



Le Système d'Archivage de l'Information Géographique à l'IGN

Présentation Groupe PIN – Mardi 18 mai 2010

Yann LE DISEZ

Responsable du Département Archivage et Diffusion

Service Information En Ligne – Institut **G**éographique **N**ational

Le Siel

Service de L'Information en Ligne

- En charge de l'archivage des données cartographiques produites par l'IGN ainsi que de celles diffusées.
- Deux types de données :
 - des données de type vecteur (des bases de données pour SIG),
 - des images raster, stockées aux formats Tiff et ECW (Enhanced Compression Wavelet)

• BD ORTHO®



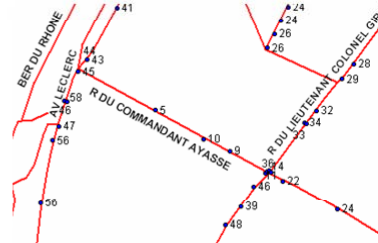
• BD TOPO®



• BD PARCELLAIRE®



• BD ADRESSE®

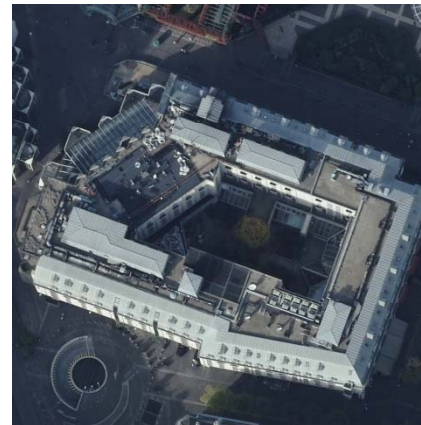
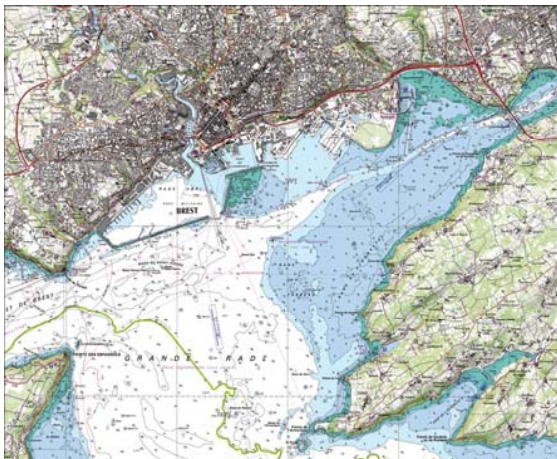




Le Siel

Service de L'Information en Ligne

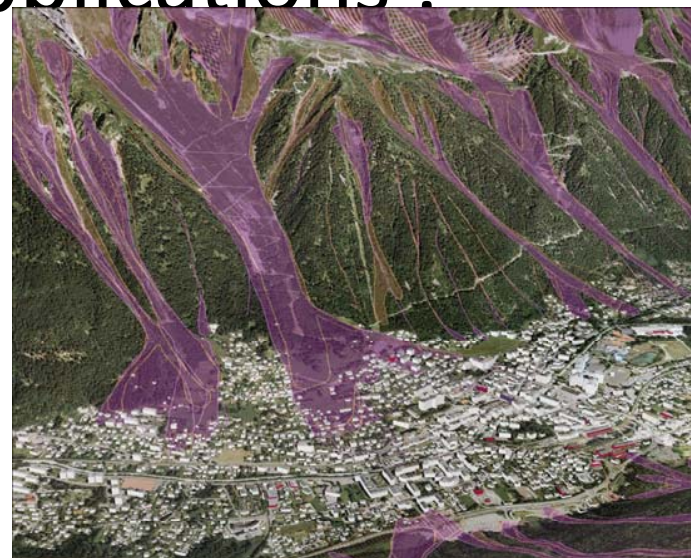
- Le workflow
 - Les unités de production envoient les données et leurs métadonnées sous la forme de SIP (Submission Information Package) au format TAR.
 - Les flux les plus importants sont ceux des images raster.





