

# Les collections audiovisuelles à l'heure du Numérique

# Un cas unique, une expérience à partager

- BnF, département de l'Audiovisuel :
  - Les collections audiovisuelles de la Bibliothèque nationale de France dépassent le million de documents : 900 000 audio, 250 000 vidéo et 100 000 multimedia (diapositives, CD-Roms, jeux vidéo et logiciels). Chaque année, grâce au dépôt légal, à des dons et des acquisitions, environ 25 000 documents supplémentaires enrichissent les collections.



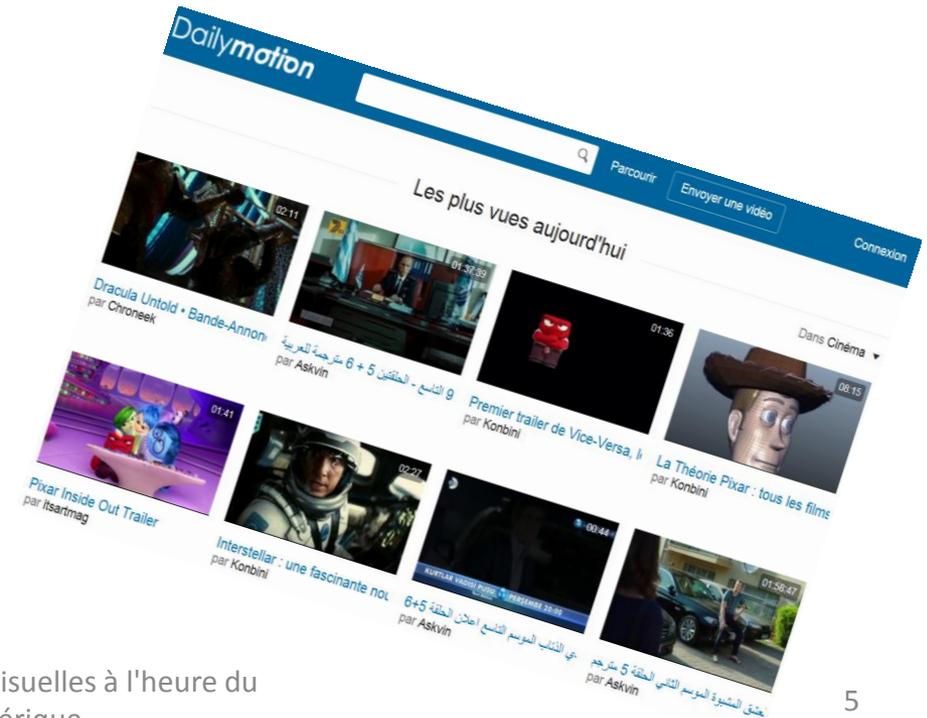
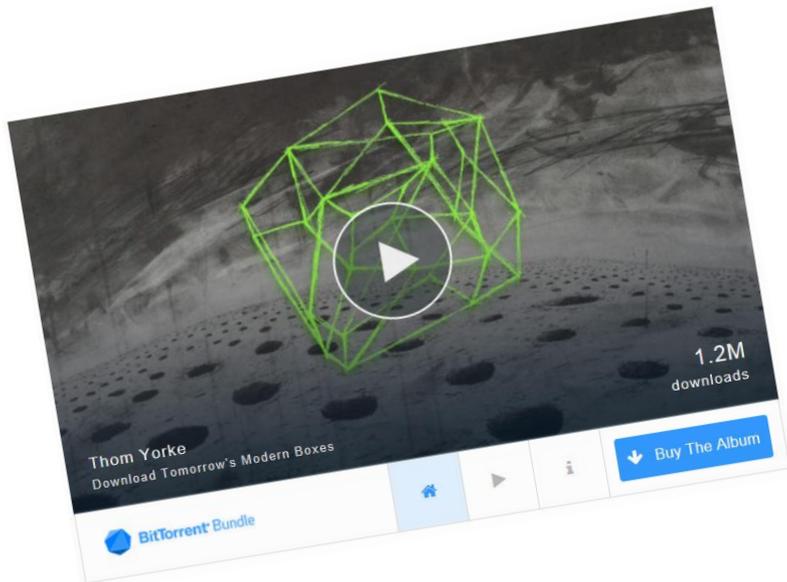
# Un cas unique, une expérience à partager

- Une collection majoritairement basée sur des supports analogiques, puis numériques
  - Du cylindre de 1877 au disque BluRay



# Un cas unique, une expérience à partager

- Une collection qui s'élargit à des formes dématérialisées
  - Du fichier audio auto-produit à la collecte des sites sur le Web



# Un cas unique, une expérience à partager

- Depuis les années 1980 pour l'audio, les années 1990 pour la vidéo, le numérique apporte de nombreux avantages :
  - Reproductibilité potentiellement infinie
  - Portabilité (fichiers)
  - Homogénéité des supports indépendamment des contenus (archivage de masse)
  - Identification et tri possibles (métadonnées)
  - Normalisation
- Aujourd'hui, un fichier peut être aussi bien le résultat d'un processus de numérisation que « numérique natif »

# Collecter / Conserver

## Numérisation

- La très grande diversité des technologies mises en œuvre dans l'enregistrement et la reproduction du son et de l'image peut rendre difficiles et d'autant plus urgentes (parfois même impossibles) certaines actions de sauvegarde



# Collecter / Conserver

## Numérisation

- A ce titre, ce ne sont pas forcément les technologies les plus anciennes qui sont les plus menacées.



