

La conservation numérique



Enjeux, risques, solutions

Louise Fauduet
Bibliothèque nationale de France
Département de la Conservation

louise.fauduet@bnf.fr

Un enjeu patrimonial ?



- ❧ Quoi : collection(s) numérique(s), patrimoine(s) numérique(s)
- ❧ Pourquoi : conservation, préservation
- ❧ Comment : un problème « technique » ?



Quoi ?



Collection(s) numérique(s), patrimoine(s) numérique(s)

Du numérisé



∞ Pour le ministère, c'est là le « patrimoine numérique »

∞ <http://www.numerique.culture.fr/pub-fr/index.html>

∞ Déjà plusieurs « strates » de numérisation dans les bibliothèques...

Qu'a-t-on numérisé ?



- ❧ Du papier : livres, images, presse, etc.
- ❧ Des objets : maquettes, statues, etc.
- ❧ De l'audiovisuel : cassettes audio, VHS, disquettes, CD, CD-ROM, etc.
 - ❧ qui nécessite un outil de lecture
 - ❧ qui est parfois déjà numérique
- ❧ Les originaux sont-ils rares, précieux, fragiles ?
- ❧ Les versions numériques sont-elles fidèles à l'original ?

Dans quel cadre ?



- ❧ Quel est le statut des documents numériques produits ?
 - ❧ Hors droits, sous droits ?
 - ❧ Numérisés « à des fins de conservation » ou pour « préserver les conditions de sa consultation » au sens du Code de la propriété intellectuelle (article L122-5) ?
 - ❧ Dans le cadre d'un contrat passé avec un auteur ?
 - ❧ Dans le cadre d'une convention passée avec une bibliothèque numérique collective ?

Comment ?



☞ Choix techniques :

☞ Liés aux objectifs

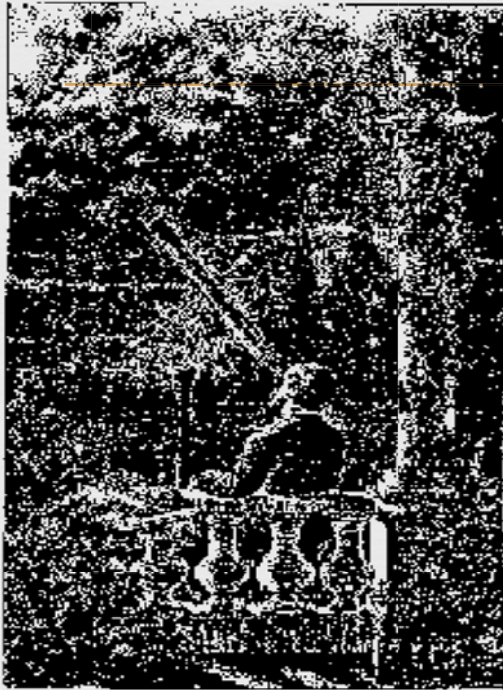
☞ Diffusion, préservation, les deux

☞ Aux moyens techniques à une date donnée

☞ Format de fichier, résolution, couleur

☞ Mode image, texte (OCR), table des matières et autres enrichissements (entités nommées, etc.)

Comparons...



Débuts de la numérisation
N&B - 300dpi - TIFF G4
1 page ~ **200Ko**



Numérisation maintenant
Couleur- 400dpi - TIFF
non compressé
1 page ~ **80Mo**

Plus de 500 fois plus !

Du né-numérique



Une masse et une diversité croissantes



Au commencement



- ❧ CD-ROMS, bases de données :
 - ❧ Encore dans les collections ?
 - ❧ Encore exploitables techniquement ?
 - ❧ Encore utilisés ?
 - ❧ Remplacés par une version en ligne ?

Livres et périodiques



- ❧ Acquis ou accessibles ?
 - ❧ Sous licence locale ou globale ?
 - ❧ Catalogués ou indexés ?
 - ❧ Libres ou protégés (DRM) ?
 - ❧ PDF, HTML, EPUB, indéterminés ?
 - ❧ Quelles relations au « papier » : copie exacte, contenu similaire, genre totalement nouveau ?
-
- ❧ Sans oublier les éventuelles éditions électroniques de la bibliothèque elle-même...

Comparons...

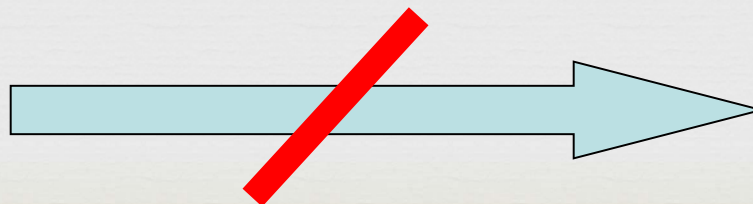


Le Monde



Le Monde.fr

France-soir



MEDIAPART



Données de la recherche



- ❧ Complexité des sources :
 - ❧ Applications spécifiques à un domaine, voire à un chercheur
 - ❧ Masse immense de données
- ❧ Et pourtant sources précieuses, et nécessité de pouvoir reproduire les expériences scientifiques
- ❧ Rejoint la problématique du « big data »

Archives



- ❧ Archives publiques
- ❧ Archives privées, achetées ou reçues

- ❧ Quels droits sur les documents ?
 - ❧ Conservation pendant la période d'utilisation courante (GED), et après ?
- ❧ Quelle variété de documents ?
 - ❧ Proiciels, environnements numériques de travail, e-mails, etc.
- ❧ Géré à l'échelle de la bibliothèque, de l'université, de la collectivité ?

Internet



- ❧ Dépôt légal de l'internet : à la BnF et à l'INA, mais aussi dans les bibliothèques de dépôt légal imprimeur
 - ❧ Sélections de sites à l'occasion des élections depuis 2004 (en 2012, 21 participantes sur 26), à terme sélections régionales
 - ❧ Possibilité de consultation sur place : arrêté en cours d'élaboration, tests avec les bibliothèques volontaires fin 2013

- ❧ Mais aussi : expositions en ligne, parcours pédagogiques...

Pourquoi ?



Conservation, préservation

Parce que



∞ Charte de la conservation dans les bibliothèques

<http://www.patrimoineecrit.culture.gouv.fr/ConsRest/Charte%20de%20la%20conservation%20v19.pdf>

∞ « Art. 8

La préservation des documents numériques se définit comme l'ensemble des principes d'organisation, des programmes informatiques et des infrastructures mis en œuvre en vue du contrôle et de la régénération des informations numériques et numérisées ainsi que des métadonnées qui leur sont associées.

∞ NOTE

Le terme de préservation est préféré à celui de conservation dans le domaine numérique en raison du caractère anticipatif de ce mode de conservation. »

Parce que



- œ **Charte sur la conservation du patrimoine numérique de l'UNESCO (2003)**
http://portal.unesco.org/pv_obj_cache/pv_obj_id_8DB243E8359B6D2A0B75BC6E777D1551013A0100/filename/Charter_fr.pdf
- œ « Le patrimoine numérique se compose de ressources uniques dans les domaines de la connaissance et de l'expression humaine, qu'elles soient d'ordre culturel, éducatif, scientifique et administratif ou qu'elles contiennent des informations techniques, juridiques, médicales ou d'autres sortes, créées numériquement ou converties sous forme numérique à partir de ressources analogiques existantes. [...]
- œ Les documents numériques [...] sont souvent éphémères, et leur conservation nécessite des mesures volontaires d'entretien et de gestion dès leur création. »
- œ **MESURES REQUISES**
- œ **Article 6 - Elaborer des stratégies et des politiques**
- œ Des stratégies et des politiques doivent être élaborées pour protéger le patrimoine numérique en tenant compte du degré d'urgence, de la situation locale, des moyens disponibles et des prévisions d'avenir. Ce sera plus facile si les créateurs, titulaires du droit d'auteur, et les détenteurs de droits voisins et autres parties prenantes travaillent en coopération à la définition de normes communes compatibles et qu'ils partagent les ressources.
- œ **Article 7 - Sélectionner ce qu'il convient de conserver**
- œ Comme pour tout type de patrimoine documentaire, les principes de sélection peuvent varier d'un pays à l'autre, même si les principaux critères appliqués pour décider des documents à conserver doivent être leur importance ou leur valeur culturelle, scientifique, de preuve ou autre, sur la durée. Il est évident que les documents "d'origine numérique" doivent avoir la priorité. Les choix opérés et tout réexamen ultérieur doivent pouvoir être justifiés et reposer sur des principes, politiques, procédures et normes bien définis.