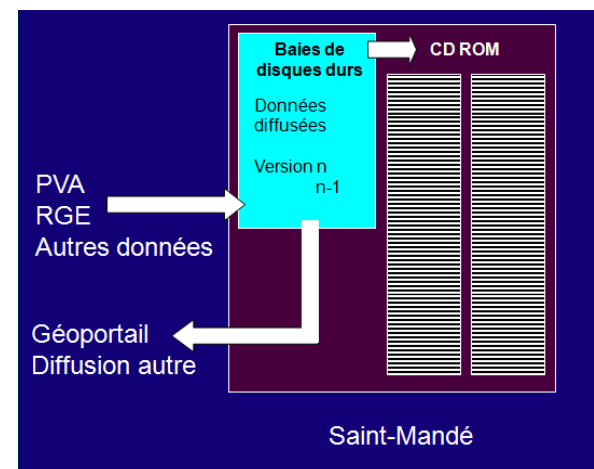
Données et applications :

- Autres données
 - Données LASER (LIDAR)
 - Dématérialisations des fonds photographiques et cartographiques
- Applications
 - Etude des risques (avalanches)
 - Aménagement (Bâti 3D)



Système existant et problématique

- Système existant
 - Système Dorotech avec stockage sur CD, DVD et bandes LTO et DT). Soit un volume total de 50 To.
21300 CD – 400 DVD
487 DLT – 520 LTO1 & 3
 - Catalogue électronique des supports externalisés
 - Temps d'archivage et de restauration
- Estimation de volume à venir, en 2007
 - 2008 : + 20 To
 - 2009 : + 30 To
 - 2010 : + 50 To
- Constat : le système ne pourra pas absorber le volume, souci de scalabilité (évolutivité)





Le projet - objectifs

- Mettre en place un système d'archivage ayant les caractéristiques suivantes
 - Stockage sur LTO
 - Évolutivité en capacité
 - Pas de modification du workflow amont : prise en compte des SIP (Submission Information Package)
 - Archivage en format non propriétaire (exemple, format TAR)
 - Accès API



Solution retenue et critères décisifs

- KOMPOSITE avec le soft Active Circle






(Active Circle : c'est un ensemble de noeuds de stockage, appelés cellules, qui communiquent entre eux.)

- Evolutivité (scalabilité) (-> 380 To nearline)
- Architecture distribuée
- Fonctionnement sur matériel standard, donc indépendance vis-à-vis des technologies
- Fonctions intégrées de HSM et réplication
- Archivage en format non propriétaire
- Accès API (Pour prise en compte de besoin spécifiques éventuels)
- Pilotage natif des lecteurs et bibliothèques de bandes de toutes technologies



Solution mise en place

- Solution déployée
 - Accès par protocoles standards ( CIFS  NFS  FTP)
 - HSM entre disque et bande
 - Format tar sur bande
 - Externalisation par génération de double LTO et envoi vers site distant qui fait partie du même cercle (avec metadata mis à jour)
- Extension développée
 - vérification d'intégrité par MD5 pour valider le dépôt de gros fichiers qui peuvent être corrompus lors de leur transfert réseau