# Collecter / Conserver

## Numérisation

Deux critères déclenchent une action de sauvegarde :

- Dégradation du support



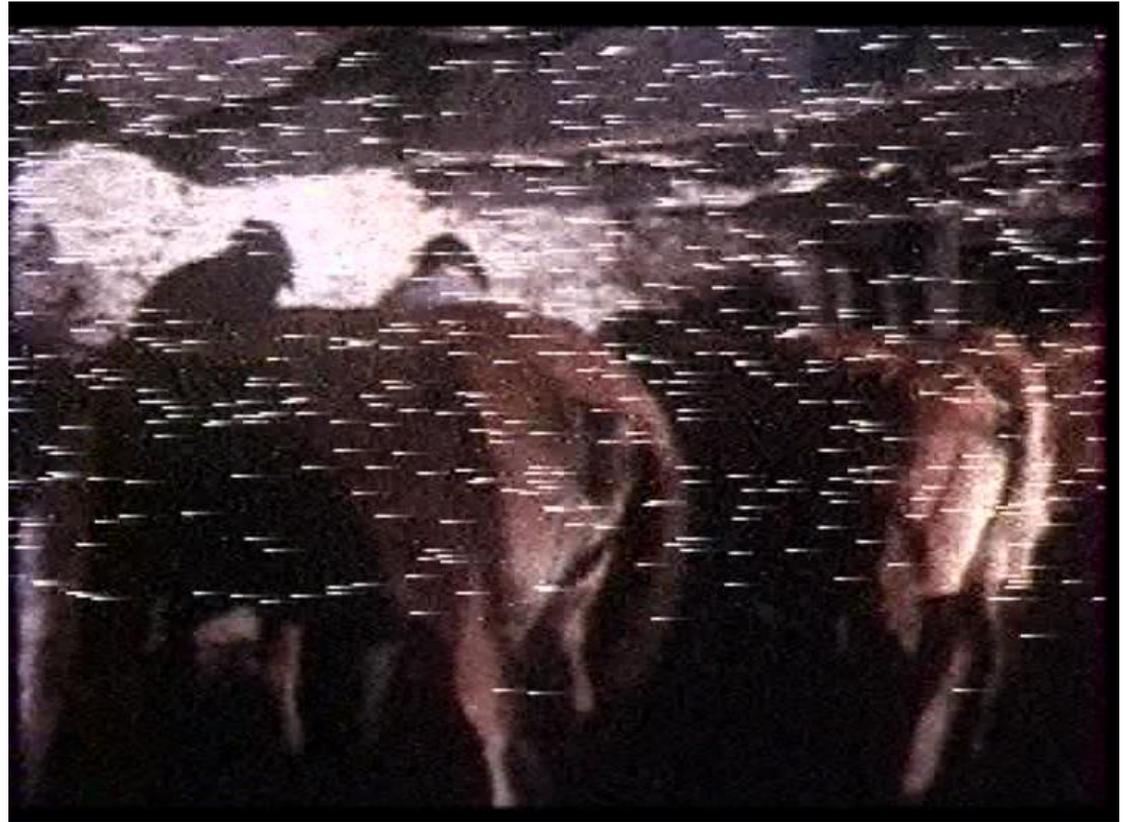
(disques à gravure directe : 1930-1950)

# Collecter / Conserver

## Numérisation

Deux critères déclenchent une action de sauvegarde :

- Dégradation du support



(vidéocassette U-Matic : 1970-1985)

# Collecter / Conserver

## Numérisation

Deux critères déclenchent une action de sauvegarde :

