



Le service d'archivage numérique pérenne du CINES

Marion Massol / Olivier Rouchon – CINES
massol@cines.fr / rouchon@cines.fr

Journées ANF Documentation Mathématique 2013 – CIRM – 23 Mai 2013



- **Le CINES et sa mission de préservation**



- **Pourquoi archiver ?**

- **Les solutions du CINES**

- Des infrastructures
- Des engagements
- Des pré-requis pour les partenariats
- Nos partenaires actuels



- **Un modèle économique**



- **Les perspectives**





Centre Informatique National de l'Enseignement Supérieur

- Basé à Montpellier (Hérault, France)
- EPA créé en 1999, succédant au CNUSC (Centre National Universitaire Sud de Calcul) – créé en 1980
- Placé sous la tutelle de la DGRI (Direction Générale de la Recherche et de l'Innovation) et de la DGESIP (Direction Générale pour l'Enseignement Supérieur et l'Insertion Professionnelle) du Ministère de l'Enseignement Supérieur et de la Recherche
- Missions
 - Calcul numérique intensif
 - Archivage pérenne de documents électroniques
 - *Activité transverse : hébergement d'environnements informatiques*
- Plus d'information : <http://www.cines.fr/>





Infrastructures sécurisées

- 5 salles machines : 1 400 m²
- Locaux techniques : 2000 m²
- 2 lignes ERDF pour un total de 12,5 MW
- Données en double alimentation ondulée et sécurisée par groupe électrogène
- Copies et sauvegardes dans des salles distinctes + copie à distance



Ressources

- Un supercalculateur de niveau mondial
- Capacités de stockage de plusieurs PetaOctets
- Des accès réseau performants
- Des équipes d'experts





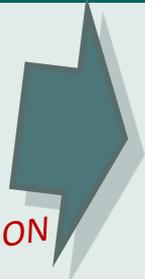
Les défis, orientations et choix pour l'archivage électronique

Voici un document
que j'ai créé en
1998...



*De quoi s'agit-il déjà ?
Est-ce bien ce qui est
indiqué sur la
disquette ?*

**METADONNEES DESCRIPTIVES + IDENTIFICATION
UNIQUE et PERENNE**



*La disquette est-elle
toujours en bon état ?*

**SUPPORT : VEILLE +
MIGRATION PHYSIQUE**



*Mon portable, acheté
en 2009, n'a pas de
lecteur de disquette...*

**ENVIRONNEMENT MATERIEL :
VEILLE TECHNO et ANTICIPATION**



*Ça marche ! Mais j'ai
perdu toute ma mise
en forme...*

INTEGRITE AUTHENTICITE



*J'ai trouvé le logiciel,
mais puis-je l'installer
et l'utiliser sous
Windows XP ?*

