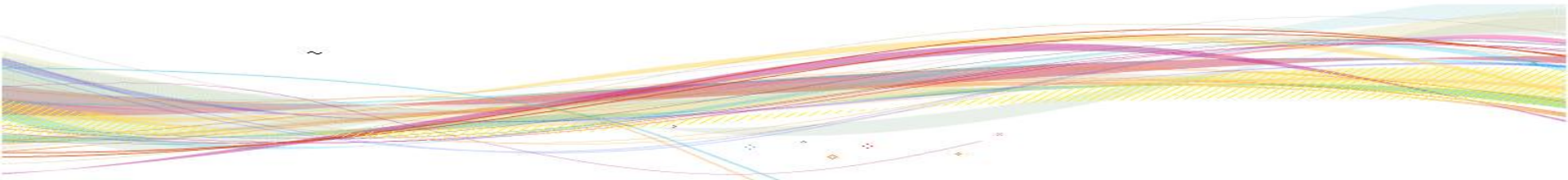


Médiaquitaine - Urfist

La conservation des documents numériques

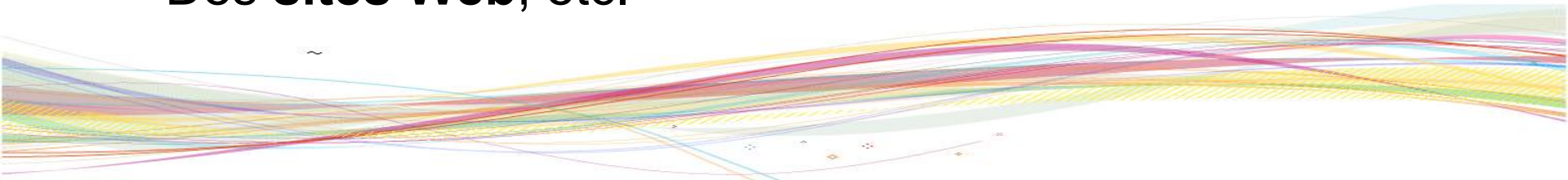
23 mai 2008

Olivier de Solan
Direction des archives de France



Que sont les documents numériques ?

- Des **images, sons, vidéos**
- Des **bases de données**
- Des systèmes de **gestion électronique de documents (GED)** intégrant des documents numérisés ou des fichiers bureautiques...
- Des documents échangés dans le cadre de **téléprocédures**
- Des **documents bureautiques**
- Des **messages électroniques**
- Des **sites Web**, etc.