- Obsolescence des matériels



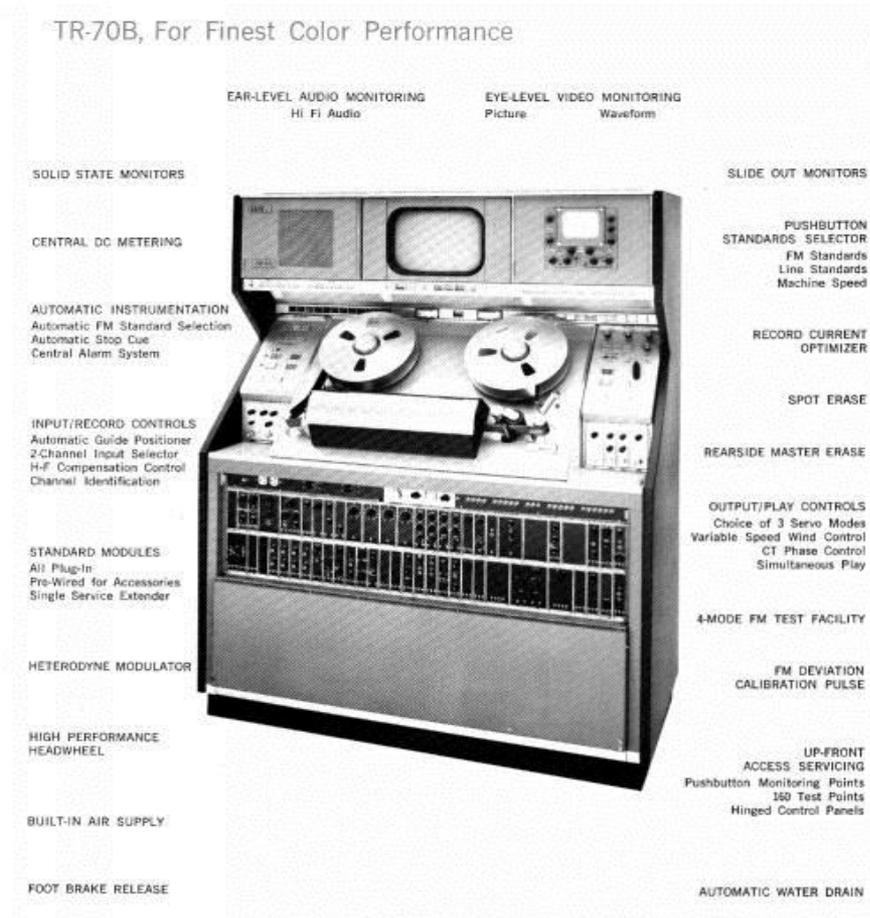
(magnétophone Studer : 1960-1980)

# Collector / Conserver

## Numérisation

Deux critères déclenchent une action de sauvegarde :

- Obsolescence des matériels



(magnétoscope 2 Pouces : 1960-1975)

# Collecter / Conserver

## Numérisation

### Plan de sauvegarde BnF :

- Fin 2013, 465 000 supports audio, 175 000 supports vidéos, 120 000 images fixes et 5 000 documents électroniques ont été sauvegardés sous la forme de fichiers numériques, représentant un volume de plus d'un Pétaoctet (soit 1.000To). La majorité de ces documents est accessible dans les emprises de la BnF (consultation est réservée aux chercheurs)

# Collecter / Conserver

## Fichiers numériques « natifs »

- La dématérialisation conduit à une substitution des supports par des fichiers.
- Les fichiers sont en règle générale reçus au format d'origine.
- *Mais qu'est-ce que le format « d'origine » d'un fichier audio ou vidéo?*  
[List of codecs - Wikipedia, the free encyclopedia](#)
- ...et convertis, si nécessaire, dans un format plus répandu (souci de pérennité) ou plus adapté au mode de diffusion souhaité.
- Un contrôle qualité / conformité est appliqué lors de la réception. Des opérations de tri et de renommage des fichiers peuvent être nécessaires.
- En parallèle, des données documentaires accompagnant les fichiers viendront enrichir les catalogues (moyennant souvent un travail de « mapping »).

# Collecter / Conserver

## Fichiers numériques « natifs »

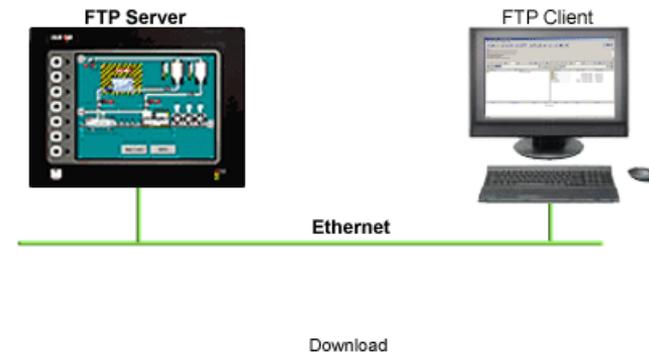
- La dématérialisation dans les environnements de production et de diffusion conduit à des échanges de fichiers (import et export) sous formes de supports de stockage de type CD ROM, DVD ROM, mais aussi disques durs...



# Collecter / Conserver

## Fichiers numériques « natifs »

- ...Ou, plus simplement, par le réseau via un serveur FTP (File transfert protocol).



