



les missions d'archivage pérenne du CINES

- le CINES est chargé de la conservation à long terme des documents numériques du patrimoine scientifique Université-Recherche
- 2 projets pilotes
 - les thèses électroniques (environ 10000 thèses / an = 100 Go / an)
 - les revues SHS du portail Persée (environ 4000 revues / an = 10 To / an)
- ... et d'autres projets à suivre
 - Elsevier : abonnements « interrompus »
 - Publications du Centre d'édition numérique scientifique (CENS) du CNRS
 - Fonds de documents vidéo ViSA de l'INRP
- besoin d'une solution générique

Groupe PIN 11/10/2006
(M. Auffret, Ph. Prat)



le projet PAC (plate-forme d'archivage du Cines)

- début : printemps 2004
- objectif : doter le CINES d'une plate-forme et d'un service d'archivage numérique pérenne apte à répondre à la majorité des besoins : éviter les approches au cas par cas, minimiser le spécifique
- l'équipe
 - 3 ingénieurs à mi-temps
 - 1 archiviste depuis le printemps 2006

11/10/2006 CINES
(M. Auffret, Ph. Prat)



Projet PAC : quelques étapes significatives

- début 2005 : il n'existe pas de solution « sur étagère »
- été 2005 (expertise Ourouk) : attention au manque de compétence archivistique
- les travaux réalisés à ce jour
 - identification d'un noyau de métadonnées
 - sélection d'un ensemble de formats « durables »
 - définition (grosse maille) des fonctionnalités attendues
 - développement interne d'une plate-forme provisoire

11/10/2006 CINES
(M. Auffret, Ph. Prat)



Projet PAC : état des lieux (octobre 2006)

- Appel d'offres envisagé pour l'acquisition d'une plate-forme complète : logiciel, serveur et stockage
- En parallèle : veille sur l'avancement des projets menés par le CNES, par la BnF et par la DAF
- Développement interne d'une plate-forme provisoire pour valider les services attendus sur les 2 projets pilotes : premiers tests d'archivage des thèses prévus à partir de fin octobre
- Référentiels privilégiés
 - le modèle OAIS
 - le standard d'échange DAF-DGME
 - les travaux du groupe PIN

11/10/2006 CINES
(M. Auffret, Ph. Prat)



PAC : une plateforme d'archivage pérenne

- Développement en interne
- Langage java
- JHOVE outil open source utilisé en tant que module d'analyse des fichiers
- Interface : shell, SFTP, web en phase de test
- Sauvegarde : TSM baie HDS puis sur bandes (Robothèque STORAGETEK STK 9310)

11/10/2006 CINES
(M. Auffret, Ph. Prat)



PAC : 3 serveurs logiques actuellement

- **Versement**
 - Réception des SIP
 - Vérification des SIP
 - Création et transfert des AIP sur le serveur de stockage
- **Stockage**
 - Sur disque
 - Sauvegarde TSM sur baie HDS, puis sur Robothèque
- **D'accès**
 - Communication d'une archive grâce à son identifiant

11/10/2006 CINES
(M. Auffret, Ph. Prat)



PAC : Organisation physique

pac/projet
pac/projet/PROJET_X

pac/projet/PROJET_X/VERS

pac/projet/PROJET_X/VERS/SIP_ID
pac/projet/PROJET_X/VERS/SIP_ID/DEPOT
pac/projet/PROJET_X/VERS/SIP_ID/sip.xml
pac/projet/PROJET_X/VERS/SIP_ID

pac/storage
pac/storage/STOCKAGE_X/ID

pac/projet/PROJET_X/ACCESS

11/10/2006 CINES
(M. Auffret, Ph. Prat)



PAC : 3 étapes du serveur de versement

- **Réception**
 - détection d'un nouveau SIP déposé via SFTP.
 - envoi d'un accusé de réception par mail au service versant
- **Vérifications**
 - structure de base
 - conformité sip.xml/sip.xsd)
 - correspondance bi-univoque sip.xml/document
 - utilisation de JHOVE pour détecter le format
- **Transfert au serveur de stockage**
 - création du fichier aip.xml (cf aip.xsd)
 - création de l'identifiant de l'archive
 - copie des fichiers sur le serveur de stockage
 - Envoi du certificat d'archivage.

11/10/2006 CINES
(M. Auffret, Ph. Prat)