SYSTÈME D'EXPLOITATION



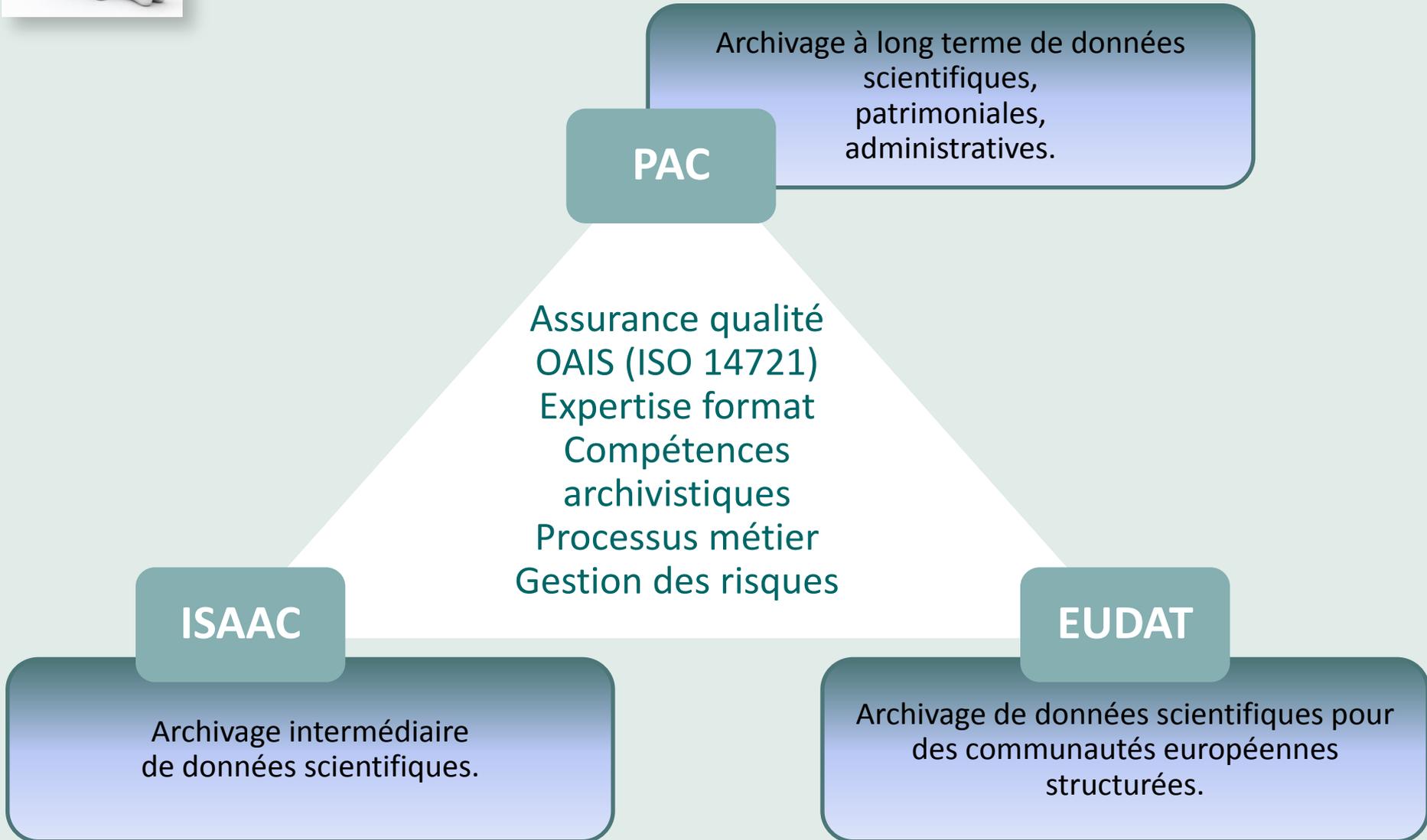
*J'ai créé ce document avec Claris
Works. Comment retrouver ce
logiciel ? Quel est le format du
document ?*

**ENVIRONNEMENT LOGICIEL : privilégier
les FORMATS DURABLES + MIGRATION
LOGIQUE**





Les infrastructures d'archivage du CINES



La sécurisation de l'information : un ensemble de moyens, une même finalité



CINES : site sécurisé

Onduleurs



Salles machines



Equipements d'archivage : PAC, ISAAC, EUDAT@CINES

Service Versant



Groupe Electrogène



10 MW

2.5 MW



Supervision + Astreinte
24/24 – 7/7



Pare-feu + système d'authentification



La sécurisation de l'information : un ensemble de moyens, une même finalité

Plateforme d'archivage

1. Réception



2. Vérification de la qualité des données reçues



3. Ajout d'informations (PID, empreintes, date d'archivage...)



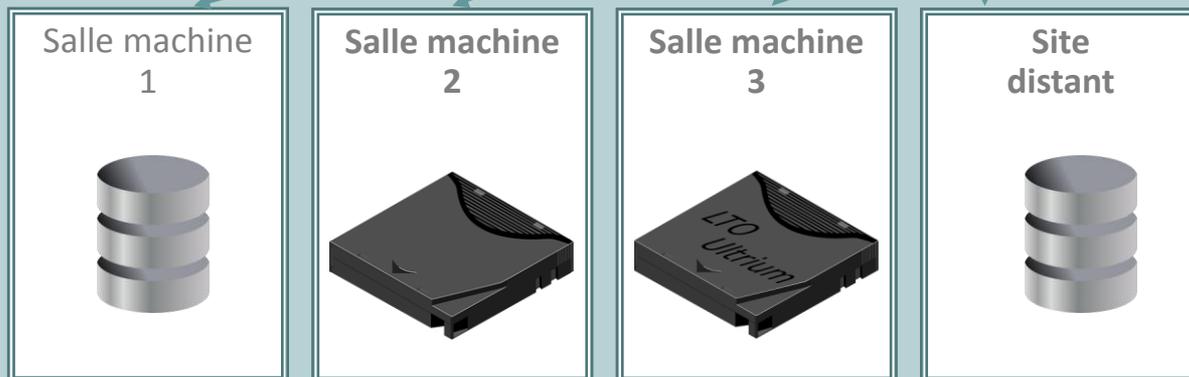
4. Traitements complémentaires (récupération d'informations associées, etc.)