# Communiquer / Valoriser

- La diffusion de fichiers issus de la numérisation de documents audiovisuels analogiques ou bien de fichiers numériques natifs présente de nombreux avantages (sous réserve du règlement des éventuels aspects juridiques)
  - Mise à disposition aisée grâce au téléchargement ou au « streaming »
  - Grand choix de formats possibles (avec ou sans compression)
  - Mise en valeur d'un fonds jusque-là dormant ou difficilement accessible
  - Gestion des données juridiques et commerciales facilitée grâce aux métadonnées

# Communiquer / Valoriser

- Communiquer et conserver sont deux actions étroitement liées :
  - Conserver est une entreprise à long terme dont la principale finalité est de pérenniser la communication du document. Numériser un document audiovisuel signifie réaliser un fichier « pivot » à partir duquel seront générés toutes les déclinaisons possibles de fichiers de diffusion. Il en va de même pour un fichier numérique natif qu'il faut réceptionner, contrôler et pérenniser.
  - Autrement dit : ***Conserver pour continuer à communiquer***

# Communiquer / Valoriser

- Communiquer et conserver sont deux actions étroitement liées :
  - Communiquer est à priori une action à court terme. Cependant , la pérennité du fichier pivot dont dépend cette action ne sera garantie que par des opérations de migration programmées. Opérations forcément coûteuses qui passeront par la valorisation (commerciale ou non) du fonds concerné.
  - Autrement dit : ***Communiquer pour continuer à conserver***

# Communiquer / Valoriser

- Une sélection de plus de 3 000 fichiers son est disponible dans Gallica.
  - Pour des questions essentiellement juridiques, la mise en ligne de documents son de la BnF se limite aujourd'hui sur Gallica aux documents libres de droit, ou dont les droits de diffusion ont été négociés.

The screenshot displays the Gallica website interface. At the top, the BnF logo and navigation links are visible. The search bar contains the query "Danceries françaises du 16e siècle Claude". Below the search bar, the results page shows a single result for a sound document. The document title is "Danceries françaises du 16e siècle / [Claude Gervaise, anonyme, comp.] ; Orchestre de violons, altos, gambes et basses ; Curt Sachs, dir. --1935". The document type is "Document sonore" and the author is "Gervaise, Claude". A thumbnail image of a vinyl record is shown next to the document information. The page also includes navigation elements like "Affiner", "Ma recherche initiale", and "Rechercher dans ces résultats".

# Communiquer / Valoriser

- Les fonds numérisés à cet effet sont constitués essentiellement de disques 78T.
- Diffusion sur Gallica au format MP3 (192 Kb/s), accompagnée de l'image du disque numérisé. Il s'agit, dans un format compressé, de la version restaurée du fichier d'origine (réduction des bruits de surface et des rayures).



Source gallica.bnf.fr / Bibliothèque nationale de France

# Communiquer / Valoriser

- Des documents directement issus de fonds d'archives peuvent être également mis en ligne, sous une forme « brute »



<http://gallica.bnf.fr/ark:/12148/bpt6k128287w/f1.item.pagination.r=deleuze.langFR>

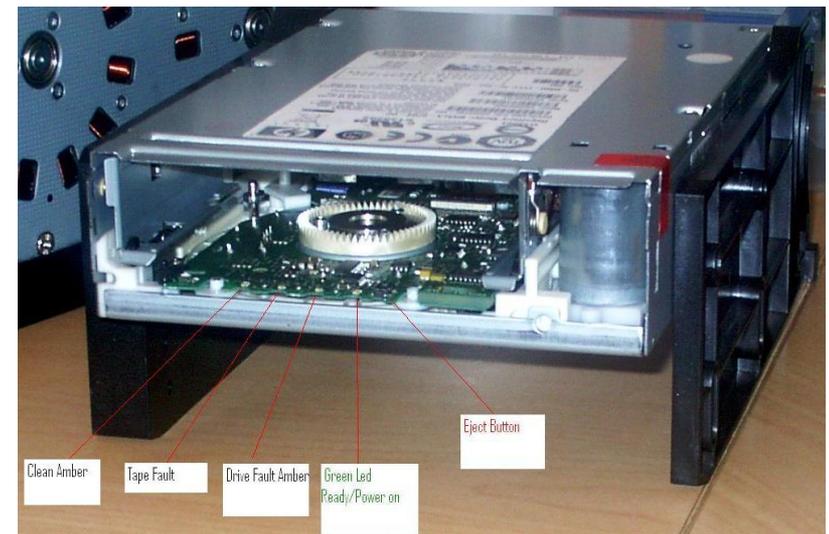
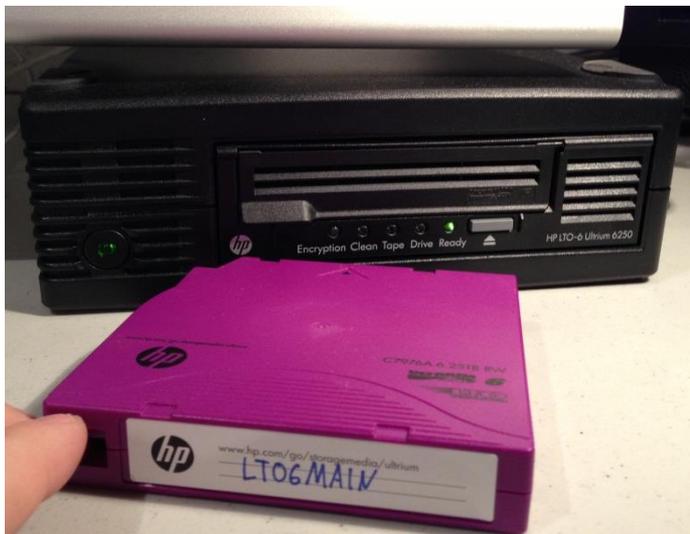
# Supports