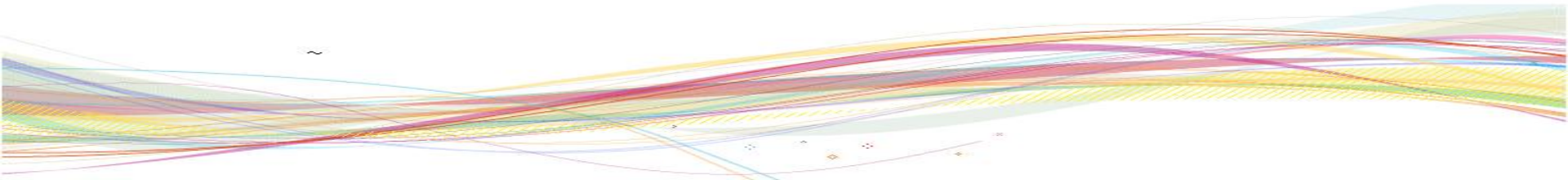


La conservation numérique : une question importante

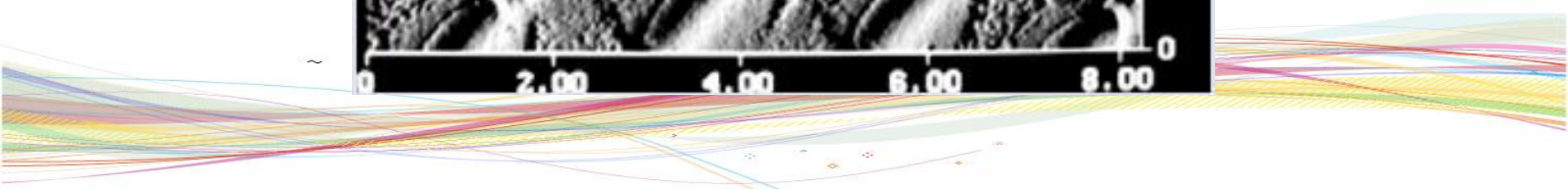
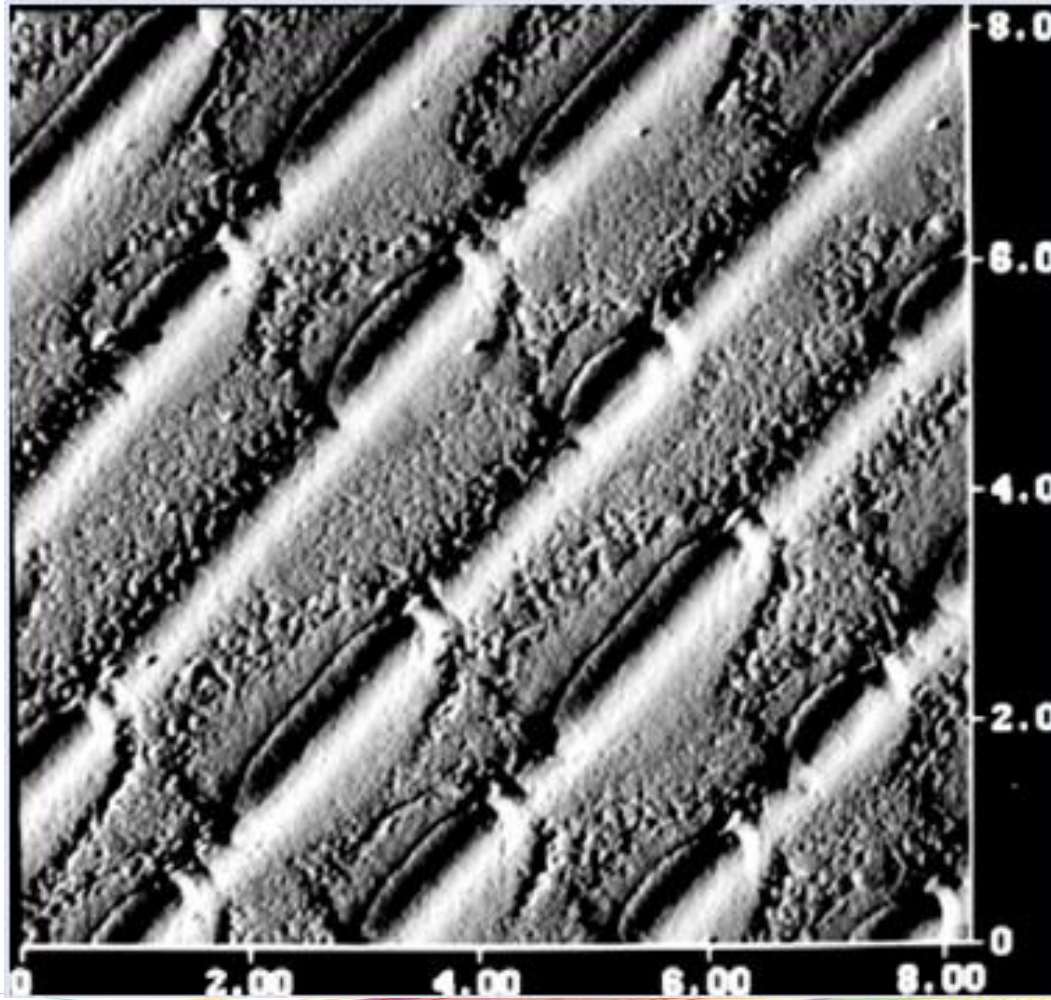
- Pour les originaux numériques
- Pour les documents numérisés

La numérisation dans les bibliothèques et les archives a débuté depuis une dizaine d'années