6. Vérification périodique de tous les exemplaires archivés



5. Stockage de l'archive en plusieurs exemplaires





Des engagements de service côté CINES :

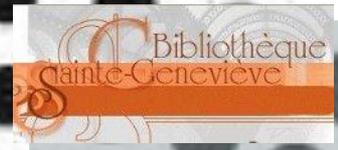
- Tous les niveaux de service liés à un Centre Informatique National
- L'accompagnement de l'établissement pendant la phase de mise en place du projet : participation d'un archiviste et d'experts ;
- L'assistance de niveau 1 et 2 (incidents, support, suivi) pendant l'exploitation du projet.

Des pré-requis côté service versant :

- Stratégie pour un projet de préservation de documents numériques (et non pas une sauvegarde ou un stockage sécurisé de données) ;
- Souhait de conserver des documents en lien avec le monde de l'ESR répondant à certains critères qualité (métadonnées, formats de fichiers HD, etc.).



Nos partenaires actuels





2 niveaux de services

- Niveau 1 (accès fréquent) : copies sur disques et cartouches (une > 300 km)
- Niveau 2 (accès occasionnel) : copies uniquement sur cartouches (une > 300 km)

A ce jour seul le niveau 1 est opérationnel ; niveau 2 disponible en 2014

Tarifs internes

• Etablissements publics dépendant du MESR

Fonctionnement

Tarifs partenaires

• Autres établissements publics de recherche

Fonctionnement

Amortissements

Tarifs externes

• Etablissements privés de recherche

Fonctionnement

Amortissements

Personnels

Niveau 1 (tarifs 2013) :

5000€ TTC / an / To
archivé

6500€ TTC / an / To
archivé.

13 000€ TTC / an / To
archivé.

Niveau 2 en cours d'évaluation

→ Tarifs dégressifs au-delà de 10To (25%) et de 50To (50%)



Le CINES est un acteur reconnu du domaine de la préservation numérique

- Mission nationale confiée par le MESR :
 - centre officiel d'archivage des thèses électroniques,
 - mise en œuvre de l'appel à projets d'archivage.
- Au sein de l'initiative BSN : définition au niveau national d'une stratégie pour l'archivage pérenne des documents électroniques produits par la communauté Enseignement Supérieur - Recherche ;
- Nombreuses sollicitations émanant d'universités, de bibliothèques ou de laboratoires, pour divers services dans ce domaine.



Objectifs 2013-2014 :

- La poursuite et le développement de tous ces projets, avec un focus particulier sur les « données scientifiques » ;
- La mutualisation des moyens via des partenariats entre les acteurs publics :
 - Infrastructures de stockage : 
 - Etudes/veille technologique : 
- La consolidation et la professionnalisation du service, des relations de confiance avec les institutions ;
- La poursuite de la démarche de certification du service d'archives.



Questions & Réponses

rouchon@cines.fr

- Phase 1 : Développement interne d'une première plate-forme pour valider les services attendus sur les 2 projets pilotes – premiers tests d'archivage des thèses en Mars 2007, déploiement en production fin Juin 2007

Protocole transfert	Soumission de SIP sur serveur de transfert via SFTP
Langage	Composants des serveurs développés en Java
Base de données	Base de données PostgreSQL
	Utilisation de bibliothèques OpenSource (module d'analyse de fichiers, calcul des empreintes numériques, etc.)
	Déclenchement des routines (accusé de réception, vérification, archivage) à intervalles réguliers – traitement par lots
Interface utilisateur	Administration en mode web