- Le stockage des fichiers (contenu et métadonnées) peut se faire sur trois différents medias: disques optiques (CDR, DVDR, BDR),



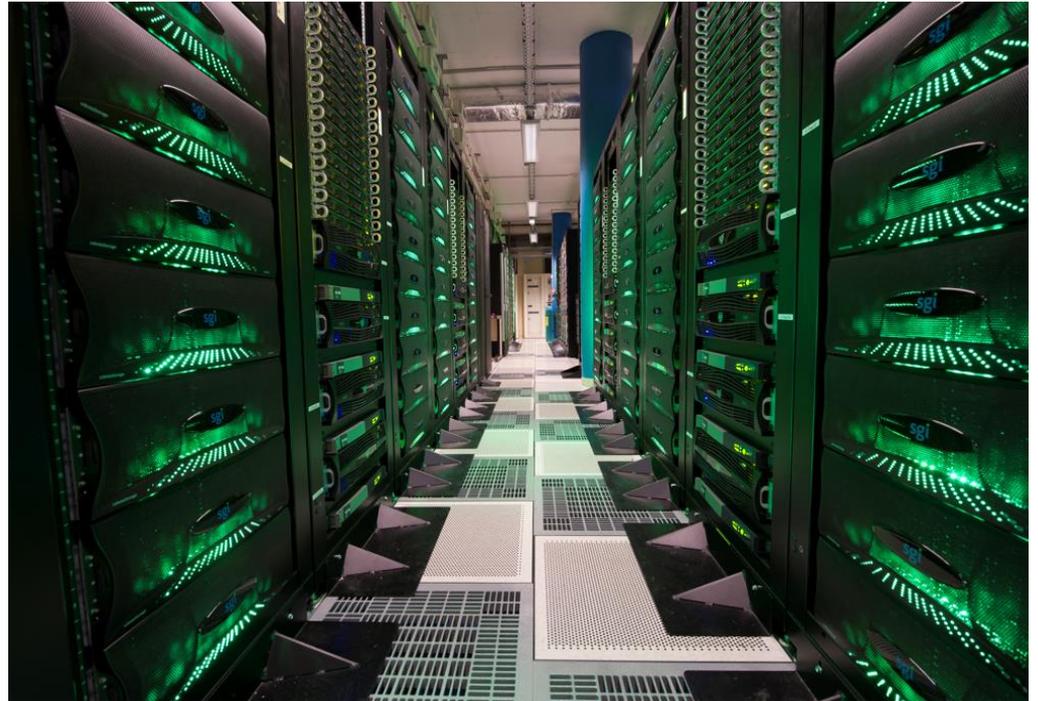
# Supports

- Le stockage des fichiers (contenu et métadonnées) peut se faire sur trois différents medias: disques optiques (CDR, DVDR, BDR), cartouches de données...



