



HAL
open science

La documentation électronique - Archivage pérenne et conservation numérique

Elizabeth Cherhal

► **To cite this version:**

Elizabeth Cherhal. La documentation électronique - Archivage pérenne et conservation numérique. École thématique. Hammamet, Tunisie. 2008, pp.33. hal-03722146

HAL Id: hal-03722146

<https://hal.univ-grenoble-alpes.fr/hal-03722146v1>

Submitted on 13 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La Documentation électronique

- Plan ->
- **Définitions**
- Production
 - Formats
 - Spécificités
- La trouver
 - Sources
 - Moteurs
- Bibliothèque numérique
 - Définitions
 - Métadonnées
- **Archivage pérenne ->**

- Plan des cours : = 3 séances
 - 1) **Documentation électronique**
 - Définitions
 - Production
 - Formats de fichier
 - Spécificités mathématiques: LaTeX, mathML
 - La trouver ?
 - Sources
 - Moteurs de recherche
 - 2) **Bibliothèque numérique**
 - Définitions
 - Métadonnées
 - Technique
 - 3) **L'archivage pérenne ou conservation numérique**
 - **Problématique**
 - **Risques et solutions**
 - **Idées et modèles**
 - **Projets**

La documentation électronique

Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion

Conservation numérique (Digital preservation)

Conservation numérique (Digital preservation)

Contexte

Problématique

->

Risques
et solutions

Idées
et modèles

Projets

Conclusion

- 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

Conservation numérique (Digital preservation)

Définition, exemples

Problématique

->

Risques
et solutions

Idées
et modèles

Projets

Conclusion

- 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.
- Les problèmes de préservation sont **les mêmes** pour les éditeurs commerciaux que pour les bibliothèques numériques et archives ouvertes.
- Voir la [vidéo de kopal](#)

Conservation numérique (Digital preservation) petit rappel utile

Problématique

->

Risques
et solutions

Idées
et modèles

Projets

Conclusion

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

Conservation numérique (Digital preservation)

Les risques connus

Problématique

Risques ->
et solutions

Idées
et modèles

Projets

Conclusion

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

Conservation numérique (Digital preservation)

Les risques connus

Problématique

Risques ->
et solutions

Idées
et modèles

Projets

Conclusion

- 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 ...

Conservation numérique (Digital preservation)

Les risques connus

Problématique

Risques ->
et solutions

Idées
et modèles

Projets

Conclusion

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

Conservation numérique (Digital preservation)

Les risques connus

Problématique

Risques ->
et solutions

Idées
et modèles

Projets

Conclusion

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

Conservation numérique (Digital preservation)

Les remèdes aux risques

Problématique

Risques
et solutions ->

Idées
et modèles

Projets

Conclusion

- 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

Problématique

Risques
et solutions ->

Idées
et modèles

Projets

Conclusion

- 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

Conservation numérique (Digital preservation)

Les remèdes aux risques

Problématique

Risques
et solutions ->

Idées
et modèles

Projets

Conclusion

- Transparence :
 - 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

Conservation numérique (Digital preservation)

Les remèdes aux risques

Problématique

Risques
et solutions ->

Idées
et modèles

Projets

Conclusion

- Diversité :
 - Des supports
 - Des formats de fichiers (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

Problématique

Risques
et solutions ->

Idées
et modèles

Projets

Conclusion

- **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

Conservation numérique (Digital preservation) un remède théorique

Problématique

Risques
et solutions ->

Idées
et modèles

Projets

Conclusion

- 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

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

Problématique

Risques
et solutions

Idées
et **modèles** ->

Projets

Conclusion

- OAIS : un modèle
- 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.

