

## STOCKAGE.

# Un patrimoine cartographique bien conservé à l'IGN

*Pour archiver de manière hiérarchisée et sécurisée les dizaines de téraoctets d'images collectées et de cartes produites, l'IGN a adopté la solution de l'éditeur français Active Circle.*

## L'ENTREPRISE ÉTUDIÉE

## Institut géographique national (IGN)

ACTIVITE : production et ventes de cartes et de bases de données géographiques.

SIEGE : Vincennes (94).

EFFECTIF : environ 1700 salariés.

BUDGET 2008 : 143,1 M€.

## PROBLÈME À RÉSOUDRE

• Archiver l'ensemble des données géographiques du portail de l'IGN.

## SOLUTIONS DÉPLOYÉES

• Active Circle, logiciel de stockage et d'archivage de l'éditeur français éponyme. Bandothèques Neo 8000 d'Overland. Serveurs NAS Exanet.

## DIFFICULTÉS RENCONTRÉES

• Suivi des clés de sécurité MD5.  
• Archivage des fichiers de plus de 800 Go.

## LES COÛTS

• 238 000 € pour l'ensemble de la solution.

Le service de l'information en ligne (Siel) est en charge de l'archivage des données cartographiques produites par l'Institut géographique national ainsi que de celles diffusées sur son site, le géoportail. Il doit ainsi gérer deux types de données : des images vectorisées – les cartes – et des images raster, stockées aux formats Tiff et ECW (Enhanced Compression Wavelet). « Les unités de production de l'IGN nous envoient leurs données sous la forme de SIP (Submission Information Package) qui incluent les images, compressées au préalable au format TAR, et leurs métadonnées. Les flux sont bien sûr beaucoup plus importants pour les images raster », précise Yann Le Disez, le responsable du département diffusion et archivage au Siel.

Auparavant, la division de l'IGN utilisait le système Dorostore conjointement à un catalogue électronique des supports externalisés. Les temps d'archivage et de restauration étaient très longs. En 2007, pour faire face à la croissance exceptionnelle des informations à archiver (50 To jusqu'en 2007 et 25 To estimés à l'époque pour 2008), le Siel décide

de changer de système et d'opter pour un archivage unifié, sécurisé et évolutif sur cartouche LTO (Linear-Tape Open). Pour cela, il lance un appel d'offres auquel répondent quatre intégrateurs.

## LES BESOINS

### S'adapter aux meilleures pratiques

Sur les quatre solutions proposées, celle que propose l'intégrateur Komposite remporte la faveur du Siel : c'est la plus à même de s'adapter aux développements spécifiques déjà réalisés auparavant. « Conformément à un appel d'offres européen ouvert, nous avions défini trois critères de sélection : la technique, le prix et les délais de livraison. Un intégrateur très renommé nous proposait une offre avec l'intégralité des briques du système, mais cela ne convenait pas car nous voulions un outil qui nous permette de conserver les briques que nous avions déjà produites », explique Yann Le Disez.

Le département archive en effet ses données selon des procédures bien définies au travers du modèle conceptuel OASIS (Open Archival Information System). Celui-ci impose notamment

## LE CALENDRIER DU PROJET

AOÛT 2007	Fin de l'analyse des besoins et lancement de l'appel d'offres.
DÉC. 2007	Choix de la solution et mise en place d'une salle spécifique.
FÉV. 2008	Déploiement d'Active Circle 3.
MARS 2009	Migration vers la version 3.1.
FIN 2009	35 To d'archives sont ajoutés aux 25 To générés en 2008 et aux 50 To résultant de la migration du précédent système.
A VENIR 2010	Prévision de 50 To à archiver.
2011	Prévision de 100 To à archiver.

de réaliser, avant de procéder à l'archivage proprement dit, des paquets d'informations compressés – les fameux SIP – reposant sur un standard de compression ouvert, plutôt que de reposer sur les fonctions et capacités de compression propriétaires du système d'archivage. L'opération est donc réalisée au niveau des unités de production par un programme développé en interne.





**Yann Le Disez**, responsable du département diffusion et archivage au Siel, division de l'IGN

« Le système est bien adapté à la croissance de nos besoins »

« Depuis la mise en production d'Active Circle, nous n'avons fait que corriger quelques petits bugs. Nous avons fait migrer toutes les anciennes archives, soit près de 50 To. Avec des flux d'entrée qui sont en moyenne de 1 To par semaine, nous disposons désormais d'environ 110 To de données archivées. Il nous arrive d'avoir à restaurer des fichiers, environ deux à trois fois par mois, avec une volumétrie qui dépend largement du besoin, très spécifique. Cela peut représenter une vingtaine de teraoctets par an. C'est la préparation des données et l'administration qui nécessitent le plus d'attention. Pour le reste, une fois le travail en amont réalisé, il suffit de pousser les données et de laisser les procédures d'archivage s'exécuter. »

en panne, le système continue à travailler sur les deux autres le temps de la réparation.