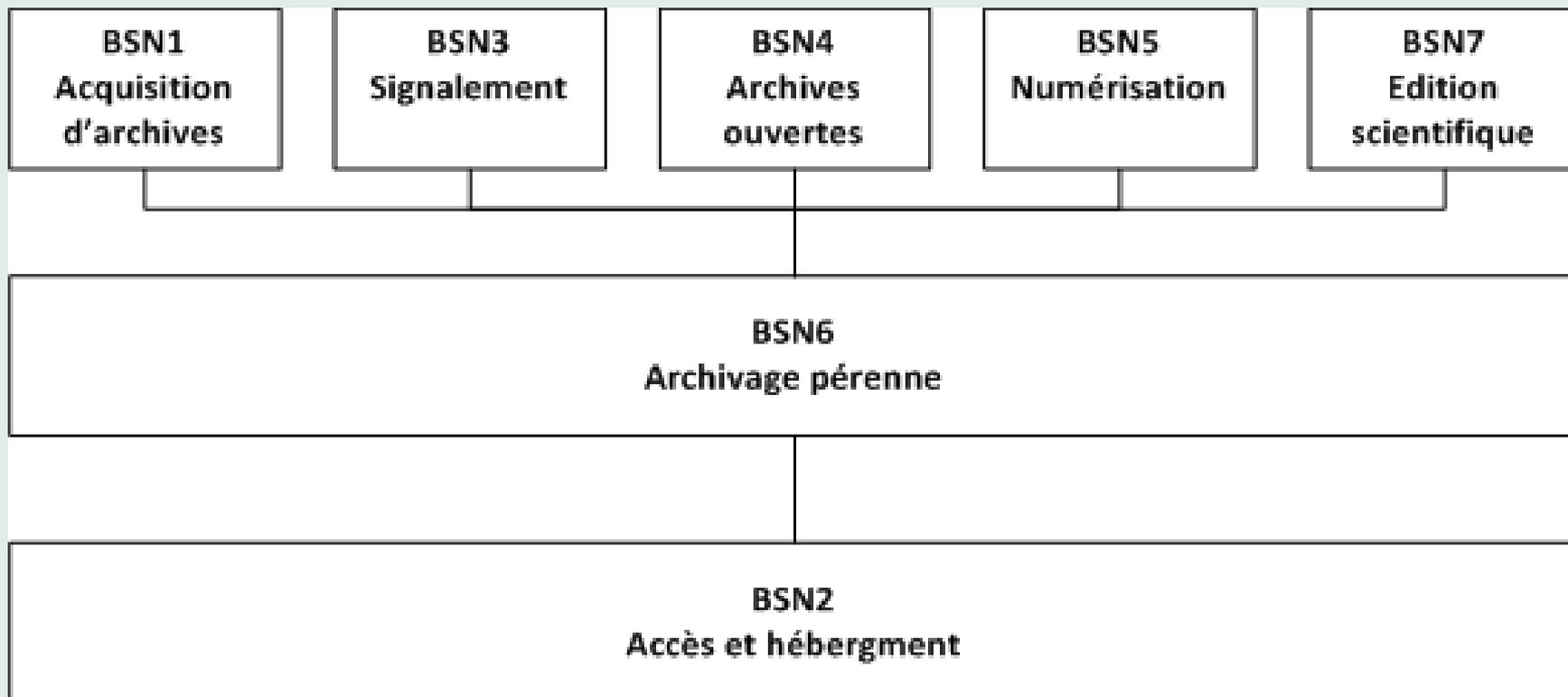
- Phase 2 : Appel d'offres en cours pour l'acquisition fin 2007 d'une plate-forme de stockage capable de gérer de larges volumes (40 To)

Plusieurs types d'objets numériques ont besoin d'être conservés :

- la production des établissements : les thèses, les documents pédagogiques, la production scientifique, et les archives administratives ;
- les données issues d'opérations de numérisation : documents et thèses analogiques numérisés, collections patrimoniales comme sources de la recherche, etc. ;
- les ressources acquises auprès d'autres producteurs : édition électronique, livre numérique, etc.
- les données brutes de la recherche : en dehors du périmètre de la BSN.

Le champ de réflexion du groupe BSN6 :

- Comment BSN peut-elle devenir le lieu visible d'une politique nationale portée au niveau interministériel et facilitant des coopérations internationales, notamment européennes ?
- Quel est le coût à venir de l'archivage pérenne ?



- Extension de la mission de conservation du CINES à l'ensemble de la production numérique susceptible d'intégrer la BSN ;
- Mise en place d'un label BSN validant l'intérêt scientifique et la qualité technique d'un corpus numérique en contrepartie duquel les services d'archivage et d'accès seraient proposés, semblable au modèle en place pour l'attribution d'heures de calcul intensif ;
- Mise en place d'un financement direct par l'Etat des services proposés dans le cadre du label BSN ;
- Soutien du MESR et participation de la BSN pour représenter la communauté ESR au sein d'un groupement indépendant regroupant les acteurs publics nationaux de l'archivage pérenne tel que préconisé dans le rapport Racine, le rapport Quenet.
- Promouvoir les collaborations entre acteurs lorsque les corpus impliquent plusieurs communautés – ex. de l'IRHT (BU/BM)