



Conservation numérique (Digital preservation)

Elizabeth CHERHAL-CLEVERLY
Cellule MathDoc
Rencontres RNB M 1-5 Octobre 2007



Conservation numérique (Digital preservation) Contexte

- Des siècles d'expérience dans la conservation des collections sur papier
- Extrême jeunesse des collections numériques
- Malaise sur l'évolution des rôles
- Volatilité de l'information en ligne

Contexte

Risques et solutions

Idées et modèles

Projets

Conclusion



Conservation numérique (Digital preservation) Définition, exemples

- Collection numérique:
 - Ensemble de documents (fichiers) associé à des métadonnées généralement rangées dans une base de données
- Exemples :
 - Gallica, NUMDAM, ArXiv, HAL
 - JSTOR, Springer Link, etc.

Problématique
Risques et solutions
Idées et modèles
Projets
Conclusion



Problématique
Risques et solutions
Idées et modèles
Projets
Conclusion

Conservation numérique (Digital preservation) petit rappel utile

- "Un document numérique n'est pas accessible à un utilisateur humain" [1]
- De nombreux intermédiaires
 - Support
 - Lecteur
 - Système d'exploitation
 - Logiciel applicatif
 - Péipherals
 - Mode d'emploi
 - (réseau)
- Mais, de fait il est souvent plus accessible à un instant T qu'un document papier



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Les risques connus

○ Pannes :

- Panne support (media failure) :
 - impossibilité de lire le support physique (disque, CD, DVD...)
- Panne machine (hardware failure):
 - Panne de la machine sur laquelle réside la collection
- Panne logicielle (software failure)
 - Bugs potentiels dans le logiciel de consultation de la collection



Problématique
Risques
et solutions
Idées
et modèles
Projets
Conclusion

Conservation numérique (Digital preservation)

Les risques connus

○ Obsolescences :

- Des supports et matériels (media and hardware) : tout support et tout matériel devient obsolète. Risque de non communication entre différents composants d'un système (bandes DAT, DLT...)
- Des logiciels (software) : impossibilité de décoder l'information, formats de fichiers, compatibilité des versions, mise à jour logicielle ...



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Les risques connus

- Risques naturels :
 - Incendie, tremblement de terre, inondation...
- Erreurs humaines : sur la collection numérique, sur le système ou les applications sous-jacentes...
- Attaques informatiques
- ...



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Les risques connus

- Problèmes économiques et organisationnels:
 - Impossibilité de maintenir le système: personnels, coûts de fonctionnement..
 - "restructuration" de l'entreprise ou équipe...



Problématique

Risques
et **solutions**

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Les remèdes aux risques

○ RéPLICATION :

- Copie sans perte d'information:
 - Sauvegardes classiques
 - Sauvegardes déportées
 - Miroirs
- Copies sélectives :
 - LOCKSS (chaque participant archive ce qui l'intéresse)

Conservation numérique (Digital preservation)

Les remèdes aux risques

○ Migration :

- Des supports : bandes DAT à DLT, de CD à DVD...
- Des formats de fichier: si un format devient obsolète, passer tous les fichiers de ce format à un autre plus récent.
- Des versions du système et des logiciels

Problématique

Risques
et **solutions**

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Les remèdes aux risques

○ Transparency :

- Il n'existe pas de "format pérenne", mais ...
 - Formats largement utilisés (open source ou spécifications connues)
- Il n'existe pas de "logiciel pérenne", mais ...
 - Logiciels open source et largement utilisés

Problématique

Risques
et **solutions**

Idées
et modèles

Projets

Conclusion



Problématique

Risques
et **solutions**

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Les remèdes aux risques

○ Diversité :

- Des supports
- Des formats de fichier (tiff, pdf, source...)
- Géographique
- Des versions (ne pas être trop lié à telle ou telle version du système ou du logiciel)
- Economique

Conservation numérique (Digital preservation)

Les remèdes aux risques

- Audit (vérification que cela marche encore !) :
- Système consulté fréquemment : commentaires des utilisateurs
- Système de préservation : peu utilisé, doit avoir des procédures de vérification
 - Audit lors de procédures de réPLICATION ou d'ingestion dans d'autres systèmes

Problématique

Risques et **solutions**

Idées et modèles

Projets

Conclusion



Problématique

Risques
et **solutions**

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation) un remède théorique

○ L'émulation

- Conserver le document "original" et tout ce qui est nécessaire à sa visualisation (support, applicatif, système ...)
- Mais... l'émulateur doit lui-même être migré



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation) système d'archivage "pérenne"

- Open Archival Information System
- OAIS : un modèle conceptuel
- Système différent de la réPLICATION
- Les "objets numériques" sont insérés avec (une partie de) leur contexte
- Une base théorique pour les ébauches de systèmes actuels.