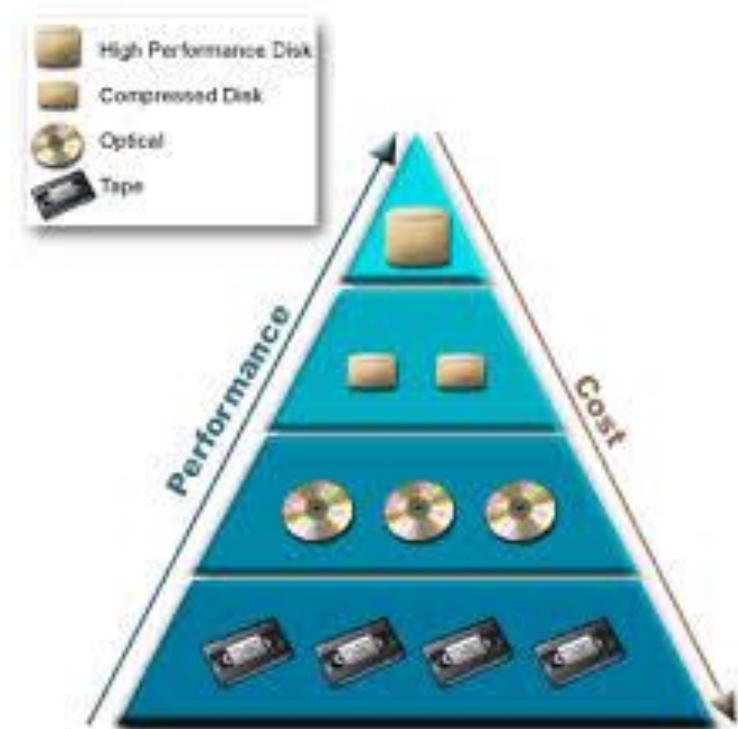
# Supports

- Le stockage des fichiers (contenu et métadonnées) peut se faire sur trois différents medias: disques optiques (CDR, DVDR, BDR), cartouches de données ou disques durs (à terme, cartes mémoire?)



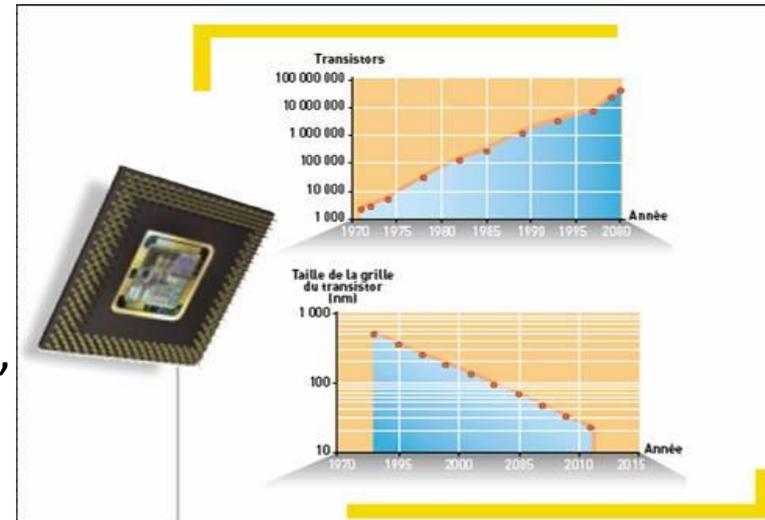
# Supports

- Chacun de ces supports de stockage offre des avantages et des inconvénients, en termes de temps d'accès, de fiabilité, de coût. Il s'agit avant tout d'établir des priorités; c'est ce que permet le « Hierarchical Storage Management » (HSM).
- Le HSM place les fichiers jugés importants sur les supports rapides et les autres sur des supports moins performants.



# Supports

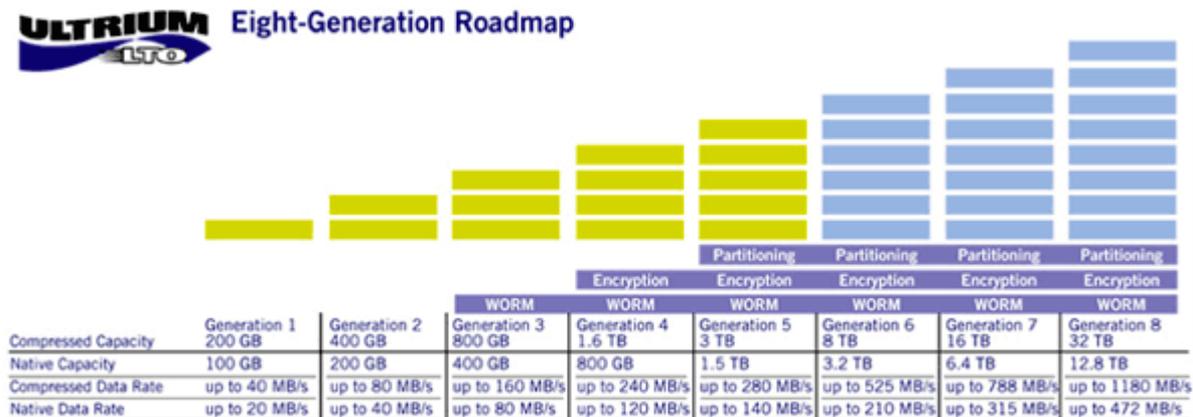
- **Aucun de ces supports de stockage n'offre une garantie de pérennité**; à cela, deux causes : dégradation et/ou usure des supports d'une part, obsolescence des supports et des équipements (loi de Moore)
- La "loi de Moore" a joué un rôle important dans l'histoire de l'informatique. Formulée pour la première fois par Gordon E. Moore dans un article de 1965, elle postule le doublement annuel des performances des circuits intégrés (mémoires et processeurs).
- Cette "loi", fondée sur un constat empirique, a été vérifiée par la suite. Il est estimé qu'elle se poursuivra jusqu'en 2017, date à laquelle elle devrait rencontrer des contraintes physiques.