Et parce que



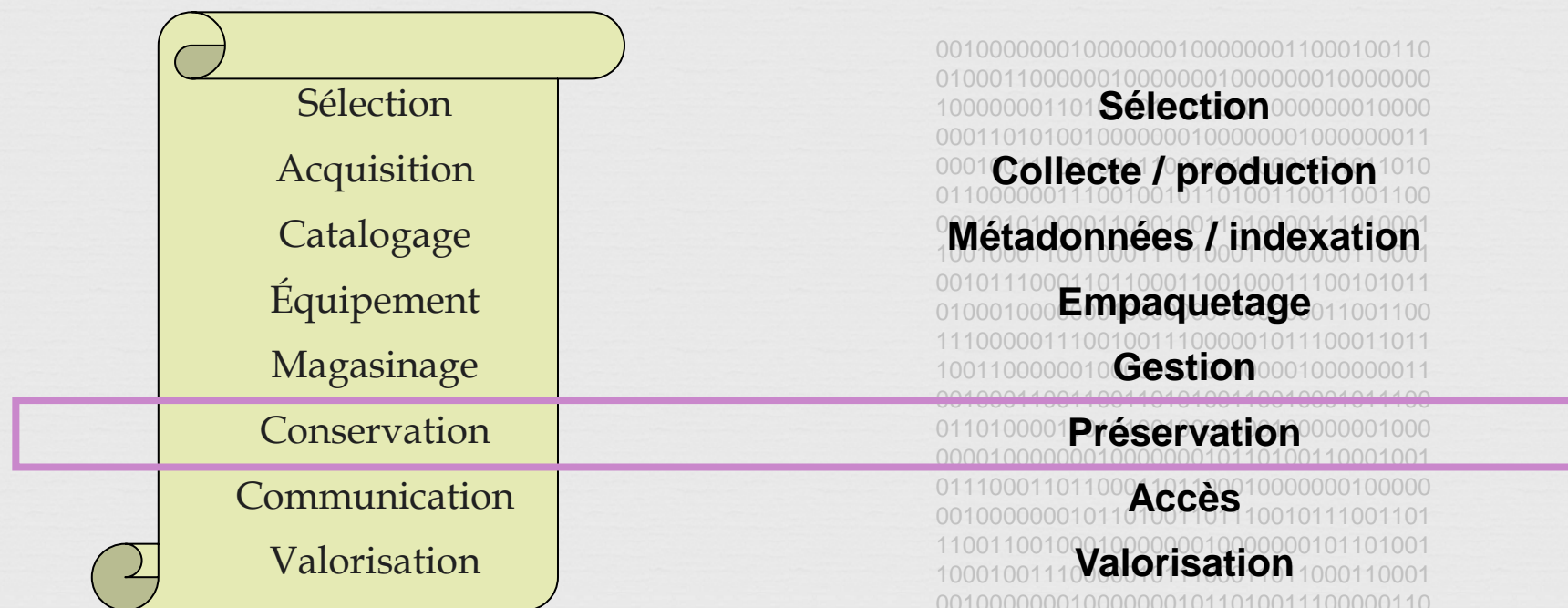
- œ Complétée lors de la conférence The Memory of the World in the Digital age: Digitization and Preservation (2012, Vancouver)
http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/mow/unesco_abc_vancouver_declaration_en.pdf
 - œ “Digitization, while not in itself a major means of preservation, can protect invaluable documents from handling and further deterioration. For some types of material, it is the only means of ensuring their survival”

- œ Communication de la Commission (européenne) sur la numérisation et l’accessibilité en ligne du matériel culturel et la conservation numérique, 2006
 - œ http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=162

Pourquoi nous ?



« conservation » et « préservation » : même missions, mêmes stratégies, quelques risques en plus





Gestion des fichiers

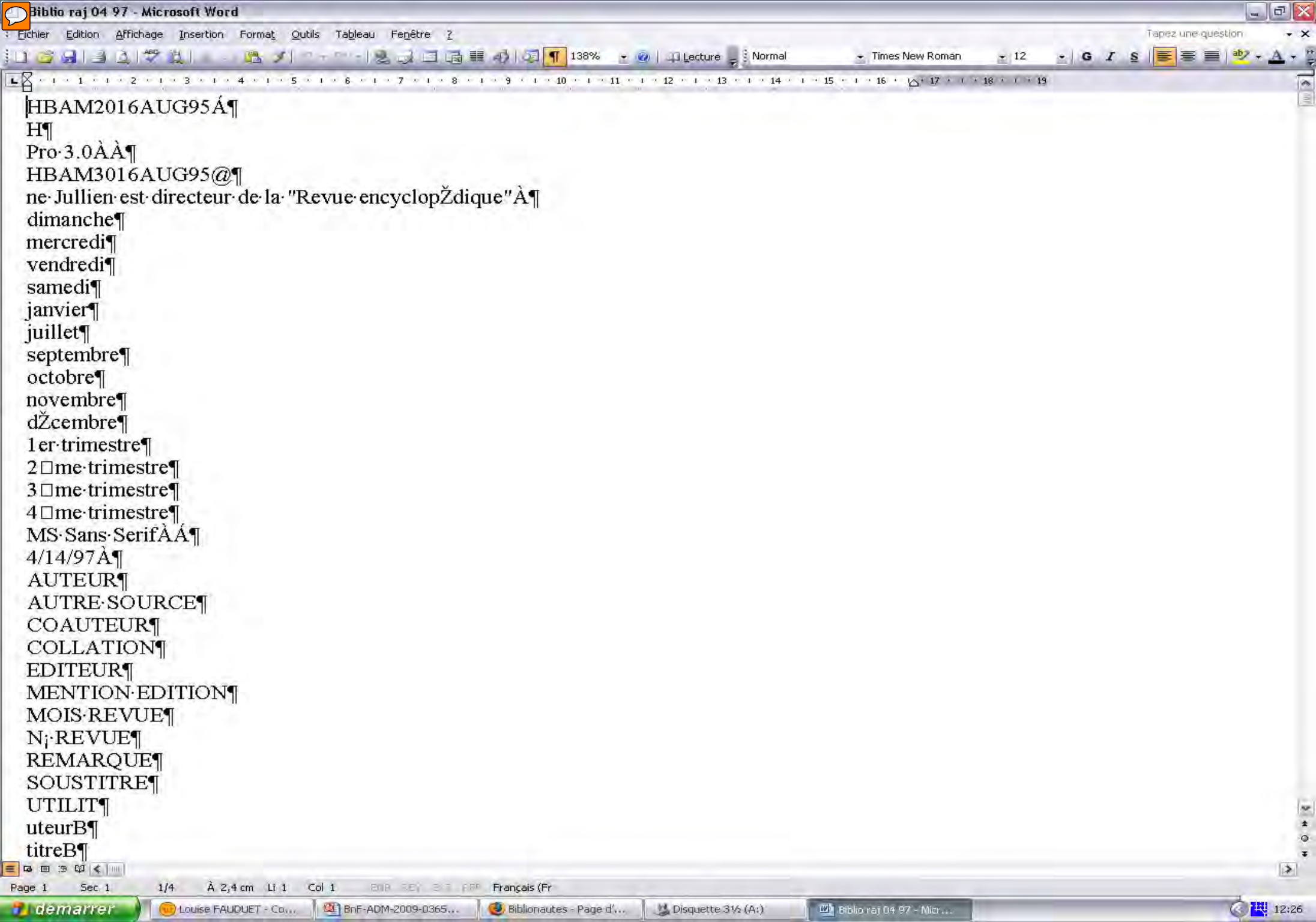
- Créer un nouveau dossier
- Publier ce dossier sur le Web