Problématique

Risques et solutions

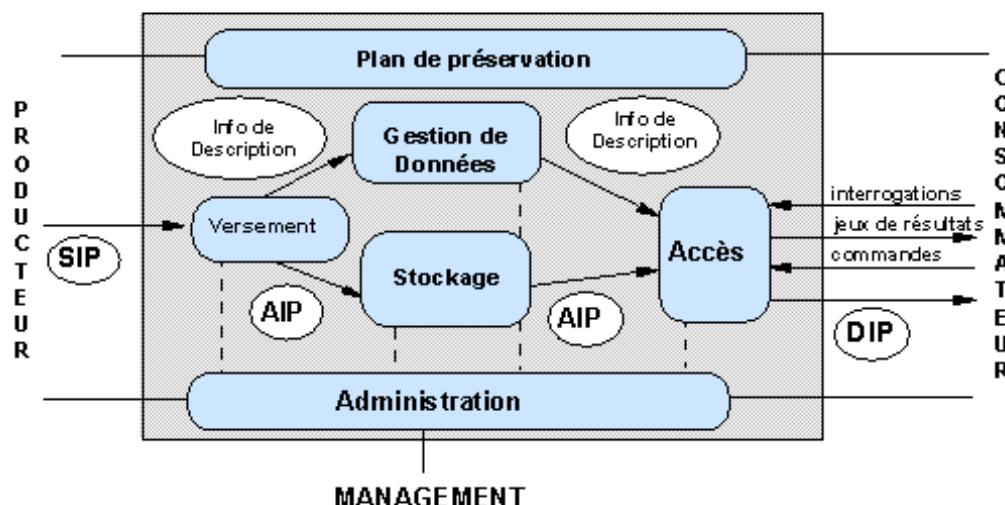
Idées et **modèles**

Projets

Conclusion

Conservation numérique (Digital preservation) système d'archivage "pérenne"

Le modèle fonctionnel OAIS



SIP

Submission information package
Paquet de versement

AIP

Archival Information Package
Paquet d'archivage

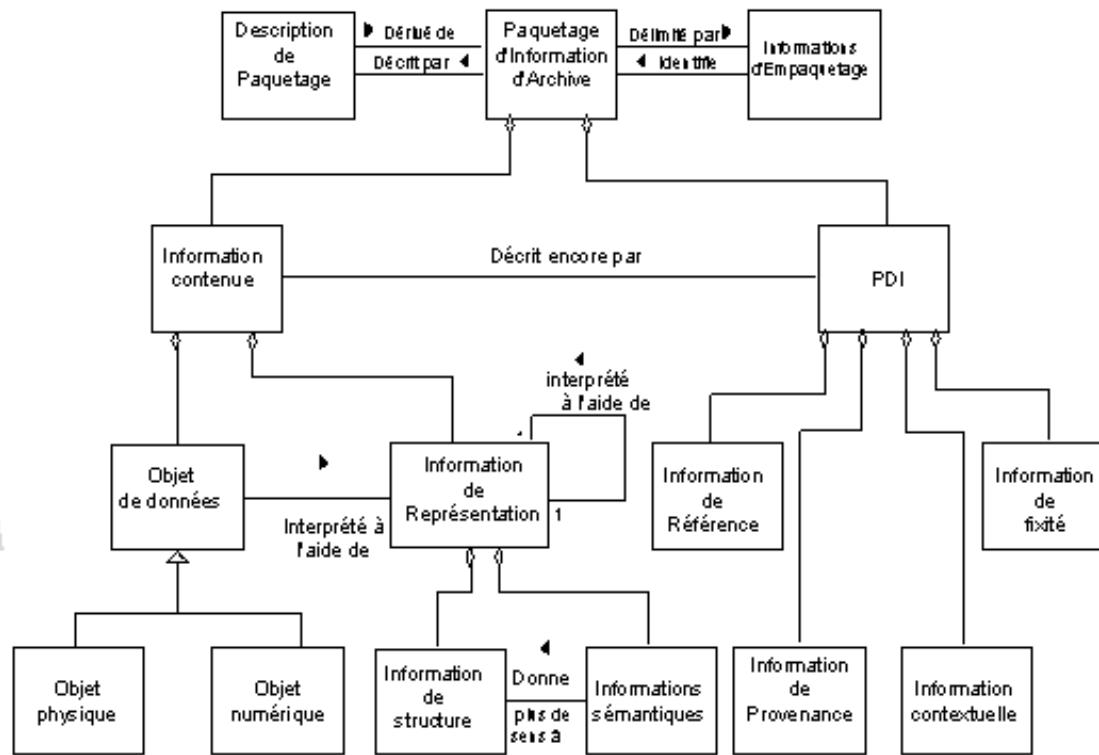
DIP

Diffusion Information Package
Paquet de diffusion



Conservation numérique (Digital preservation) système d'archivage "pérenne"

