web, qui elle-même repose sur le système de fichier global d'Active Circle. Cette application dispose donc d'un système de fichiers extensible et sécurisé. L'accès est transparent, de l'ordre de la seconde si la donnée est sur disque, et de l'ordre de la minute si elle se trouve sur bande LTO. Une donnée accédée remonte sur disque, les lectures suivantes sont donc immédiates. Les temps d'accès sont totalement acceptables pour une lecture dans un tel volume de données.

Les données sont sécurisées au fil de l'eau

« L'un des intérêts de la solution choisie est de sécuriser les données sans faire de sauvegarde. Chaque donnée nouvelle est dupliquée et stockée physiquement en deux endroits. Cette duplication automatique au fil de l'eau est la bonne solution. » souligne Henri Pham, administrateur de la solution. En effet Active Circle gère globalement deux espaces de stockage qui sont physiquement dans deux bâtiments. Pour chaque fichier logique, Active Circle stocke une copie physique sur chacun de ces deux espaces, ce qui est transparent pour les utilisateurs.

Stocker pour diffuser la connaissance

Maintenant que le Grand Herbier est numérisé, de nouveaux besoins s'expriment : « Avec ce système de stockage, nous envisageons l'avenir sereinement. Nous pouvons faire face aux nouveaux besoins qui émergent, comme numériser en 3D selon des procédés de coupes tomographiques, et étendre la numérisation à d'autres collections. Avec la numérisation, nous facilitons l'accès au public et aux chercheurs, et nous remplissons notre mission de diffusion de la connaissance. Et pour mener à bien cette numérisation, nous avons besoin d'un système de stockage sécurisé, économique et évolutif : c'est Active Circle » déclare Henri Michiels.

Créée en 2002, la société Active Circle est un éditeur de logiciel spécialisé dans le stockage et l'archivage numérique. Active Circle s'adresse aux organisations qui gèrent de grands volumes de contenu vidéo, images, données scientifiques et techniques, et leur permet de maîtriser le cycle de vie des données tout en optimisant les coûts de possession et d'administration du stockage.

Active Circle est une marque déposée par Active Circle SA. Les autres noms ou représentations sont mentionnés à des fins d'identification et appartiennent à leurs propriétaires respectifs. © ACTIVE CIRCLE 2013 - Ce document ne peut être copié ou reproduit sans accord écrit. - www.active-circle.com