Autres emplacements

- ALEXANDRIE\BNF0015067 sur PADM1249
- Mes documents
- Favoris réseau

Détails

- | | | | |
|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| BOSQ2
Fichier
1 Ko | BOSQ21
Fichier
1 Ko | BOSQ2R
Fichier
6 Ko | BOSQ3
Fichier
1 Ko |
| BOSQ3R
Fichier
1 Ko | BOSQ4
Fichier
1 Ko | BOSQ4R
Fichier
1 Ko | BOSQ5
Fichier
1 Ko |
| BOSQ5R
Fichier
5 Ko | BOSQ6
Fichier
1 Ko | BOSQ6R
Fichier
3 Ko | CIBLEKM
Fichier
1 Ko |
| CLAS4 | CLAS5 | CLAS6 | CLAS7 |
| CLAS8 | CLAS9 | CLAS10 | CLAS11 |
| CLAS12 | CLAS13 | CLAS14 | CLAS15 |
| CLAS16 | CLAS17 | CLAS18 | CLAS19 |
| CLAS20 | CLASF.CCC
Fichier CCC
1 Ko | CLASN°3
Fichier
1 Ko | CLASSES
Fichier
1 Ko |
| DUPLIC2
Fichier
1 Ko | F3415
Fichier
1 Ko | F4312
Fichier
1 Ko | MF1
Fichier
1 Ko |
| MF1R
Fichier
1 Ko | N°3
Fichier
1 Ko | N°3CLAS
Fichier
1 Ko | N°3CLASR
Fichier
1 Ko |
| N°3CLSS
Fichier
1 Ko | N°3CLSS
Fichier LIB
1 Ko | N°3CLSS.CCO
Fichier CCO
1 Ko | N°3CLSS.CLI
Fichier CLI
3 Ko |
| N°3CLSS.COL
Fichier COL
1 Ko | N°3CLSS.DON
Fichier DON
1 Ko | N°3CLSS.IDE
Fichier IDE
1 Ko | N°3CLSS.LIG
Fichier LIG
1 Ko |
| N°3CLSS.PAR
Fichier PAR
1 Ko | N°3CLSSR
Fichier
1 Ko | N°3R
Fichier
1 Ko | QANT18
Fichier
1 Ko |
| QANT18R
Fichier
1 Ko | QANT22
Fichier
1 Ko | QANT22R
Fichier
1 Ko | QANT182
Fichier
1 Ko |
| QANT182R
Fichier
1 Ko | QUA7DUP
Fichier
1 Ko | QUA7DUPR
Fichier
1 Ko | QUANDUP
Fichier
1 Ko |
| QUANDUPR
Fichier | QUANT14
Fichier | QUANT14R
Fichier | QUANTI
Fichier |



HBAM2016AUG95

H

Pro 3.0

HBAM3016AUG95@

ne Jullien est directeur de la "Revue encyclopédique"

dimanche

mercredi

vendredi

samedi

janvier

juillet

septembre

octobre

novembre

décembre

1er trimestre

2me trimestre

3me trimestre

4me trimestre

MS Sans Serif

4/14/97

AUTEUR

AUTRE SOURCE

COAUTEUR

COLLATION

EDITEUR

MENTION EDITION

MOIS REVUE

N° REVUE

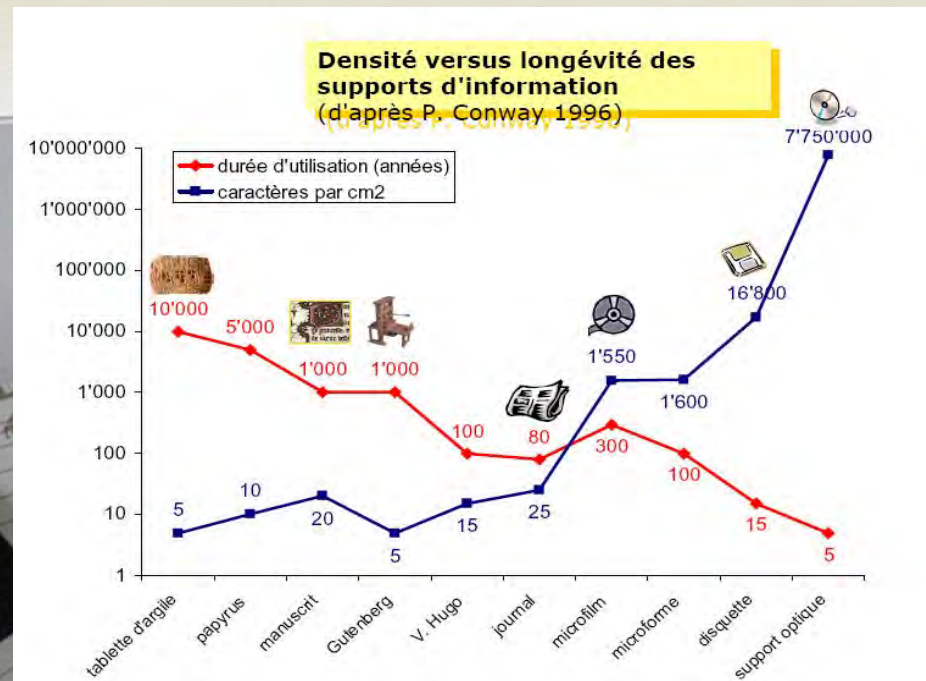
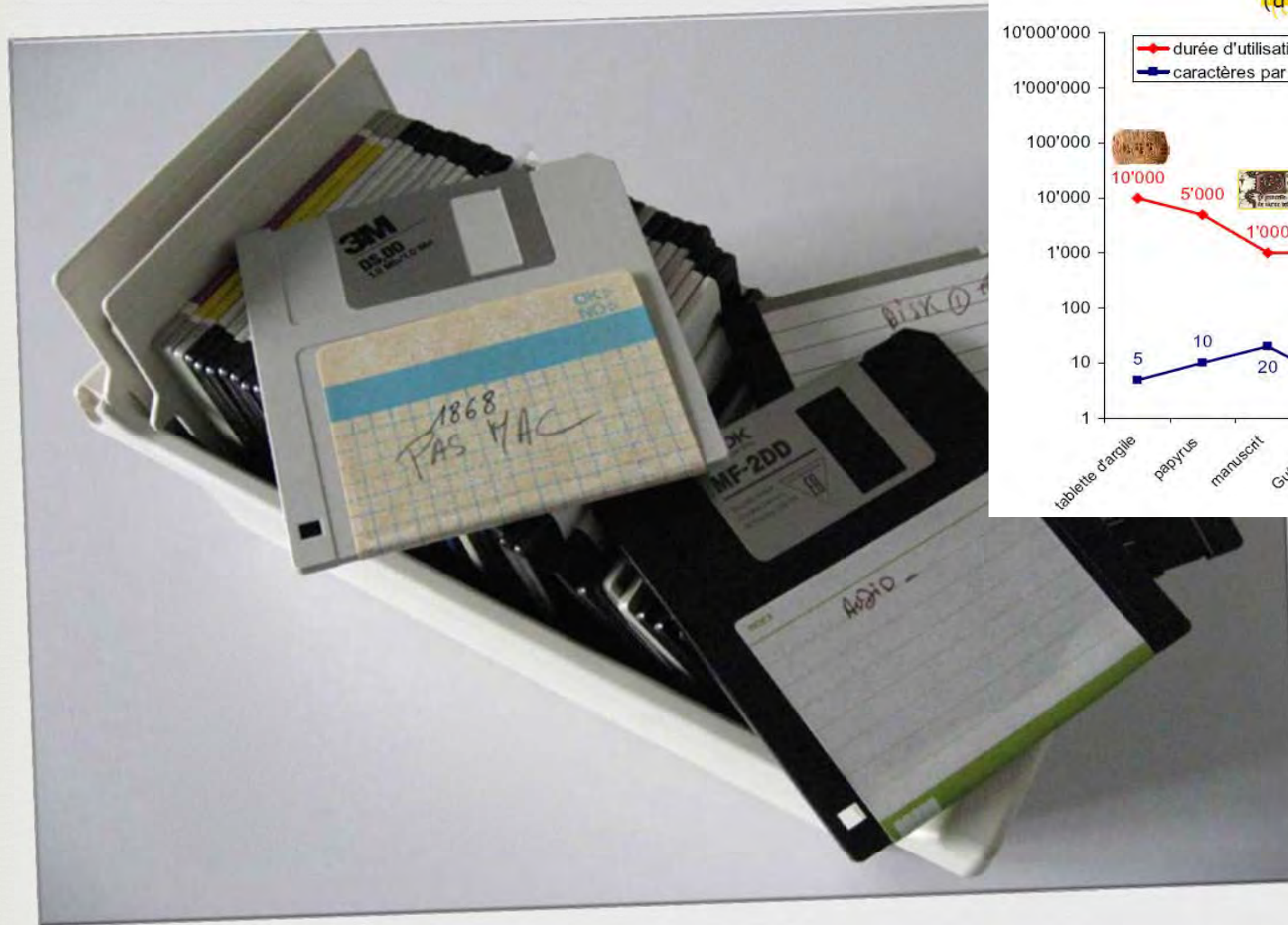
REMARQUE