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

Problématique

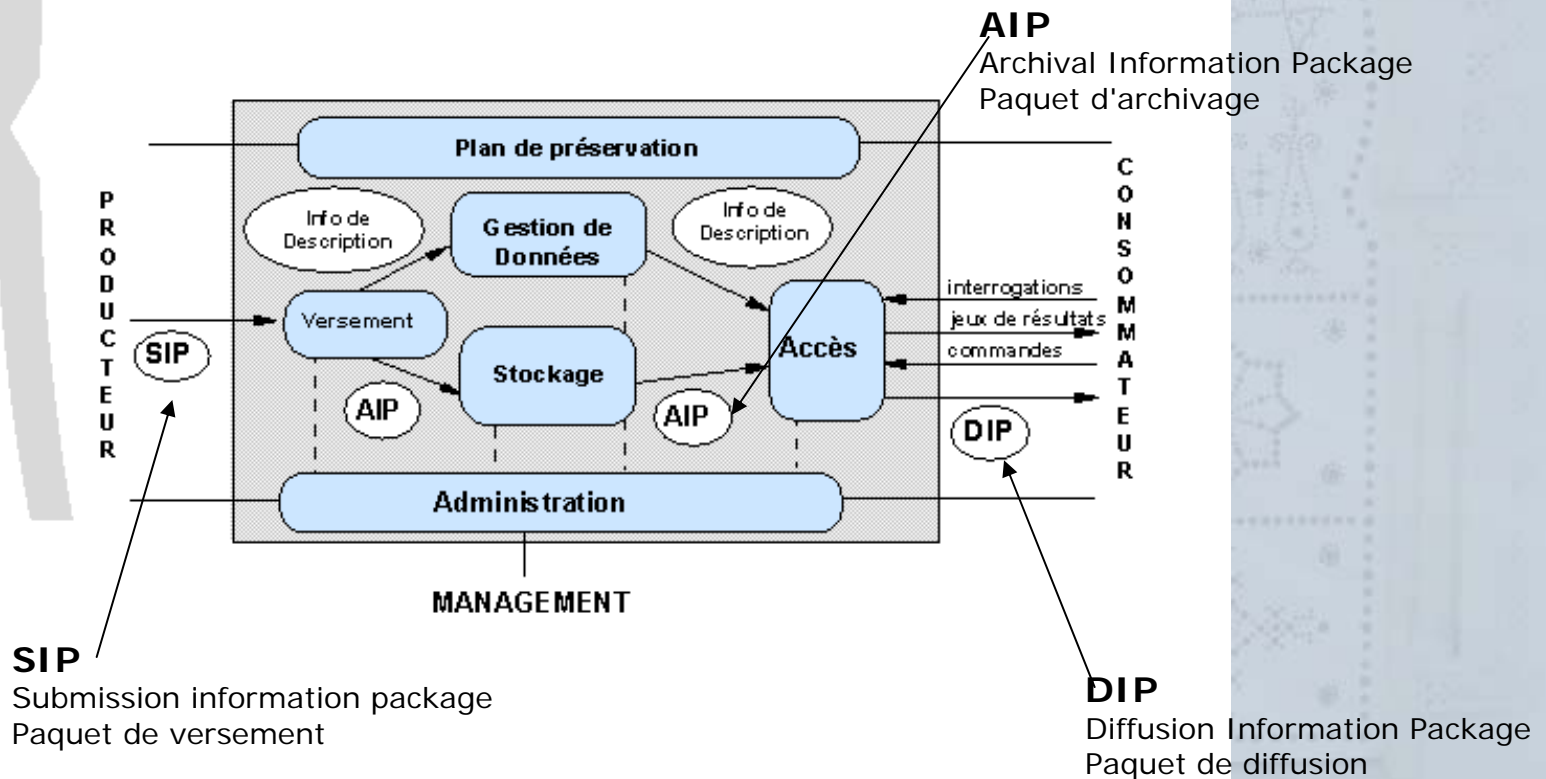
Risques
et solutions

Idées
et **modèles** ->

Projets

Conclusion

Le modèle fonctionnel OAIS



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