Le modèle d'information OAIS



Conservation numérique (Digital preservation)

Métadonnées pour l'archivage

- METS : metadata encoding and transmission standard
 - Une implémentation de référence pour OAIS.
 - Un "objet METS" va comprendre:
 - La ressource numérique
 - Les métadonnées de structure, administratives et descriptives

Problématique

Risques
et solutions

Idées
et **modèles**

Projets

Conclusion



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation) Quelques projets et systèmes

- Distinguons les "archives" destinées à la consultation quotidienne, (JSTOR, NUMDAM...) des systèmes d'archivage à plus long terme.
- Notons que certains logiciels de gestion d'entrepôt de document sont basés sur le modèle OAIS (D-Space, Fedora)



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Quelques projets et systèmes

○ LOCKSS (lots of copies keep stuff safe) : un super système de réPLICATION.

- LOCKSS is open source software designed to ensure that important scholarly assets remain available in a distributed, self-repairing, robust, digital preservation system. We work with libraries and electronic publishers to preserve materials published online for long term access.
- Entre bibliothèques et éditeurs
- Entre bibliothèques et bibliothèques



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Quelques projets et systèmes

○ LOCKSS - idées de base :

- La menace majeure pour la conservation numérique est économique: personne n'a suffisamment de fonds pour mettre en place l'archive parfaite
- Archive vivante // archive sombre (dark archive)
- Le contenu est obtenu par des moteurs de recherche (pull // push).
- Un système "d'audit" intégré assure la cohérence.
- Des bibliothèques qui veulent préserver le même contenu peuvent coopérer



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation) Quelques projets et systèmes

- LOCKSS est fait pour tourner seul sur une machine (LOCKSS box)
- Collection des nouveautés sur le site éditeur (après accord)
- Comparaison de contenu avec celles des autres "boîtes LOCKSS"
- Interface d'administration via navigateur web