SOUSTITRE

UTILIT

uteurB

titreB

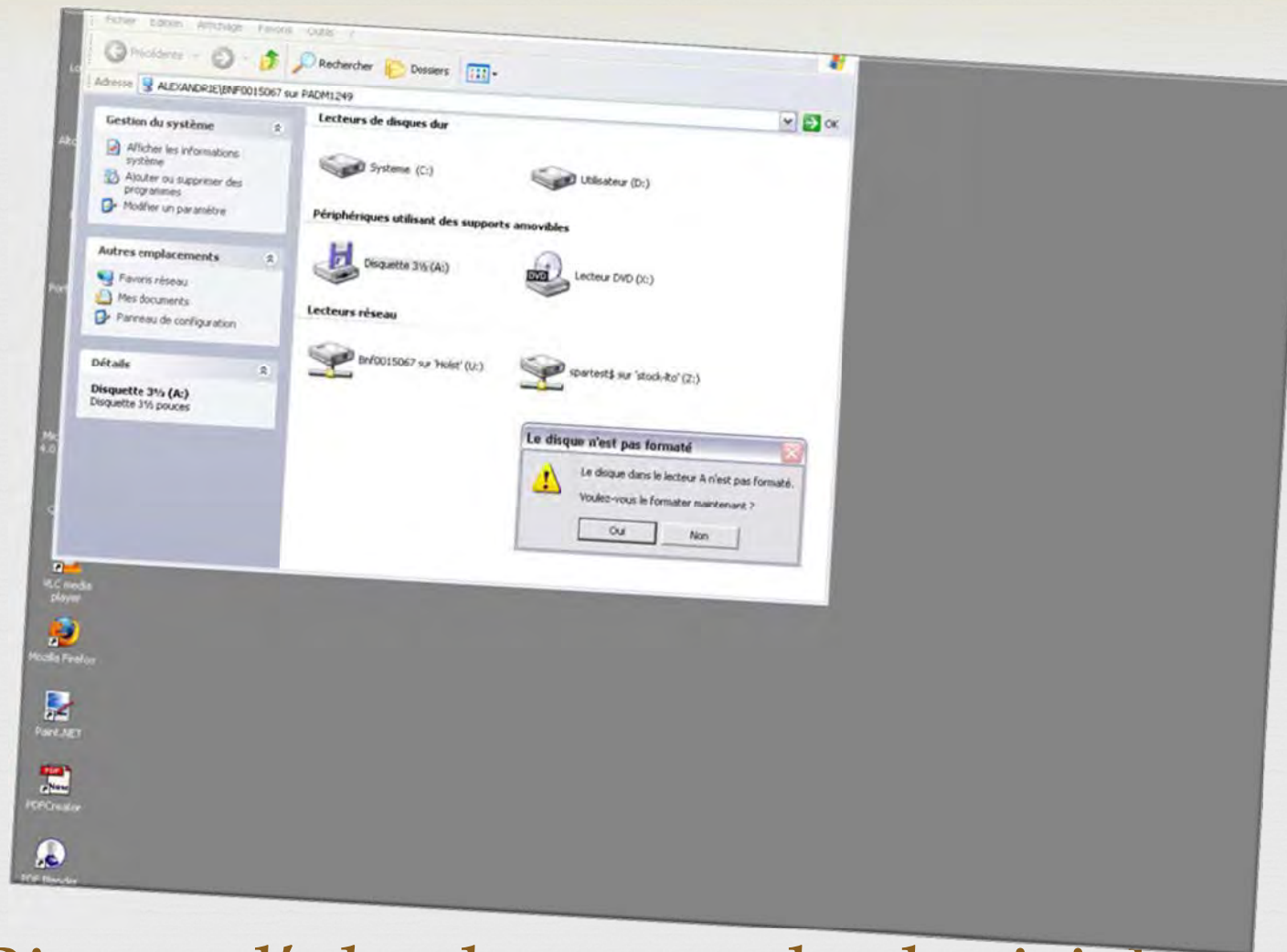


Risque d'obsolescence des supports
(fragilité, usure...)



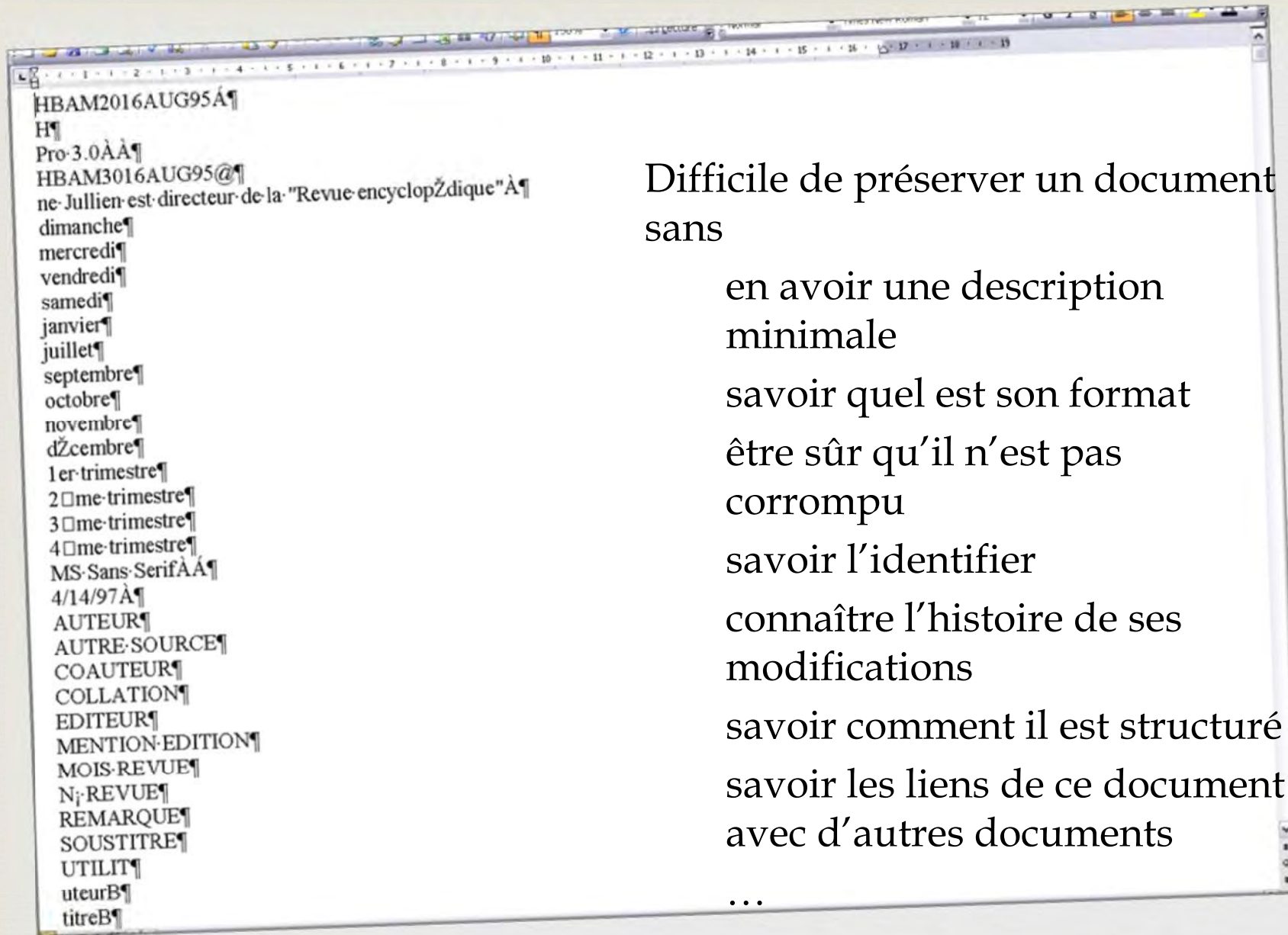
Risque d'obsolescence de l'environnement informatique

(disparition du matériel capable de lire les supports
ou de faire marcher les logiciels nécessaires à la lecture des données)



Risque d'obsolescence des logiciels

(disparition des logiciels capables de lire les données dans le format où elles sont conservées)



Risque d'obsolescence des données



Risques organisationnels

(ressources financières et humaines, gestion du projet)

Un « plan de conservation » numérique ?



<http://www.flickr.com/photos/nate/284184160/>

- ∞ Les informaticiens fournissent un service technique, mais pas la définition du service (définition du besoin)
 - ∞ ce n'est pas la même chose de mettre en place un bon système de climatisation, et de savoir quelle est la température et l'hygrométrie idéales d'un magasin.
- ∞ Les méthodologies de conservation sont les mêmes pour le papier et le numérique :
 - ∞ stratégies préventives,
 - ∞ plans de conservation,
 - ∞ choix de priorités en fonction de la valeur/fragilité des collections,
 - ∞ gestion de risques,
 - ∞ alertes et résolutions de problèmes,
 - ∞ plans d'urgence...
- ∞ Ce sont les bibliothécaires qui doivent établir des niveaux de service



<http://www.flickr.com/photos/jurvetson/855530042/>

Un « restaurateur » numérique ?