Problématique

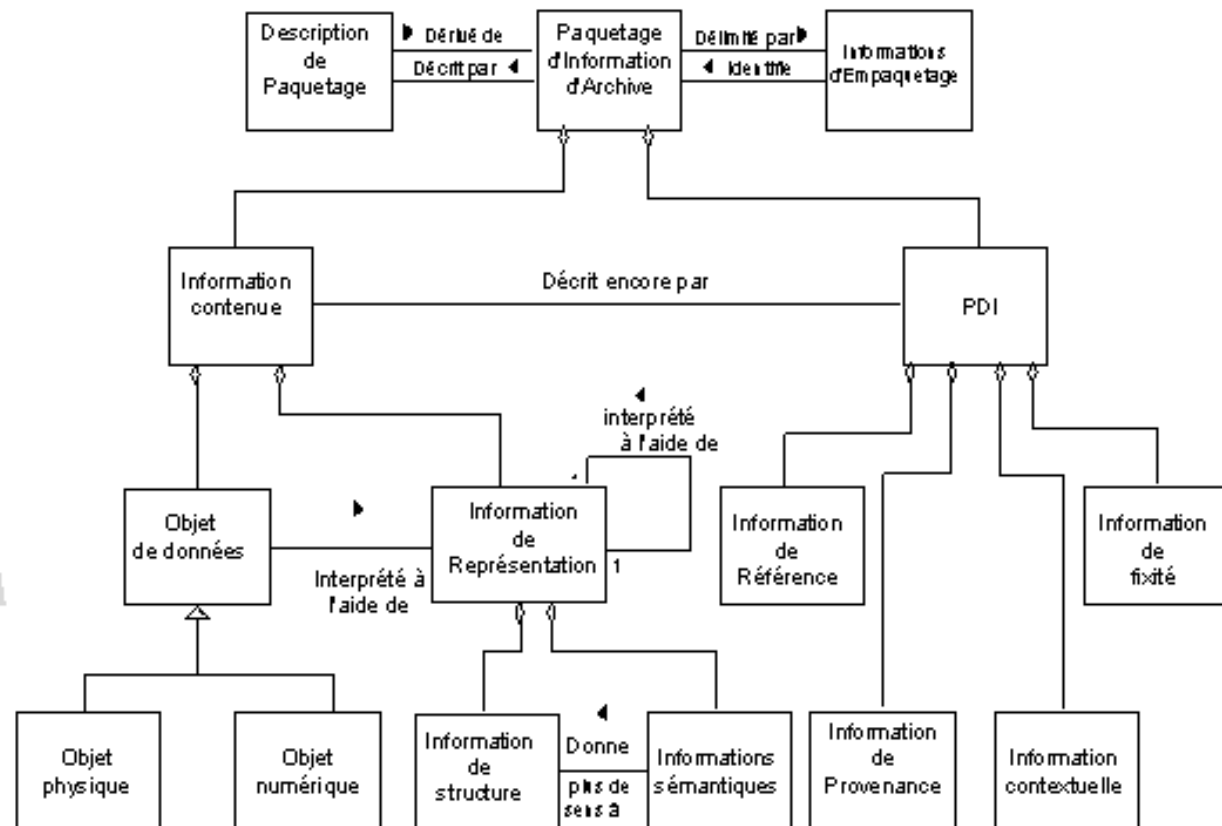
Risques
et solutions

Idées
et modèles ->

Projets

Conclusion

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 et les métadonnées de structure, administratives et descriptives