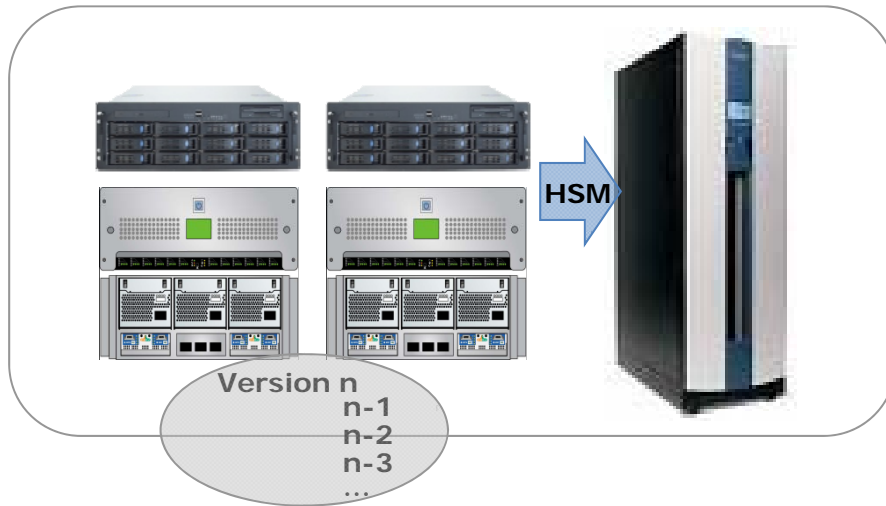


La solution déployée à l'IGN

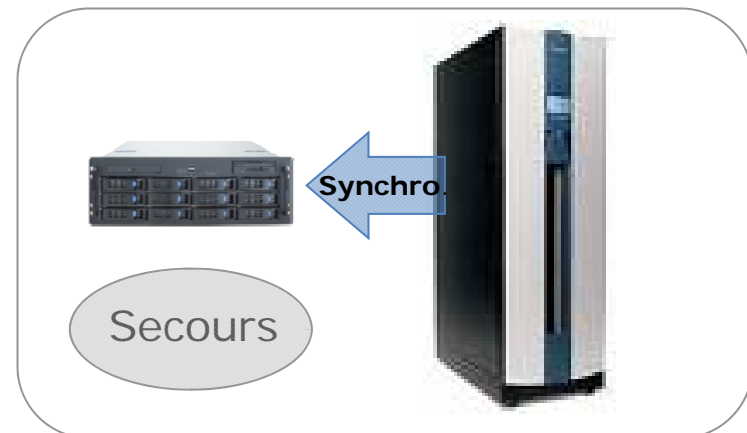
SAINT-MANDÉ



NFS/CIFS/FTP



VILLEFRANCHE



Réplication
Méta-data





Historique Projet

- Historique :
 - Décembre 2007 : choix de la solution
 - Février 2008 : déploiement par l'intégrateur Komposite. Installation de la solution hardware et software.
(Le système est opérationnel mais vide)
 - Migration des anciennes archives, soit 50 TO (3 mois avec l'aide du prestataire)
 - Mars 2009 : migration vers la version 3.1 d'Active Circle
 - 2009 : 40 To archivés
 - Quelques bugs corrigés (clôture de bandes)

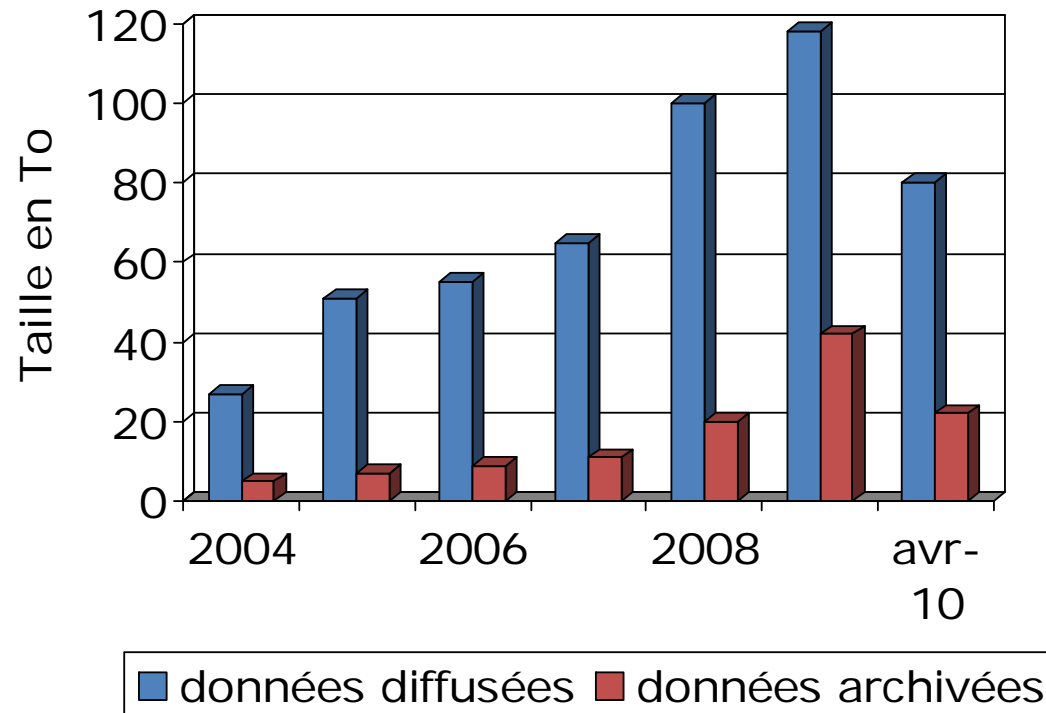


Quelques améliorations :