- ❧ La conservation du papier, c'est technique
 - ❧ (désacidification de masse, encres ferrogalliques, qualité des cuirs, effets des gommes à effacer sur la cellulose du papier...)
- ❧ Conserver des documents numériques c'est surtout les préserver
 - ❧ (en général, avec le numérique, après = trop tard)
- ❧ Stocker/sauvegarder ce n'est pas archiver !





<http://www.flickr.com/photos/yarhargoat/>

Un « magasin » numérique ?

- ⌘ Le numérique est lui aussi stocké sur un support physique
- ⌘ L'infrastructure de stockage coûte cher et prend de la place, il faut la sécuriser et la répliquer
- ⌘ À partir d'une certaine échelle, il faut une couche logicielle pour gérer les fichiers et les données



Un « catalogue numérique » ?

∞ Le catalogage ne se limite plus au descriptif :

∞ Métadonnées techniques, administratives (dont droits, historique), etc.

∞ Heureusement, les parties techniques sont les plus faciles à automatiser : extraction

∞ On jongle avec plusieurs granularités :

∞ d'une logique de document à une logique d'objet

```
- <metadata>
- <oai_dc:dc>
  <dc:identifiant>http://gallica.bnf.fr/ark:/12148/bpt6k10
- <dc:title>
  Les 100 jours de voyage en zigzag : de Paris à Venise
</dc:title>
<dc:creator>Abadie, Louis (18..-19.. ; avocat)</dc:cre
<dc:publisher>impr. de A. Lebon (Paris)</dc:publisher>
<dc:date>1862</dc:date>
<dc:subject xml:lang="fre">France -- Descriptions et
<dc:subject xml:lang="fre">Italie -- Descriptions et v
<dc:language>fre</dc:language>
<dc:type xml:lang="eng">text</dc:type>
<dc:type xml:lang="fre">monographie imprimée</dc:ty
<dc:type xml:lang="eng">printed monograph</dc:ty
<dc:rights xml:lang="fre">domaine public</dc:right
<dc:rights xml:lang="eng">public domain</dc:right
<dc:format>application/pdf</dc:format>
</oai_dc:dc>
</metadata>
</record>
- <record>
- <header>
  <identifiant>oai:bnf.fr:gallica/ark:/12148/bpt6k39258m</
  <datestamp>2007-01-19</datestamp>
  <setSpec>gallica</setSpec>
  <setSpec>gallica:9</setSpec>
  <setSpec>gallica:9:944</setSpec>
  <setSpec>gallica:monographies</setSpec>
</header>
- <metadata>
- <oai_dc:dc>
  <dc:identifiant>http://gallica.bnf.fr/ark:/12148/bpt6k39
- <dc:title>
  Bâtiments de la Bibliothèque nationale : rapport / M
</dc:title>
- <dc:contributor>
  Barthélemy-Saint-Hilaire, Jules (1805-1895). Rédact
</dc:contributor>
<dc:creator>France. Ministère de l'instruction public
<dc:publisher>Impr. nationale (Paris)</dc:publisher>
<dc:date>1879</dc:date>
<dc:subject xml:lang="fre">Bibliothèque nationale (B
<dc:language>fre</dc:language>
```


Comment ?



Un problème « technique » ?

Préserver = connaître



- ❧ Le point de départ de toute planification de la préservation, c'est la connaissance des collections
 - ❧ D'un point de vue documentaire → priorités selon valeur
 - ❧ D'un point de vue matériel → priorités selon état

- ❧ Pour mettre en œuvre la préservation, on s'appuie sur :
 - ❧ Les volumes
 - ❧ Les formats
 - ❧ Les métadonnées disponibles
- ❧ → à prévoir autant que possible...

Check-list : l'OAIS



- œ Open archival information system (ISO 14721)
- œ Le modèle d'information
 - œ Comment décrire mes objets en vue de leur préservation ?
- œ Le modèle fonctionnel
 - œ Quelle organisation mettre en place pour assurer cette activité de préservation ?
- œ Les stratégies de préservation
 - œ Quelles sont les méthodes pour éviter l'obsolescence technologique ?
- œ V1 en français : http://pin.association-aristote.fr/lib/exe/fetch.php/public/documents/norme_oais_version_francaise.pdf
- œ V2 en anglais : <http://public.ccsds.org/publications/archive/650x0m2.pdf>

Modèle d'information



Monde réel :