Un investissement important à sauvegarder

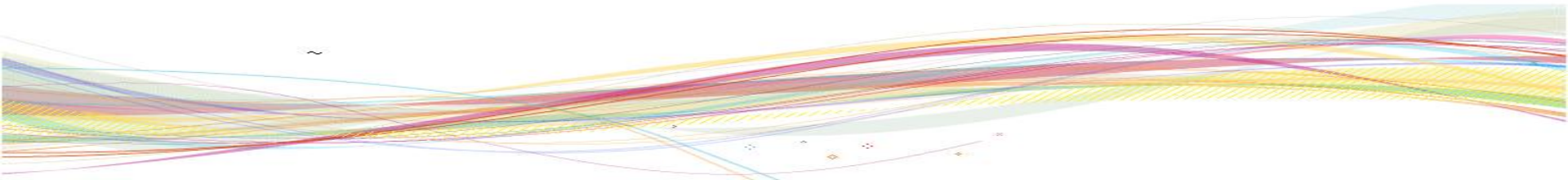


L'information numérique

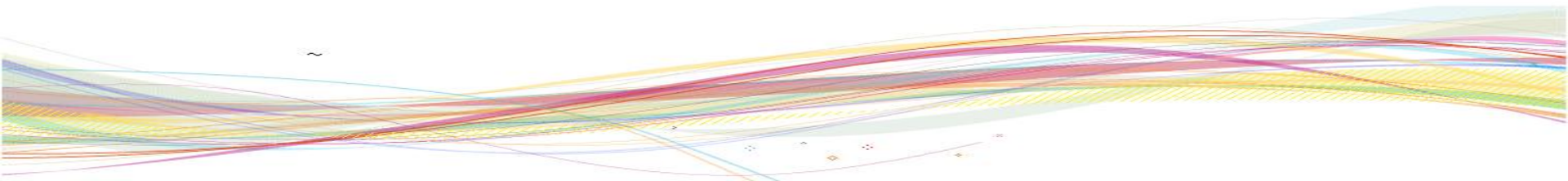


Des documents à risques ?

1. La fragilité et la durée de vie relativement faible des **supports** ;
2. L'obsolescence des applications informatiques et des **formats** de fichiers ;
3. La possibilité d'**altérer** les fichiers informatiques.