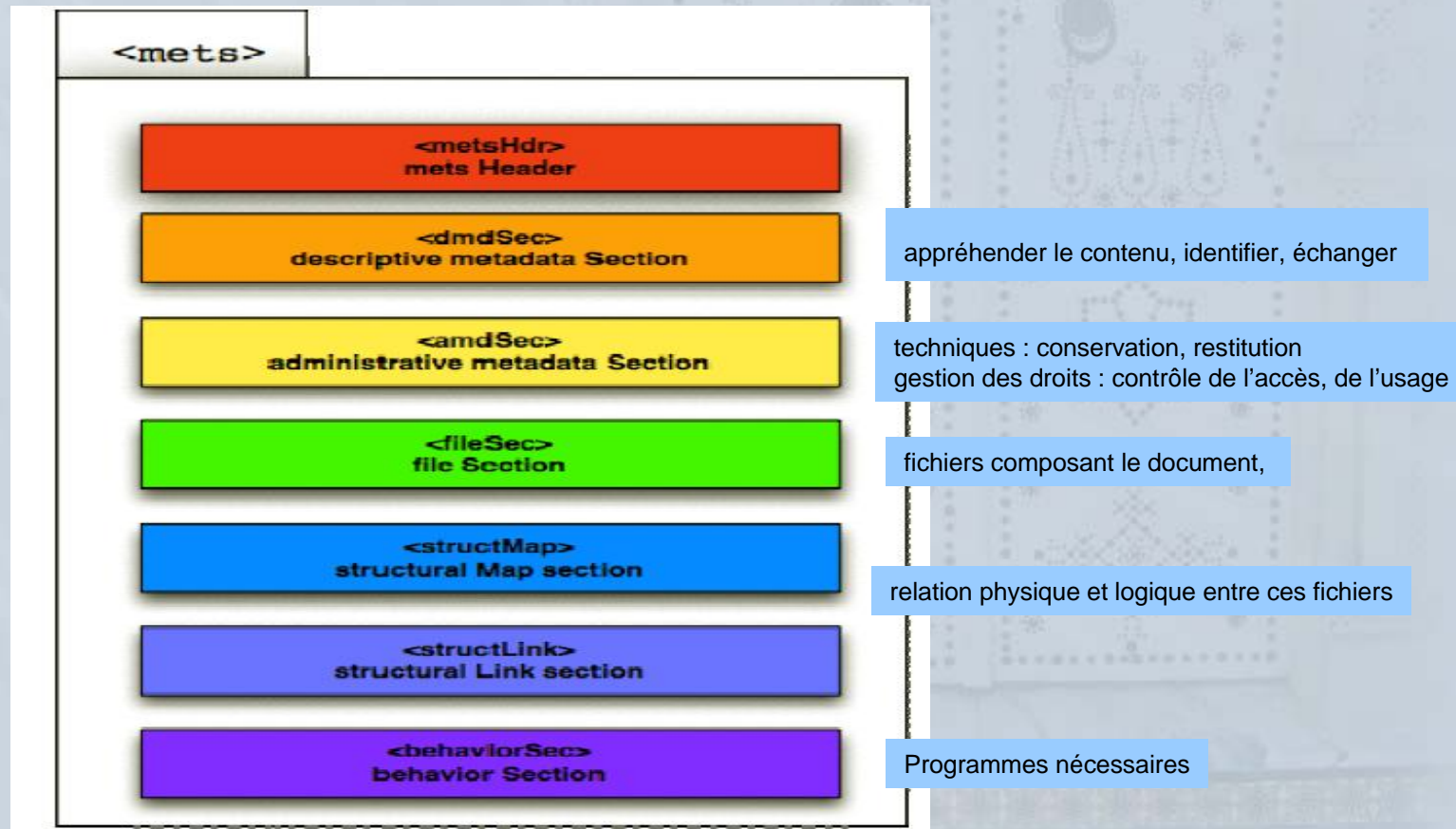
Problématique

Risques
et solutions

Idées
et modèles ->

Projets

Conclusion



Conservation numérique (Digital preservation) Quelques projets et systèmes

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- 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)

Conservation numérique (Digital preservation) Quelques projets et systèmes

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- 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

Conservation numérique (Digital preservation) Quelques projets et systèmes

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- 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

Conservation numérique (Digital preservation) Quelques projets et systèmes

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- 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 (Oct 2008)
 - 8,151,142 articles
 - 60 éditeurs participants
 - 7,967 titres
 - 469 bibliothèques participantes