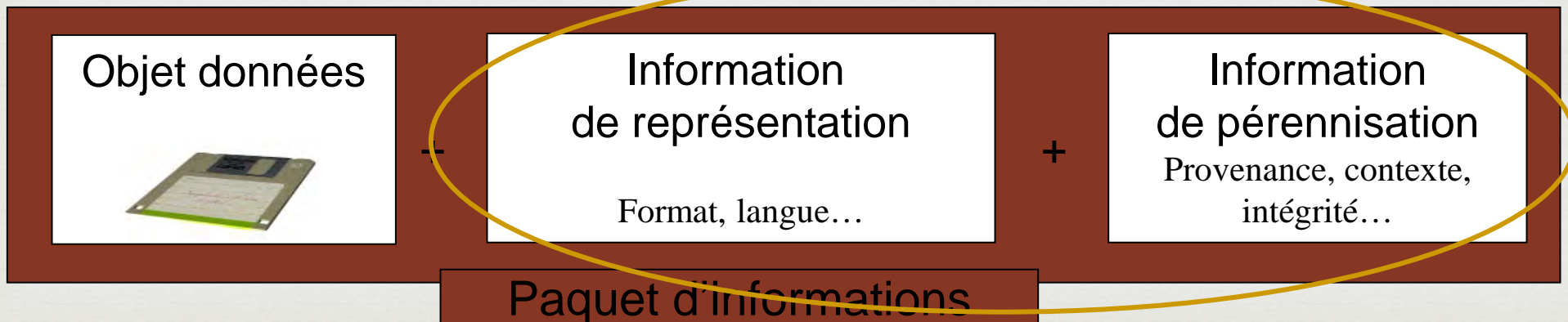
+



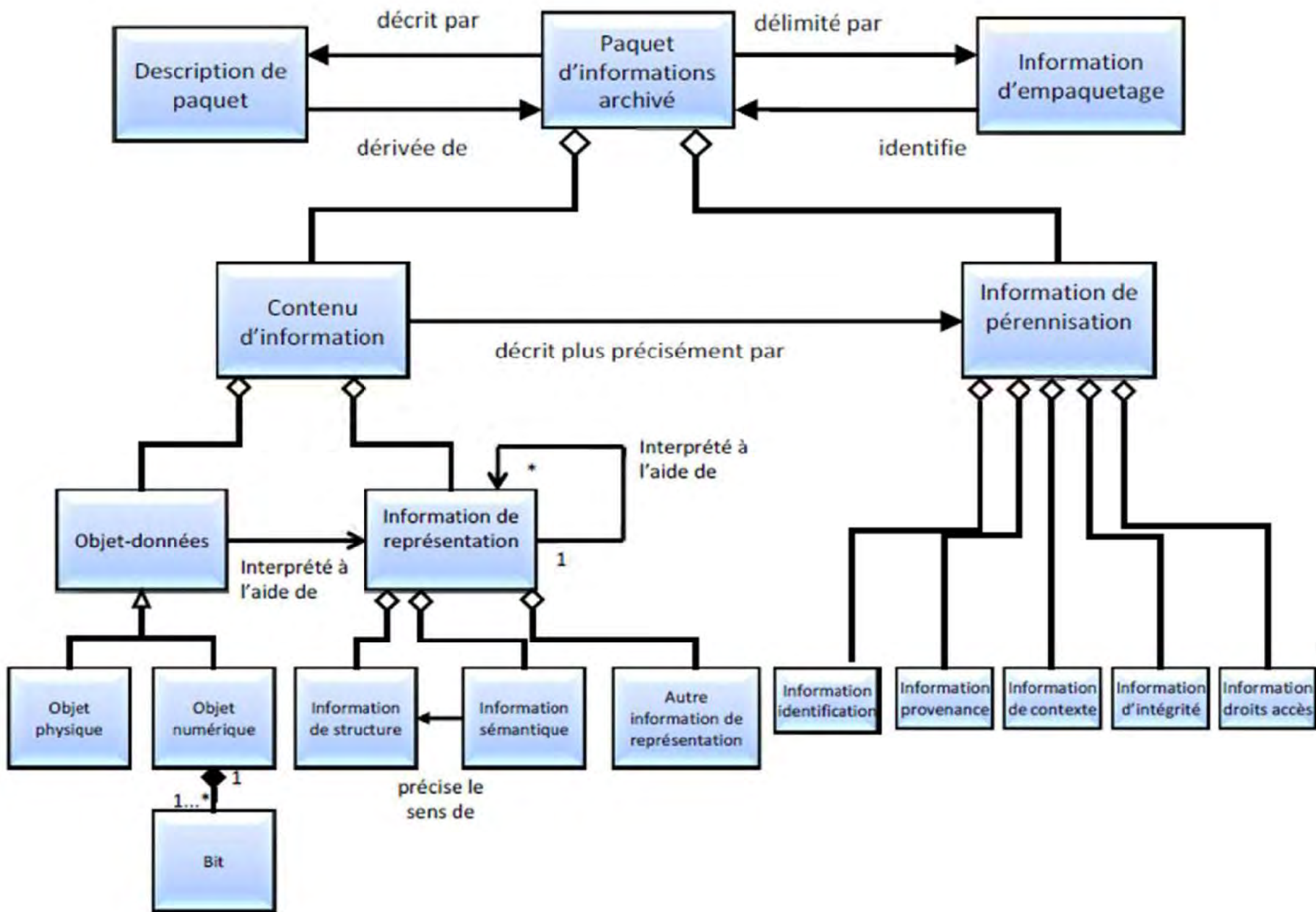
+



Monde numérique :



métadonnées



Documenter : contexte et provenance

- ❧ Documenter tous les choix effectués pour un type de document donné
 - ❧ Cahier des charges de la numérisation
 - ❧ Evolutions, choix additionnels en cours de route
 - ❧ Problèmes rencontrés sur les fichiers
 - ❧ Procédures de versement
 - ❧ Dates extrêmes des marchés
- ❧ Documenter les choix « implicites » pour la structure du document
 - ❧ Numérisation à la page ou au chapitre ?
 - ❧ Océrisation pour tout ou partie de la collection ?
 - ❧ Avoir une politique de nommage des fichiers et du document dans son ensemble

Documenter : représentation

- ❧ Inclure (ou faire inclure) des métadonnées techniques dans le fichier image
 - ❧ Date de création, application d'origine
 - ❧ Réglages de l'appareil...
 - ❧ Métadonnées JFIF (pour JPEG) ou TIFF
- ❧ Procéder au moins à un contrôle manuel par échantillonnage de la validité des fichiers

Analyser : identification et cie

- ❧ *Identification*. Processus qui consiste à déterminer le format supposé d'un objet numérique sur la base d'indices extrinsèques significatifs et de signatures intrinsèques, internes (nombre magique, etc.) et externes (extension de fichier, etc.).
- ❧ *Validation*. Processus qui consiste à déterminer le niveau de conformité aux normes syntaxiques et sémantiques définies par la spécification du format de l'objet qui fait autorité.
- ❧ *Extraction de propriétés / Caractérisation*. Processus qui consiste à consigner les propriétés intrinsèques d'un objet numérique, signifiantes à des fins de classement, d'analyse et d'utilisation.
- ❧ *Évaluation*. Processus qui consiste à déterminer le niveau de conformité d'un objet numérique à des règles définies localement, pour une utilisation spécifique.

Traduit de <https://bitbucket.org/jhove2/main/wiki/Glossary>

Outils : exemples



- œ Pour un échantillonnage à la main :
 - œ Validation : <http://facile.cines.fr/>
 - œ Identification, validation et caractérisation : <http://description.fcla.edu/>

- œ Pour faire travailler ses informaticiens à la chaîne :
 - œ Le vétéran de l'identification : DROID <http://digital-preservation.github.com/droid/>
 - œ FIDO, le petit nouveau :
<http://www.openplanetsfoundation.org/software/fido>
 - œ Identification, validation et caractérisation : JHOVE et JHOVE2
<http://sourceforge.net/projects/jhove/>
<https://bitbucket.org/jhove2/main/wiki/Home>
 - œ Mediainfo, pour les ressources audiovisuelles
<http://mediainfo.sourceforge.net/fr>

Consigner : formats de métadonnées

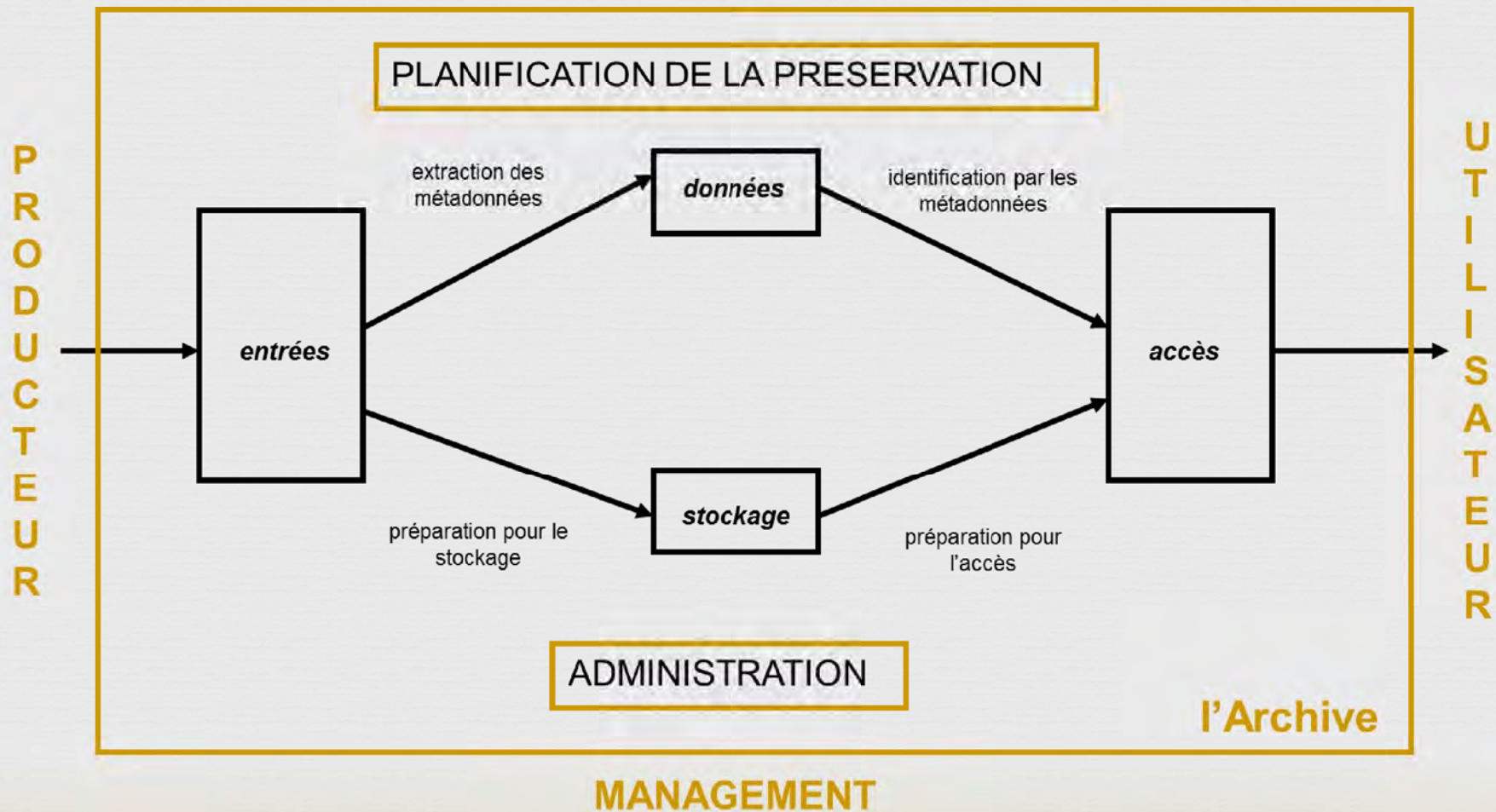
- ❧ Comme dans le domaine des catalogues :
 - ❧ D'un côté, des contenus attendus
 - ❧ De l'autre, des formats dans lesquels les exprimer
- ❧ Un effort de formalisation au départ, pour faciliter la recherche et la manipulation des métadonnées par la suite

- ❧ Des dictionnaires de données pour exprimer des informations dans un domaine spécifique
 - ❧ PREMIS
 - ❧ MIX, etc.
- ❧ Des formats « de regroupement »
 - ❧ METS, XFDU
 - ❧ MPEG-21 DIDL, etc.