# Supports

## Attention!

- Si, auparavant, 1 cassette audio analogique d'1 h numérisée en 16bits/44,1kHz = 1 CDR,
- Aujourd'hui, **1600** cassettes audio analogiques d'1h numérisées en 24bits/96kHz = 1 cartouche LTO6!



Note: Compressed capacities for generations 1-5 assume 2:1 compression. Compressed capacities for generations 6-8 assume 2.5:1 compression (achieved with larger compression history buffer).  
 Source: The LTO Program. The LTO Ultrium roadmap is subject to change without notice and represents goals and objectives only.

# Formats

- Le choix d'un format de fichier audio ou vidéo et, si nécessaire, d'un algorithme de compression, doit se faire en évitant les solutions "propriétaires" et dans la perspective des normes internationales ou "de fait".
- Ainsi, les formats WAV pour les fichiers audio, MPEG2 pour les fichiers vidéo, TIFF pour les fichiers image sont utilisés au département de l'Audiovisuel de la BnF. Par contre, l'INA a choisi le JPEG2000 comme format d'archivage de la vidéo.
- Le format d'archivage sur bande choisi par le département de l'Audiovisuel de la BnF est le TAR (créé dans les premières version d'UNIX et standardisé par les normes POSIX.1-1988 puis POSIX.1-2001)



```
azatoh@azabox:~$ tar --version
tar (GNU tar) 1.16
Copyright (C) 2006 Free Software Foundation, Inc.
This is free software. You may redistribute copies of it under the terms of
the GNU General Public License <http://www.gnu.org/licenses/gpl.html>.
There is NO WARRANTY, to the extent permitted by law.

Written by John Gilmore and Jay Fenlason.
azatoh@azabox:~$ tar -cf kernel.tar kernel/
azatoh@azabox:~$ tar -zcf kernel.tar.gz kernel/
azatoh@azabox:~$ tar -jcf kernel.tar.bz2 kernel/
azatoh@azabox:~$ du -sh kernel
1,9M   kernel
azatoh@azabox:~$ ll -h kernel.tar*
-rw-r--r-- 1 azatoh azatoh 1,7M 2007-03-20 18:43 kernel.tar
-rw-r--r-- 1 azatoh azatoh 344K 2007-03-20 18:43 kernel.tar.gz
-rw-r--r-- 1 azatoh azatoh 432K 2007-03-20 18:43 kernel.tar.bz2
azatoh@azabox:~$
```

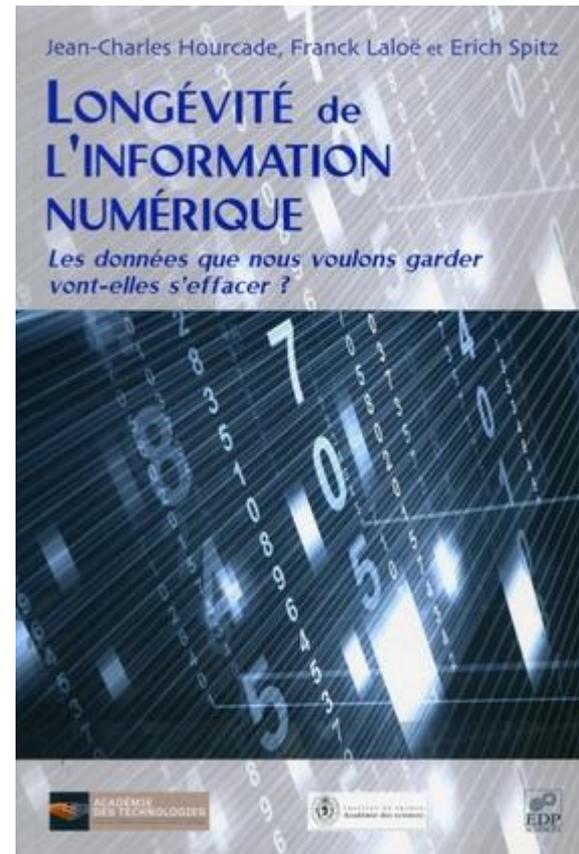
# Formats

- De nouveaux formats sont à envisager aujourd'hui pour l'archivage de la vidéo :
  - L'augmentation des capacités de stockage et de transfert conduisent à appréhender différemment la notion de compression appliquée à la vidéo (Avec perte/Sans perte?).
  - Le développement de la Haute Définition (4K aujourd'hui, 8K demain?) est en cours (poids moyen d'un fichier HD 1920x1080 d'1heure : 750Go)