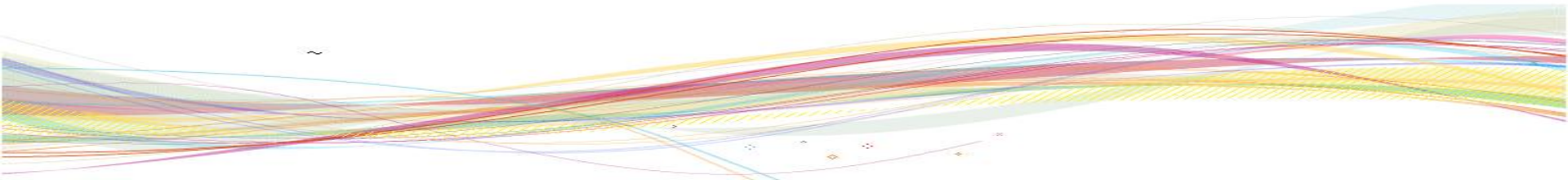


1. Les supports

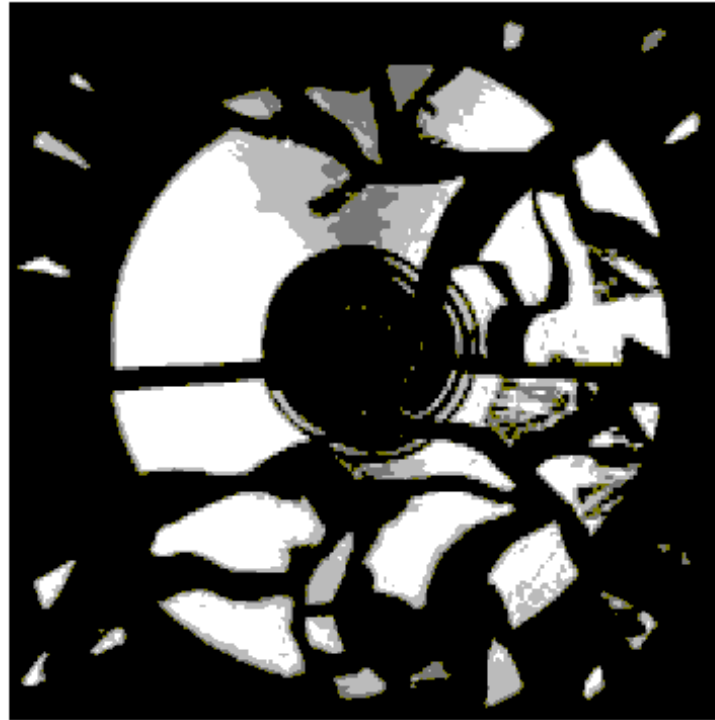


Les grands types de supports

- Disques optiques (CD-R, DVD-R...)
- Bandes magnétiques (LTO, SDLT...)
- Disques magnétiques (SCSI, S-ATA...)



Les supports sont fragiles

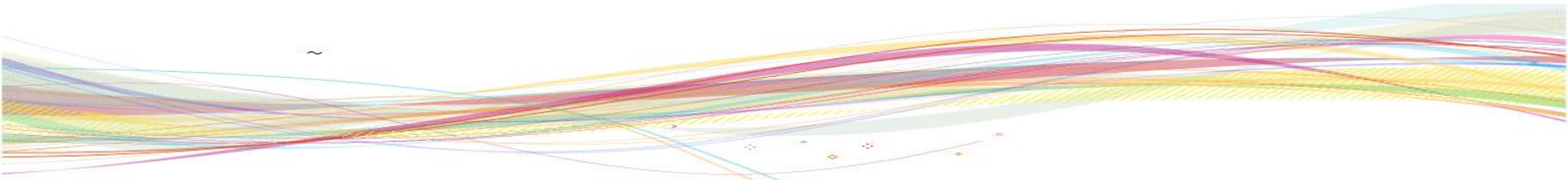


© LAM - JMF

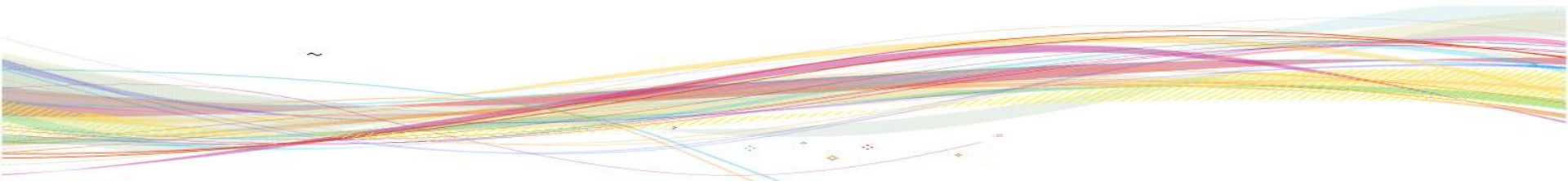
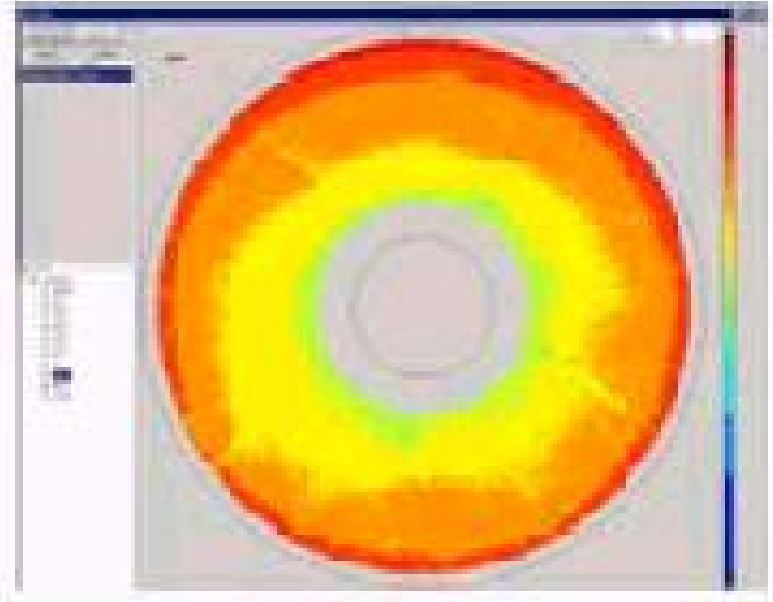
*Une très grande vitesse de rotation
peut provoquer l'éclatement du disque
dans le lecteur-enregistreur !*