Conservation numérique (Digital preservation) Quelques projets et systèmes

Problématique

Risques
et solutions

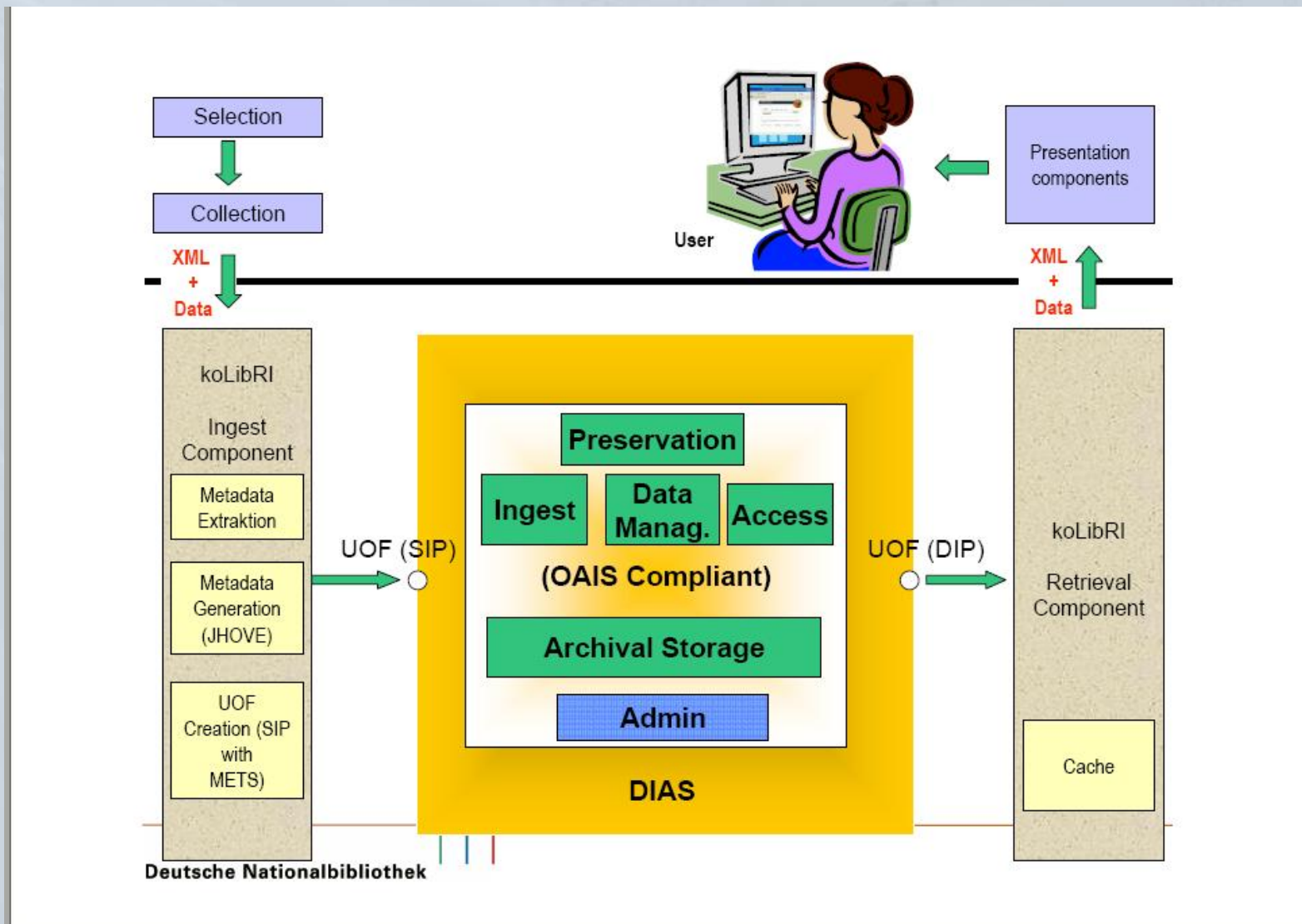
Idées
et modèles

Projets ->

Conclusion

- 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



Kopal: chiffres

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- 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
- La DNB utilise maintenant kopal en phase de production.

Conservation numérique (Digital preservation) Quelques projets et systèmes

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- 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 plateforme **PAC** (Plateforme d'Archivage au CINES), développé au CINES (Centre Informatique National de l'Enseignement Supérieur)
- **SPAR** = système de préservation et d'archivage réparti (projet de la BnF)
 - Devrait être opérationnel en 2010
 - Page officielle à la BnF
http://www.bnf.fr/pages/infopro/numerisation/num_spar.htm

Conservation numérique (Digital preservation)

Quelques projets et systèmes

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- **La plateforme PAC**
- Le système PAC a été conçu comme un ensemble de trois serveurs logiques, s'inspirant du modèle proposé par la norme ISO 14721 ([OAIS](#)).
 - un serveur de versement où le producteur des documents pourra transmettre ses archives.
 - un serveur de stockage où sont conservés les documents
 - un serveur d'accès où le producteur et les service demandeurs autorisés à consulter ses archives pourront rechercher et obtenir une archive.

Conservation numérique (Digital preservation)

Quelques projets et systèmes

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- **Système PAC**

- Prise de contact
- Établissement d'un protocole de transfert (identification des informations à pérenniser, volumétrie et formats choisis...)
- Transfert par réseau ou sur support amovible
- Contrôles
- Transfert sur serveur de stockage, renvoi d'un identifiant de l'archive
- La communication (récupération) **se fera** via le serveur d'accès.

Conservation numérique (Digital preservation)

Quelques projets et systèmes

Problématique

Risques
et solutions

Idées
et modèles

Projets ->

Conclusion

- Système PAC
- Utilisateurs :
 - Système STAR (ABES)
 - Bibliothèque numérique PERSEE (sciences sociales)
- Documentation
 - <http://www.cines.fr/spip.php?rubrique152>

Conservation numérique (Digital preservation) citations en conclusion

Problématique

Risques
et solutions

Idées
et modèles

Projets

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

Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion ->

1. Rossi, Christian.- *De la diffusion à la conservation des documents numériques*. Cahiers GUTenberg (n° 49)
<http://www.gutenberg.eu.org/publications/cahiers/r46-cahiers49/222-rossi.html>
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)

Problématique

Risques
et solutions

Idées
et modèles

Projets

Conclusion ->

- 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>
 - Vidéo d'explication :
http://kopal.langzeitarchivierung.de/downloads/kopal_Demonstrator_en.swf
- PIN : <http://vds.cnes.fr/pin/>
- COST : [http://www.revues.org/cost/index.php?title=GR4 -
_Archivage_p%C3%A9renne&redirect=no](http://www.revues.org/cost/index.php?title=GR4_-_Archivage_p%C3%A9renne&redirect=no)
- PAC : <http://www.cines.fr/spip.php?rubrique152>
- SPAR :
http://www.bnf.fr/pages/infopro/numerisation/num_spar.htm