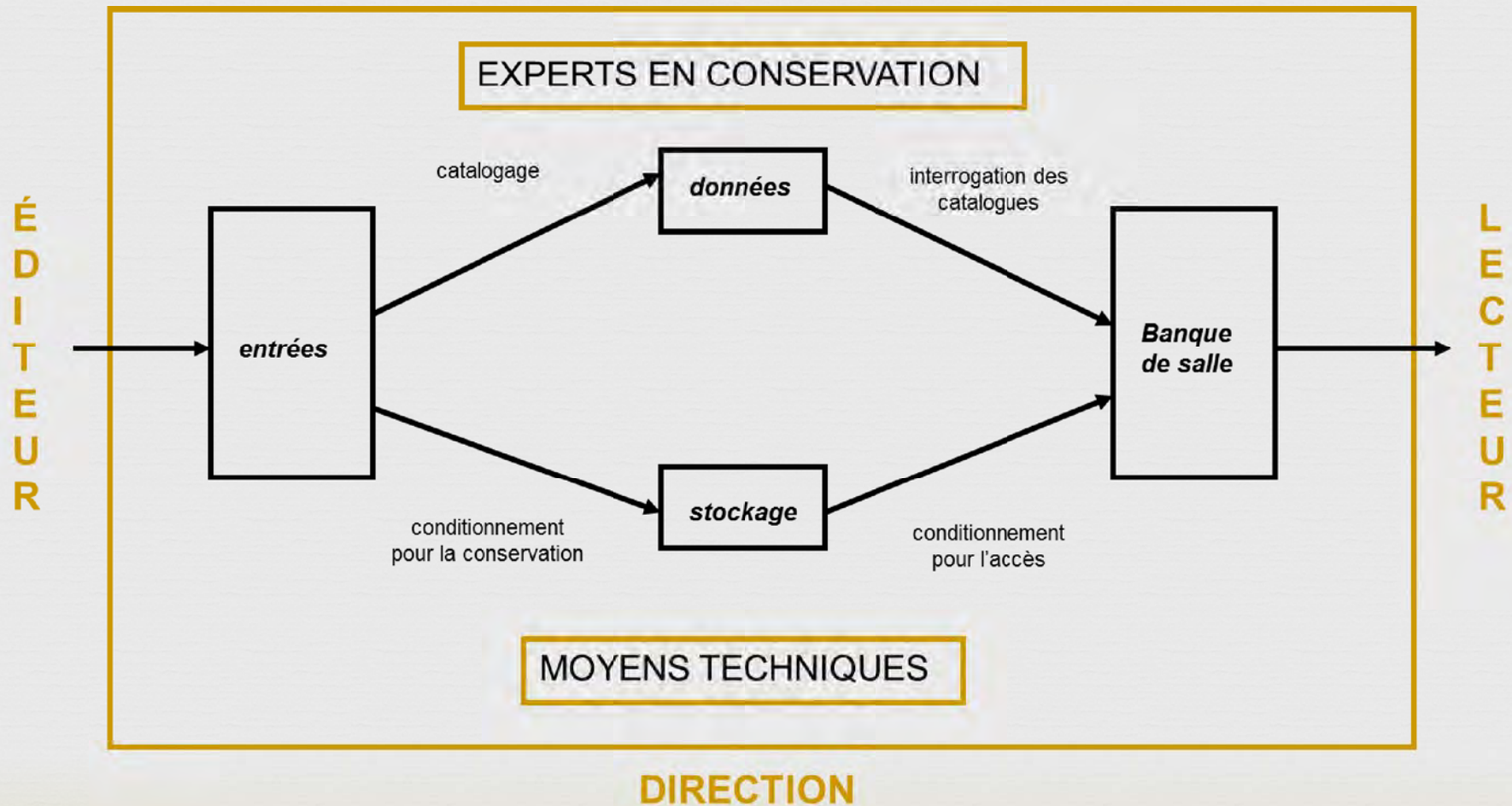
Préserver = responsabilités

- ❧ Comme dans le domaine de la conservation physique, on doit définir les rôles et responsabilités des acteurs et des services
 - ❧ Responsables de conservation, chargés de bibliothèque numérique, département des systèmes d'information...
- ❧ Là encore on peut s'appuyer sur l'OAIS

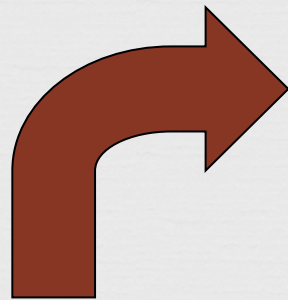
Modèle fonctionnel



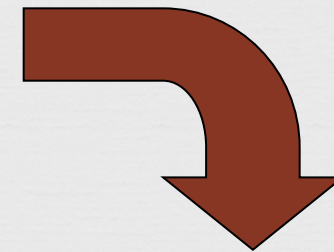
Toute ressemblance...



La notion de paquet d'information



Paquets archivés
(AIP)



Paquets à
verser (SIP)



Paquet
diffusé (DIP)

Plans de préservation ?



- œ Dans l'idéal : un plan de conservation commun pour toutes les collections, physiques et numériques

- œ Quelques initiatives intéressantes dans la définition de niveaux de services
 - œ Article de la National Library of Australia sur les intentions de préservation
 - œ <http://www.dlib.org/dlib/january13/webb/01webb.html>
 - œ Travaux de la National Digital Stewardship Alliance (organisation d'institutions américaines)

NDSA Levels of Digital Preservation

	Level One (Protect Your Data)	Level Two (Know Your data)	Level Three (Monitor Your Data)	Level Four (Repair Your Data)
Storage and Geographic Location	<ul style="list-style-type: none"> •Two complete copies that are not collocated •For data on heterogeneous media (optical disks, hard drives, etc.) get the content off the medium and into your storage system 	<ul style="list-style-type: none"> •At least three complete copies •At least one copy in a different geographic location •Document your storage system(s) and storage media and what you need to use them 	<ul style="list-style-type: none"> •At least one copy in a geographic location with a different disaster threat •Obsolescence monitoring process for your storage system(s) and media 	<ul style="list-style-type: none"> •At least 3 copies in geographic locations with different disaster threats. •Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems
File Fixity and Data Integrity	<ul style="list-style-type: none"> •Check file fixity on ingest if it has been provided with the content •Create fixity info if it wasn't provided with the content 	<ul style="list-style-type: none"> •Check fixity on all ingests •Use write-blockers when working with original media •Virus-check high risk content 	<ul style="list-style-type: none"> •Check fixity of content at fixed intervals •Maintain logs of fixity info; supply audit on demand •Ability to detect corrupt data •Virus-check all content 	<ul style="list-style-type: none"> •Check fixity of all content in response to specific events or activities •Ability to replace/repair corrupted data •Ensure no one person has write access to all copies
Information Security	<ul style="list-style-type: none"> •Identify who has read, write, move, and delete authorization to individual files •Restrict who has those authorizations to individual files 	<ul style="list-style-type: none"> •Document access restrictions for content 	<ul style="list-style-type: none"> •Maintain logs of who performed what actions on files, including deletions and preservation actions 	<ul style="list-style-type: none"> •Perform audit of logs
Metadata	<ul style="list-style-type: none"> •Inventory of content and its storage location •Ensure backup and non-collocation of inventory 	<ul style="list-style-type: none"> •Store administrative metadata •Store transformative metadata and log events 	<ul style="list-style-type: none"> •Store standard technical and descriptive metadata 	<ul style="list-style-type: none"> •Store standard preservation metadata
File Formats	<ul style="list-style-type: none"> •When you can give input into the creation of digital files encourage use of a limited set of known open file formats and codecs 	<ul style="list-style-type: none"> •Inventory of file formats in use 	<ul style="list-style-type: none"> •Monitor file format obsolescence issues 	<ul style="list-style-type: none"> •Perform format migrations, emulation and similar activities as needed

Préserver = agir

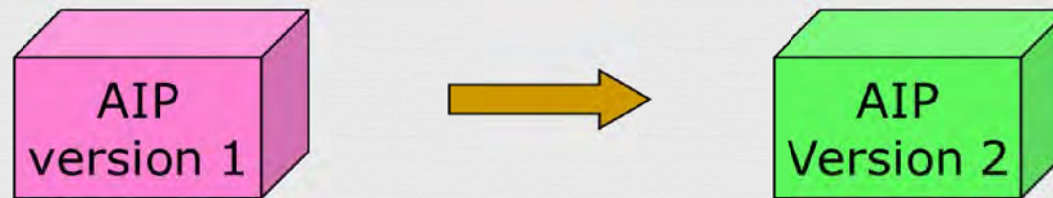


- ❧ L'OAIS donne aussi les grandes lignes des méthodes à mettre en place pour des actions de préservation