Des mauvaises conditions de stockage ou d'utilisation peuvent aboutir à une destruction rapide des supports :

- Mauvais environnement (température, humidité, lumière, poussière, champ magnétique)
- Erreurs de manipulation (rayures, produits toxiques, effacement accidentel)

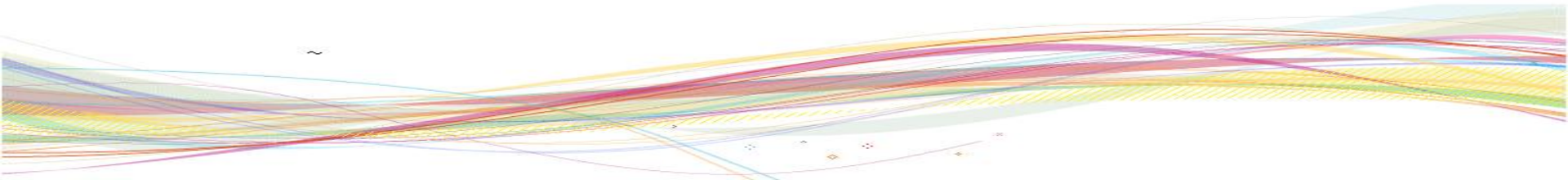


Les supports se dégradent




Quelles que soient les conditions de conservation des supports, ceux-ci se dégradent naturellement.

En général, la durée de vie des supports ne dépasse guère 5 ou 10 ans.



Les supports passent...

Adresse  <http://www.library.cornell.edu/iris/tutorial/dpm/oldmedia/disks.html>

Chamber of Horrors: Obsolete and Endangered Media

Disks

Disk Media

Click for a larger view:



8" Floppy

Storage capacity: 100KB - 1+MB

Approximate dates in use: Late 1971-1981

Comments: 8" X 8", 1 5/16" hole in center.

Read about the [decline of the floppy disk](#)

Read about the [history of the floppy disk](#)



5.25" Floppy

Storage capacity: 100KB - 1.2MB

Approximate dates in use: 1972-mid-1980s

Comments: Primary mass storage medium during the early days of microcomputing, including the Apple II, TRS-80, Commodore 64, and the original IBM PC.



3.5" Floppy

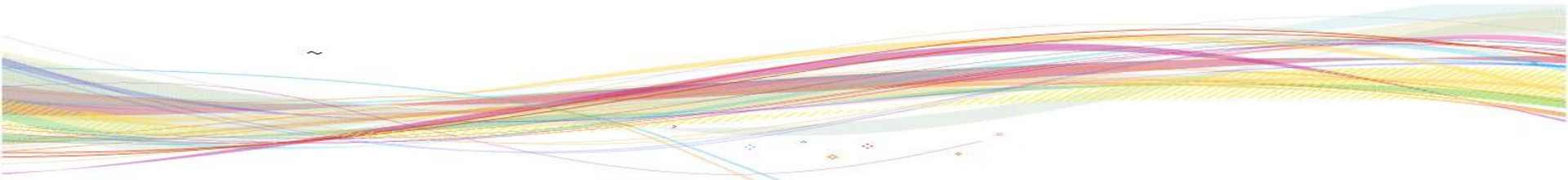
Storage capacity: 400KB - 2.8MB

Approximate dates in use: 1982-present

Comments: 3½" X 3½" (actually 90mm X 94mm). From 1995 to 2001 floppy-disk sales declined from 5 billion to 1.4 billion.

La durée de fabrication des supports et des appareils de lecture est limitée.

Le format d'enregistrement des données sur le support est souvent propre à un fabricant.