- Optimisation des mouvements de bandes magnétiques
- Localisation du traitement de COS
- Vérification de cohérence de classe de service (CoS)

Bilan

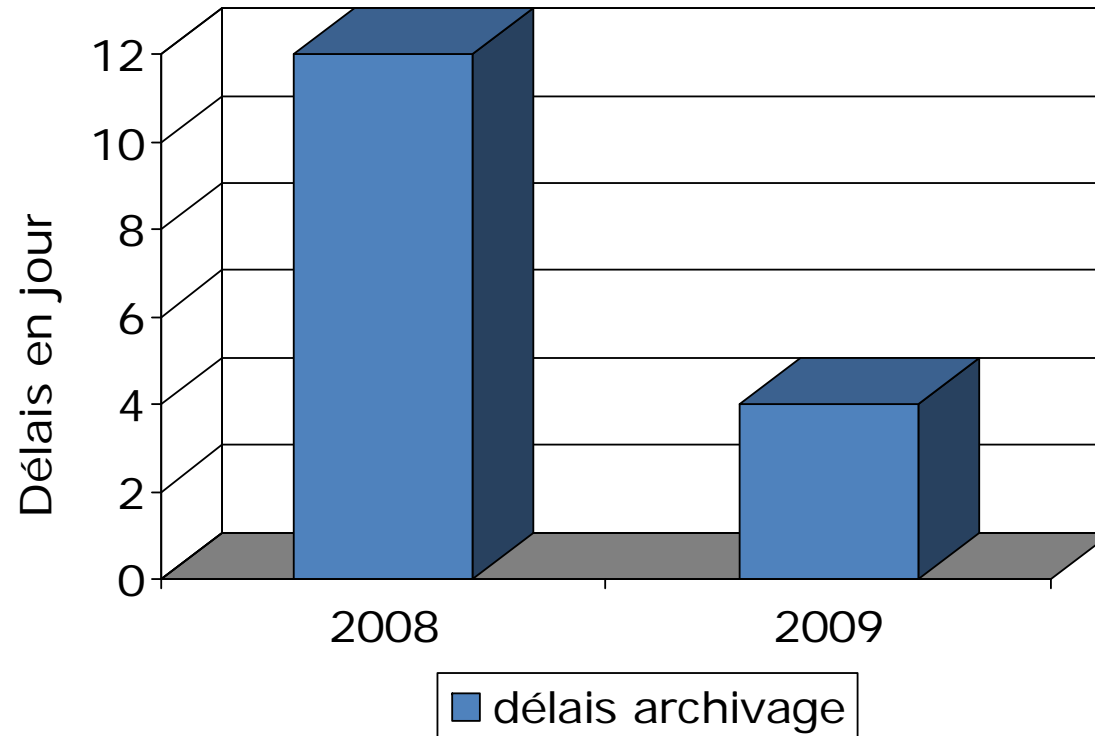
- Evolution de la volumétrie sur les 5 dernières années





Bilan

- Délais d'archivage sur 2008 et 2009 :





Bilan

- Une volumétrie toujours plus importante
 - Dématérialisation des fonds : volume = 30 To
 - Exemple des images Haute-Résolution
 - BD Ortho France entière (pixel à 50cm) : volume = 7 To
 - BD Ortho France entière (pixel à 20cm) : volume = 44 To

Nous avons pu tester en réel la fiabilité du système suite à un sinistre dans notre salle serveur :

- Données accessibles sur le site distant
- Réplication des bandes à partir du site distant



Merci de votre attention

Contact :

Tel : 01.43.98.95.80

Yann.le-disez@ign.fr