Problématique

Risques et solutions

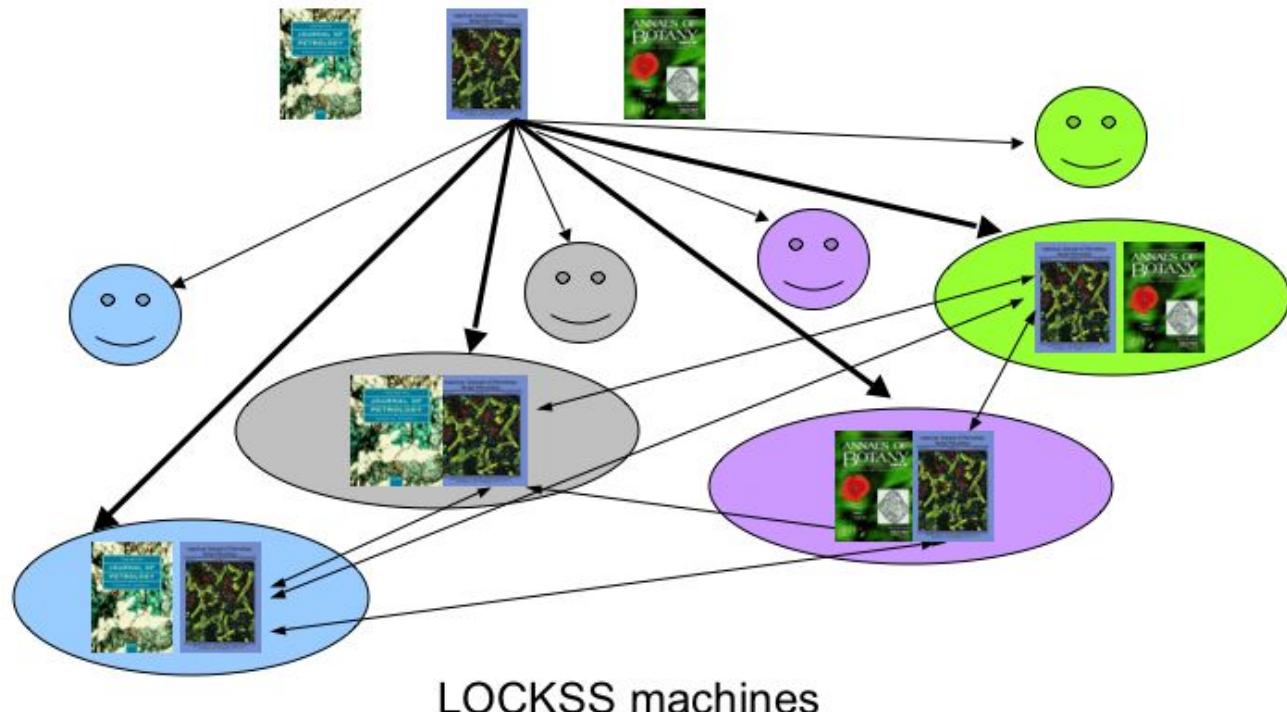
Idées et modèles

Projets

Conclusion

Schéma LOCKSS

Collection via Web Crawler





Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation) Quelques projets et systèmes

○ Portico :

- To preserve scholarly literature published in electronic form and to ensure that these materials remain available to future generations of scholars, researchers, and students.
- Orienté préservation à long terme
- Editeurs et "non profit organisations"
- Orienté journaux scientifiques
- Conservation **et** conversion de fichiers sources
- Méthodologie de préservation: migration
- Format d'archive basé sur [Journal Archiving and Interchange DTD](#)
- Contient > 200 000 articles (déc 2006)



Problématique

Risques et solutions

Idées et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Quelques projets et systèmes

- **Kopal** (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen : Constitution Coopérative d'Archives d'Informations numériques à long terme) :
 - Soutenu par le ministère de l'éducation et de la recherche.
 - système d'archivage pérenne à long terme pour des données électroniques.
 - basée sur DIAS (Digital Information and Archiving System : Information Numérique et Système d'Archivage) développé conjointement par IBM et la Bibliothèque Nationale des Pays-Bas à La Hague.
 - Utilise OAIS et METS
 - Partenaires actuels: Deutsche Bibliothek et SUB Göttingen



