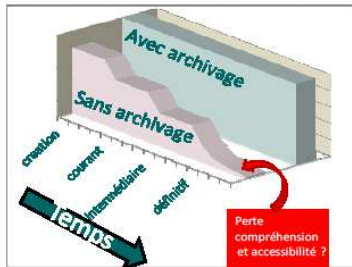


La Plateforme d'Archivage du CINES

PAC : La problématique de l'archivage numérique pérenne

- Archivage pérenne = Mise en place de procédures d'assurance qualité pour atténuer l'impact des risques lorsqu'ils se réalisent



4 risques inéluctables :

- Connaissance perdue du contenu des fichiers ;
- Format de fichier inconnu ;
- Support physique détérioré ;
- Logiciel ou matériel de lecture disparu.

L'archivage pérenne des documents électroniques consiste à conserver le document et l'information qu'il contient :

- ✓ Dans son aspect physique comme dans son aspect intellectuel,
- ✓ Sur le très long terme soit 30 ans et au-delà,
- ✓ De manière à ce qu'il soit en permanence accessible et compréhensible.

PAC : La problématique de l'archivage numérique pérenne

Contrainte	Solutions / Qualités
Perte de la connaissance du contenu	<ul style="list-style-type: none"> Utilisation de métadonnées, adoption d'un standard
Format de fichier inconnu	<ul style="list-style-type: none"> Privilégier les formats durables Identification, contrôle & validation des formats en entrée. Veille, migration logique (conversion de formats)
Support physique détérioré	<ul style="list-style-type: none"> Stockage: Veille, Capacité à conserver le train de bits constituant les fichiers (changement de support), Sécurisation.
Logiciel /matériel de lecture disparu	<ul style="list-style-type: none"> Veille technologique et anticipation

PAC : Le choix et le contrôle des formats

Pour permettre le contrôle de la qualité d'un fichier, celui-ci doit être dans un format identifié et vérifiable :

- Format publié ; ex. WAVE, SVG ;
- Format largement utilisé ; ex. XML, MPEG4 ;
- Format normalisé si possible ; ex. PDF (ISO 32000-1:2008), PNG (ISO 15948:2004).

Pour pouvoir être lisibles dans le temps, et convertibles, les fichiers doivent respecter à la lettre les **spécifications** de leur format.

Les outils libres Jhove, ImageMagick, DROID, ODF Validator permettent une identification, validation et caractérisation des **formats en entrée de PAC**.

Type	Format
Texte	HTML, PDF, PDF/A, TXT, XML, ODT
Image	GIF, JPEG, TIFF, GEOTIFF, PNG, SVG, JPEG2000
Audio	WAV, AIFF, AAC, VORBIS
Vidéo	MPEG4, THEORA, MKV



PAC : La qualité du stockage



La Plateforme d'archivage pérenne au CINES – PAC v2.0

- Capacité actuelle 40 To (25 To)
- En exploitation depuis Mai 2008
- Logiciel d'archivage** Arcsys (Infotel)
- Serveurs applicatifs et baie de stockage** SUN – Oracle
- Librairie de bandes IBM (2 x 1Po)**

La qualité du stockage garantit la conservation du train de bits composant les fichiers de données

- ✓ **Copies multiples:** 2 copies sur disques, 2 copies sur bandes, 1 copie sur site distant (CC-IN2P3 de Lyon)
- ✓ **Indépendance des supports de copies** (mélange disques/bandes, localisation géographique)
- ✓ **Audit fréquent de l'intégrité des copies**

Tous les projets d'archives à long terme partagent la même plateforme

- Mutualisation de l'infrastructure matérielle d'archivage ;
- Protocole de versement générique ;
- Diminution des coûts de mise en place et d'exploitation.