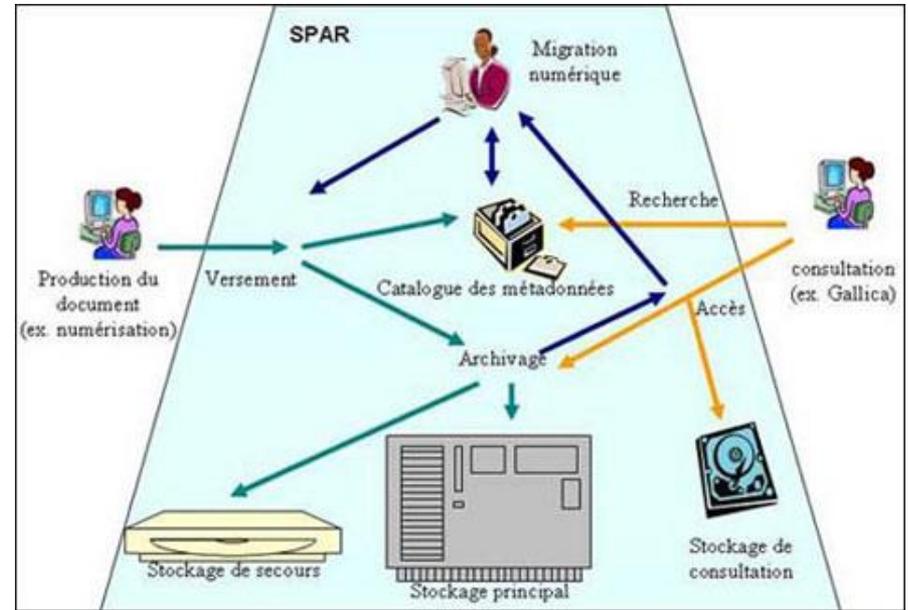
# Formats

- Ces fichiers de formats connus ou maîtrisés, correspondant idéalement à des normes internationales, seront amenés à migrer d'un support à l'autre
- La pérennité des données numériques, qu'elles proviennent d'actions de numérisation ou qu'elles soient « natives », doit donc **s'abstraire de la notion de support physique**



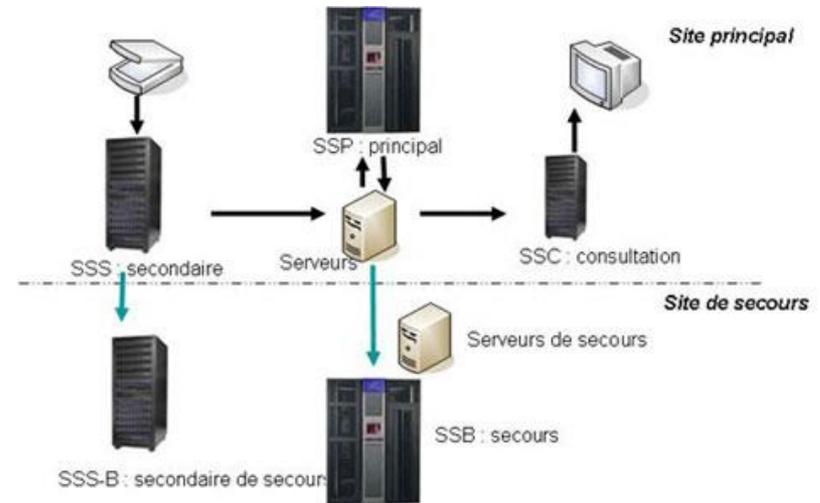
# Et ensuite?

- L'accroissement des collections numériques et la diversité de leurs formats placent la BnF devant le défi de leur conservation.
- La BnF a lancé le **projet SPAR**, véritable magasin numérique. Sa conception s'appuie sur des normes internationales faisant autorité dans le monde de la pérennisation des informations numériques. En particulier, SPAR respecte la **norme OAIS** (ISO-14721:2003), modèle de référence pour un système ouvert d'archivage d'information.



# Et ensuite?

- SPAR effectue de **multiples copies** des objets numériques et assure une **surveillance continue** de l'état des équipements et des médias d'enregistrement des fichiers permettant d'anticiper les recopies avant la perte définitive.
- Il permet également, grâce à une reconnaissance précise et complète des formats de données versées, de garantir la continuité d'accès en procédant aux **transformations** nécessaires en cas d'obsolescence technologique des outils informatiques de restitution. Ainsi, par exemple, lorsque le format d'image JPEG deviendra obsolète, SPAR sera en mesure de transformer les images concernées dans un nouveau format plus performant. Apporter cette garantie implique un travail permanent de **veille technologique** sur les formats, de prototypage et de tests des outils.
- De plus, SPAR permet à tout moment de **revenir en arrière** pour restituer les objets dans leur format d'origine.



L'infrastructure SPAR est constituée d'un ensemble d'éléments matériels et logiciels répartis sur deux sites. Cette répartition est l'assurance d'une **disponibilité élevée** des données et du système.  
Le **site principal** est l'accès privilégié au système par les producteurs et les utilisateurs des données.  
Un **site de secours** a été établi pour la sûreté de fonctionnement du système et afin de participer au processus de PRA (Plan de reprise d'activité suite à un sinistre).