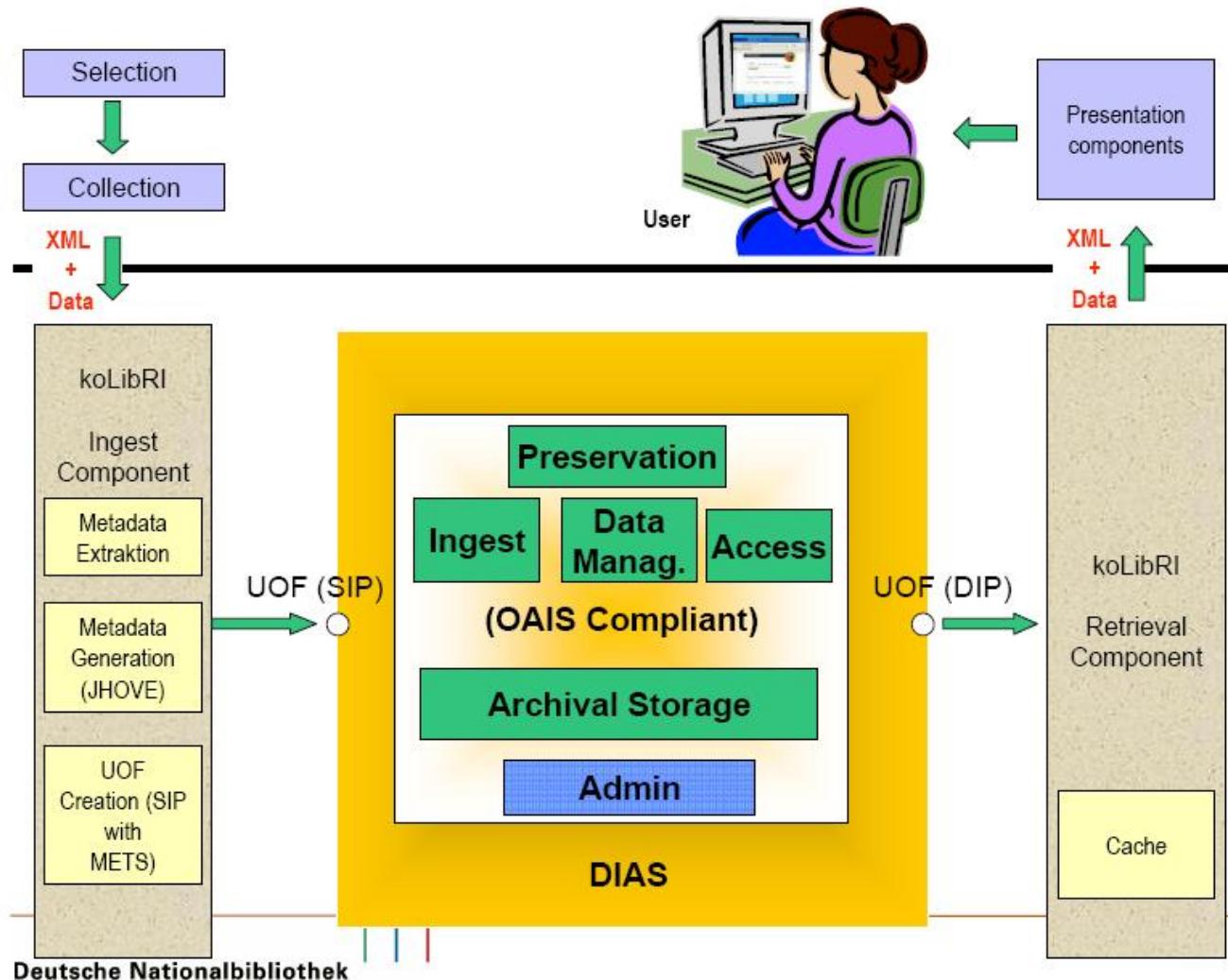
Problématique

Risques et solutions

Idées et modèles

Projets

Conclusion





Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Kopal: chiffres

- Début en Août 2006
- Traitement de 40.000 thèses
- Tout est correctement entré dans DIAS
- Le logiciel koLibRI (kopal Library for Retrieval and Ingest) fonctionne selon les spécifications
- Collections à venir :
 - Journaux électroniques de Springer
 - Echantillons de CD-ROMs et DVD (image ISO)
 - Echantillons de documents numérisés de la collection DNB



Problématique

Risques et solutions

Idées et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Quelques projets et systèmes

- Et en France ?
- Groupes de travail :
 - Le groupe de travail PIN (pérennisation de l'information numérique) a été créé en 2000 à l'occasion de la diffusion en France de la norme OAIS. Il regroupe des représentants de grandes institutions publiques (CNES, BnF, CEA, etc.) et d'entreprises qui mettent en commun leurs réflexions et leurs expérimentations concernant l'archivage de données scientifiques, patrimoniales ou opérationnelles. PIN s'efforce de développer une culture commune aux différents acteurs de l'archivage.
 - LE COST (Comité scientifique et technique du Comité stratégique inter-établissements pour la plate forme nationale d'archives ouvertes) a formé un groupe de travail sur cette question
- La suite à l'exposé suivant ...



Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation) citations en conclusion

- Chris Rusbridge, directeur du "digital curation centre" (UK) :
 - Preservation is expensive, but we are used to it and accept it as part of the cost of cultured and educated life.
 - Make your dispositions on the timescale you can foresee and for which you have funding. Preserve your objects to the best of your ability, and hand them on to your successor in good order after your lap of the relay.
- Il est difficile de projeter dans le futur la situation technologique actuelle.

Conservation numérique (Digital preservation)

Références

1. Rossi, Christian.- *De la diffusion à la conservation des documents numériques.* à paraître aux Cahiers GUTenberg (n° 49)
2. Rosenthal D.S et al.- *Requirements for Digital Preservation Systems : a bottom up approach.*
<http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>
3. Didacticiel, *Gestion des la conservation des collections numériques : stratégies à court terme pour contrer des problèmes à long terme*
<http://www.library.cornell.edu/iris/tutorial/dpm-french/index.html>
4. Rusbridge, Chris .- *Excuse Me... Some Digital Preservation Fallacies ?*
<http://www.ariadne.ac.uk/issue46/rusbridge/>

Conservation numérique (Digital preservation)

Références (liens)

- OAIS :
http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf
- LOCKSS : <http://www.lockss.org/lockss/Home>
- Portico : <http://www.portico.org/>
- Kopal :
<http://kopal.langzeitarchivierung.de/index.php.en>
- PIN : <http://vds.cnes.fr/pin/>
- COST :
<http://www.revues.org/cost/index.php?title=GR4 - Archivage p%C3%A9renne&redirect=no>