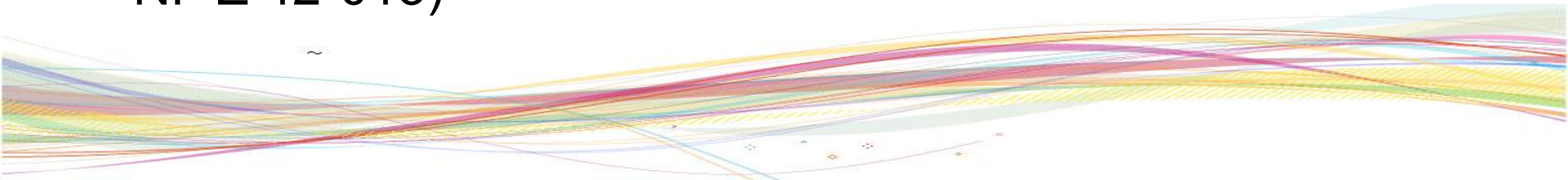


Que faire ?

- Choisir des supports robustes, normalisés, produits par plusieurs fabricants
- Créer plusieurs copies des informations
- Surveiller les conditions de stockage
- Surveiller régulièrement l'état des supports
- Surveiller l'état du marché

Recopier régulièrement les supports sur des supports de même type ou de type nouveau.

Ces opérations doivent être décrites en détail (cf. norme NF Z 42-013)



L'action du Ministère de la Culture sur les CD-R

- En 2005 : **Recommandations de la DAF sur la gravure, la conservation et la surveillance des CD-R**

<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2005.00>

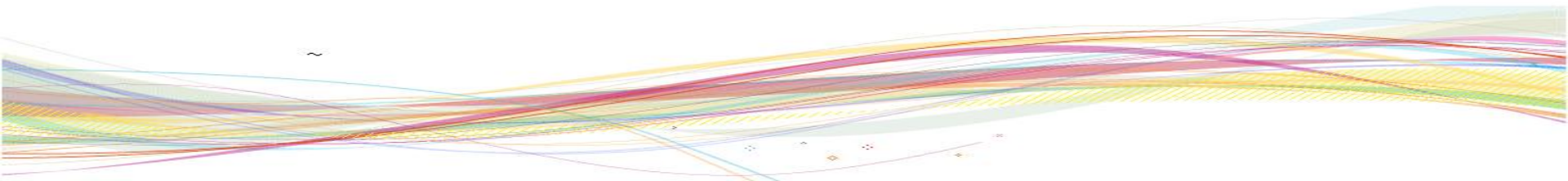
Choix du CD-R (métal, dye, capacité...)

Choix du graveur

Mode de gravure (vitesse, contrôle...)

Conditions de stockage et de manipulation

Surveillance et recopie



L'action du Ministère de la Culture sur les CD-R

- En 2005 : **étude sur l'état de conservation des CD-R** conservés par les services publics d'archives (22 AD, dont AD des Vosges)

<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2006.003.pdf>

20 % des disques présentent un risque

Résultats mitigés pour les ateliers de numérisation internes

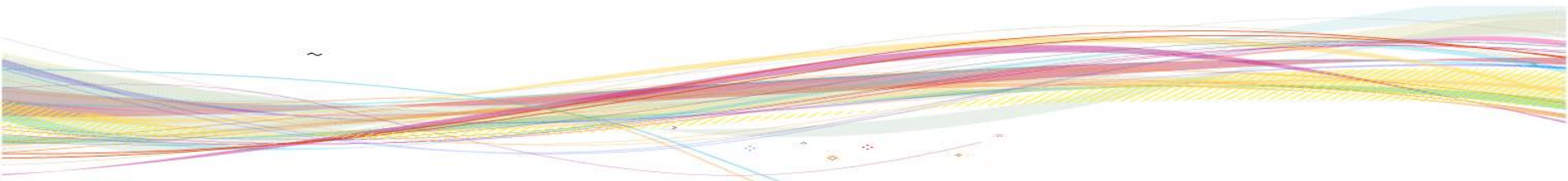
=> Recommandation de migration globale

- En 2007 : enquête de la DLL sur la conservation des collections numérisées dans les bibliothèques municipales (20 BM, dont Saint-Dié-es-Vosges)