- ❧ Il s'agit de pouvoir agir face au risque d'obsolescence des objets numériques
 - ❧ Mais l'obsolescence peut être liée à des facteurs organisationnels, légaux, scientifiques et culturels et non seulement techniques
 - ❧ De plus elle est corrélée aux moyens et aux expertises d'une structure donnée à un moment donné

Stratégies de préservation (1)

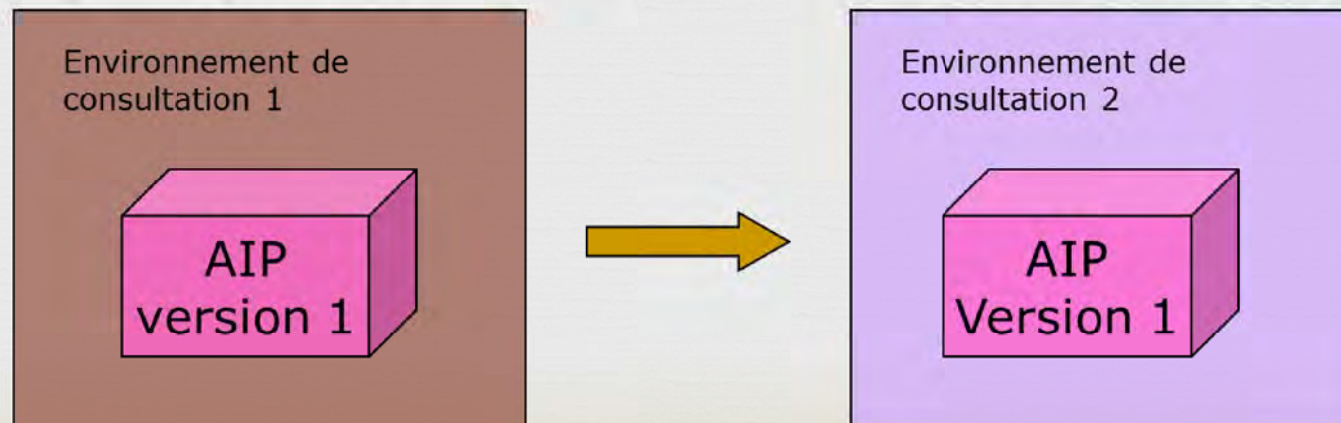
Migration : porte sur le paquet



Format propriétaire – difficile à préserver

Format ouvert et libre – facile à préserver et à transformer

Émulation : porte sur l'environnement de lecture du paquet



Stratégies de préservation (2)

☞ Plusieurs types de migrations

☞ Rafraîchissement de support



Support fragile – difficile à préserver- facile à reproduire

☞ Renouvellement de support



Support fragile n°1



Support fragile n°2 – plus de place

Support fragile n°3
– encore plus de place



d
u
p
l
i
c
a
t
i
o
n
s

Stratégies de préservation (4)

Emulation



Environnement matériel et logiciel difficile à préserver



Environnement matériel et logiciel courant

L'OAIS est abstrait



❧ C'est un outil de réflexion pour

❧ établir les responsabilités de la préservation

❧ donner à tous les acteurs un référentiel commun

❧ Mais ce n'est pas une solution clefs en main

❧ il ne donne pas de formats, schémas, règles ou techniques

❧ il ne décrit pas les applications informatiques et techniques à mettre en œuvre

❧ Il faut donc

❧ pour chaque implémentation, travailler sur les fonctions OAIS qui sont pertinentes pour son projet

❧ faire les choix techniques complémentaires : infrastructure, stockage, logiciel, métadonnées...

Alors comment ?



Concrètement ?

De la veille



- ❧ Le groupe PIN (Pérennisation des Informations numériques)
 - ❧ <http://pin.association-aristote.fr/doku.php/accueil>

- ❧ La Bibliothèque du Congrès, en tant que pilote du National Digital Information Infrastructure and Preservation Program
 - ❧ <http://www.digitalpreservation.gov/>

- ❧ Des ressources par type de collection
 - ❧ Archives de l'Internet : <http://netpreserve.org/about/index.php>
 - ❧ Audiovisuel : <http://prestocentre.eu/>
 - ❧ ...

Des outils



❧ Les outils libres des projets européens via l'Open Planets Foundation

❧ <http://www.openplanetsfoundation.org/> et son wiki
<http://wiki.opf-labs.org/display/KB/Home>

❧ Les registres de formats et les outils d'identification PRONOM et DROID

<http://www.nationalarchives.gov.uk/aboutapps/pronom/>

❧ UDFR <http://www.udfr.org/>

❧ JHOVE <http://hul.harvard.edu/jhove/> et JHOVE2
<https://bitbucket.org/jhove2/main/wiki/Home>

Des solutions : enseignement supérieur

∞ Archivage pérenne pour l'enseignement supérieur et la recherche : PAC (Plate-forme d'Archivage du CINES)

∞ <http://www.cines.fr/spip.php?rubrique219>

∞ Thèses électroniques, archives ouvertes HAL, revues en sciences humaines et sociales de PERSEE, ouvrages numérisés des bibliothèques universitaires, etc.

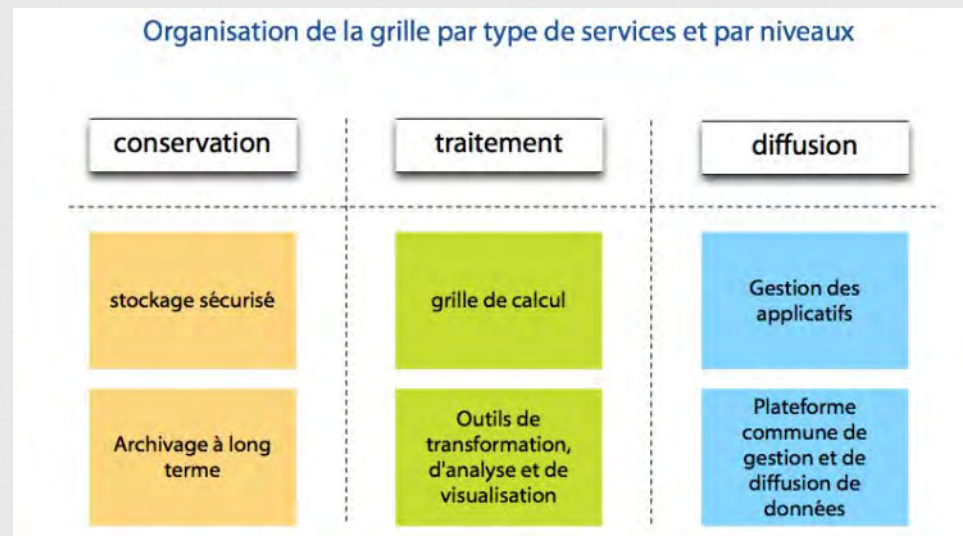
∞ Limitée à certains formats : validation du Format d'Archivage du Cines par anaLyse et Expertise

∞ <http://facile.cines.fr>

Des solutions : recherche



- Archivage intermédiaire des données de la recherche, collaboration TGE-ADONIS, CC-IN2P3 et CINES



- Grille ADONIS (<http://www.tge-adonis.fr/service/grille-adonis>) et ISAAC - Information Scientifique Archivée Au Cines
- Voir aussi : <http://www.tge-adonis.fr/article/les-presentations-des-journees-oais-sont-en-ligne>

Pour faire chez soi



∞ Safety Deposit Box (SDB) de Tessella

<http://www.digital-preservation.com/>

∞ Version 1 créée pour les Archives nationales de Grande-Bretagne en 2003

∞ Rosetta d' ExLibris

<http://www.exlibrisgroup.com/category/Rosetta>
[Overview](#)

∞ Développé en collaboration avec la bibliothèque nationale de Nouvelle-Zélande qui l'utilise depuis 2008

Pour faire faire ailleurs



❧ « Tiers-archivage »

- ❧ Venant d'institutions
- ❧ Venant d'entreprises
- ❧ Dans un entrepôt
- ❧ Dans le cloud
- ❧ Par des prestataires qui font aussi de l'archivage papier
- ❧ Ou non

❧ Dans des systèmes certifiés

- ❧ Nouvelle certification d'après la NF Z42-013 (Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes) : NF 461
- ❧ Travaux du CCSDS : Audit and Certification of Trustworthy Digital Repositories
 - ❧ Version septembre 2011 : <http://public.ccsds.org/publications/archive/652x0m1.pdf>, devenue norme ISO 16363
- ❧ Certification européenne à trois niveaux : <http://www.trusteddigitalrepository.eu/>

Des questions?



Merci de votre attention