#### L'EXPLOITATION

### Une étape de validation obligatoire

En pratique, la procédure est plus compliquée. Les fichiers soumis par les unités de production et prêts à archiver sont déposés depuis les serveurs de production dans des répertoires de ressources partagées situés sur les deux NAS. Ces répertoires sont scrutés en permanence par un agent logiciel, qui lance dans un premier temps un script de vérification du packaging SIP (contrôle d'intégrité, validation des métadonnées). Si le lot est incorrect, un message de notification d'erreur est envoyé par mail et le fichier est mis en quarantaine : il ne passe pas à l'étape suivante et doit être régénéré par les unités de production. S'il est correct, Active Circle reprend la main et le fichier est « poussé » vers l'une des deux cellules principales du système.

#### LES ÉCUEILS

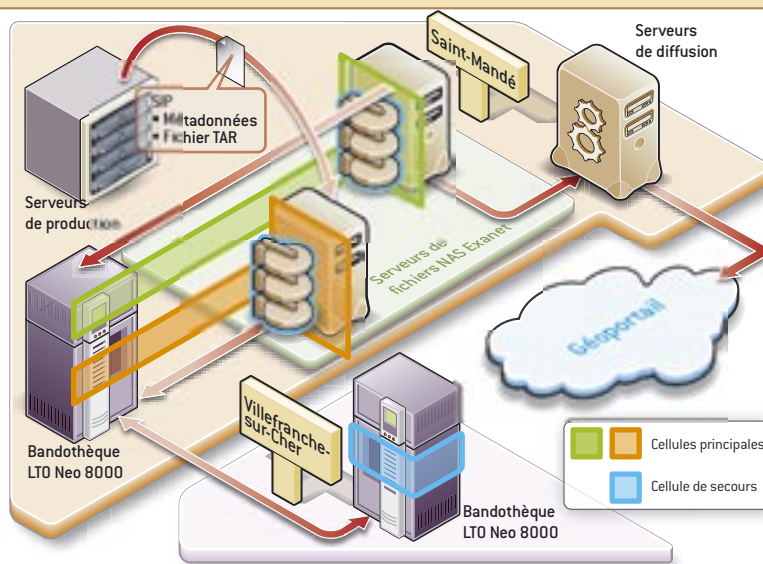
### Des processus internes contraignants

Bien que la solution d'Active Circle ait répondu en grande partie aux besoins du Siel, quelques points ont fait défaut. « La possibilité de suivre la clé d'intégrité MD5, qui accompagne le SIP, n'était pas intégrée dans Active Circle », regrette Yann Le Disez. La fonction équivalente a finalement été réalisée avec le soutien de l'intégrateur. Un autre écueil concernait non pas Active Circle mais la limitation de capacité des cartouches : le système n'était pas capable d'archiver les fichiers de plus de 800 Go. Pour ces derniers, il a donc été prévu une décomposition en deux fichiers TAR sauvegardés sur deux cartouches. Ce qui n'est pas géré avec le mécanisme intégré du logiciel Active Circle mais, en amont, au niveau de la production du contenu. ■

JONATHAN CHARTON

## Une capacité d'archivage évolutive et sécurisée

• La volumétrie importante gérée par le système est due à la fois au volume de certains fichiers mais aussi à la conservation de plusieurs niveaux de version. Il est prévu que les données restent accessibles sur bande sur les deux sites pendant une période de trente ans. Au-delà, elles seront externalisées.



#### L'ARCHITECTURE

### Un système tolérant aux pannes

Deux NAS haute performance d'origine Exanet hébergent, selon des règles hiérarchiques définies (les temps de rétention sont différents en fonction des contenus) les données cartographiques qui pourront être consultées depuis le site. Parallèlement, ils servent

de tampons pour l'archivage définitif sur les cartouches LTO4 de la bandothèque Neo 8000 installée sur le site de Saint-Mandé.

C'est le logiciel Active Circle qui gère le déplacement de ces données au sein de son système de cellules. Une cellule est constituée d'un espace de stockage sur un des NAS (4 To de cache et 25 To de stockage opérationnel) et d'une partition de la bandothèque

Neo 8000. Pour assurer un haut niveau de tolérance aux pannes, le Siel a donc activé deux cellules sur son site principal ainsi qu'une troisième, synchronisée, s'appuyant sur une seconde bandothèque, sur son site distant de Villefranche-sur-Cher. De la sorte, et c'est là l'un des avantages d'Active Circle qui fonctionne de la même manière qu'un système en grappe, si l'une des cellules tombe