80 % des collections sont stockées sur CD-R

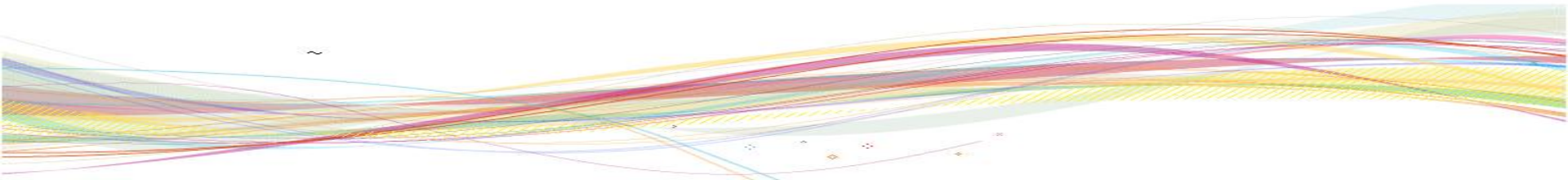
- 2007-2008 : étude DAF-LNE sur des DVD-R

2. Les formats



L 'information numérique

001111000100001000111110001111000100011001
001111010011100101010000100000010000110100
111101001100010011110101001000111101001000
100010001101000110010001100011011000110110
001100000011000000100010001111100100001001
101111011011100110101001101111011101010111
001000100000001000010011110000101111010001
100100111101001110010101000011111000111100
001011110100001000111110

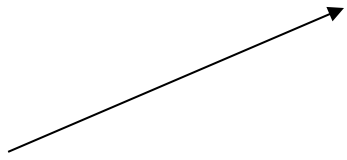


Le codage ASCII

00111100 = 60 = <



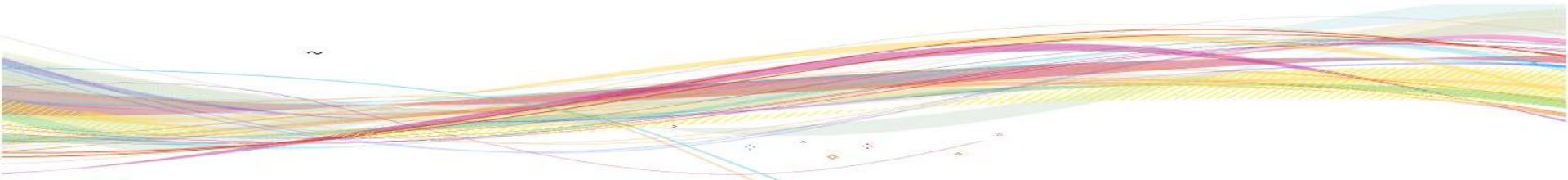
01000010 = 66 = **B** etc.



0011110001000010001111100011110001000110010011110100111
0010101000010000001000011010011110100110001001111010100
1000111101001000100010001101000110010001100011011000110
1100011000000110000001000100011111001000010011011110110
1110011010100110111101110101011100100010000000100001001
1110000101111010001100100111101001110010101000011111000
111100001011110100001000111110



Bonjour !



Le langage HTML

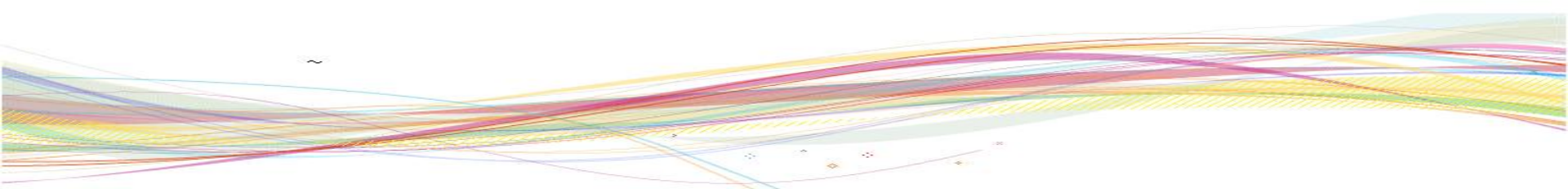
`` = Bold = **Caractères gras**

``

`Bonjour !`

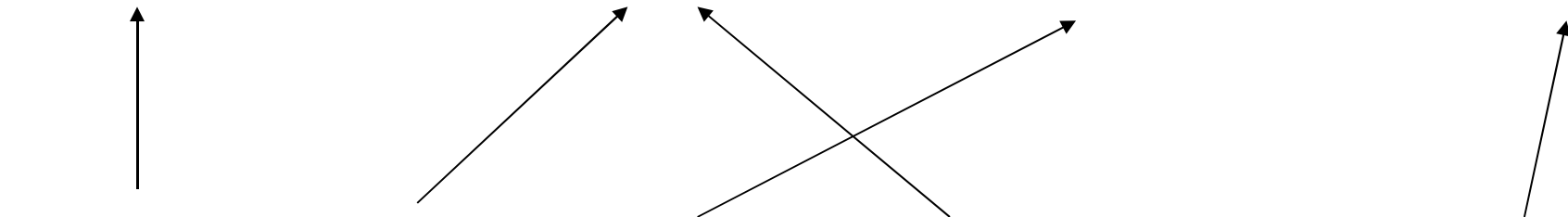
``

``

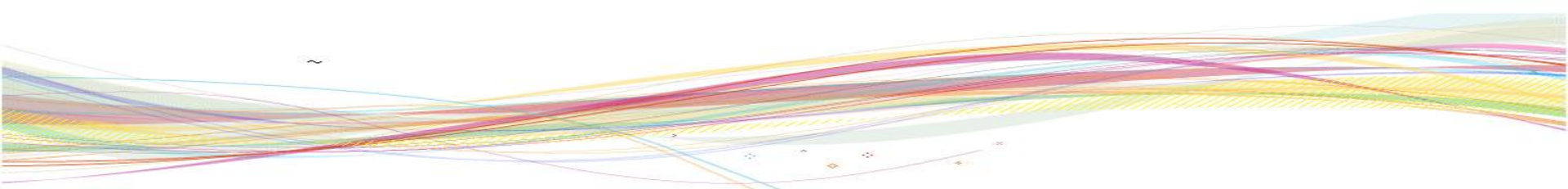


Autre exemple : un fichier JFIF

01001010 = **J** 01000110 = **F** 01001001 = **I** 1.1

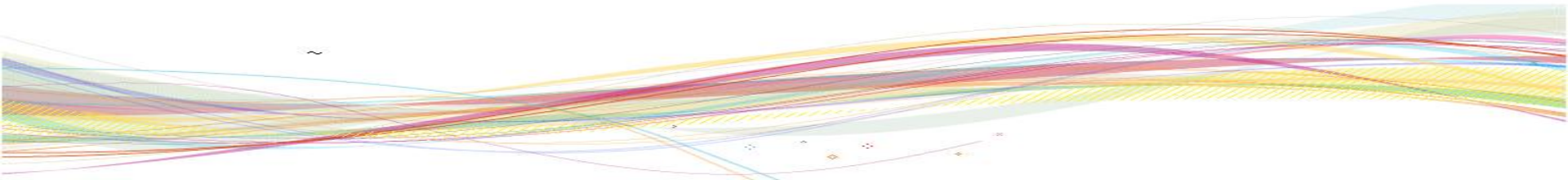


```
1111111111011000111111111110000000000000000010000
01001010010001100100100101000110000000000000010
0000001000000010000000011011000000000101101100000
00000000000000
```



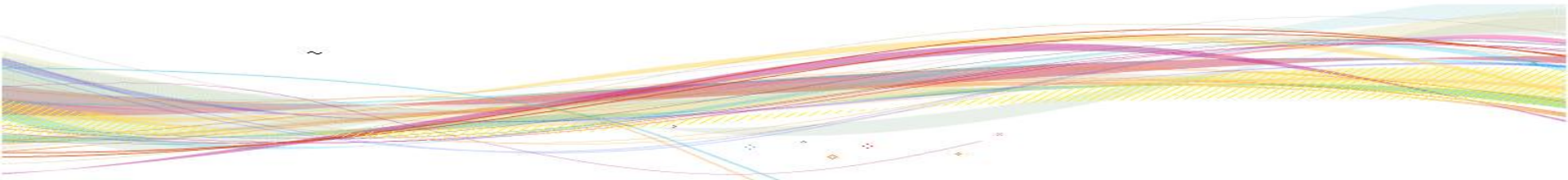
Les formats sont des systèmes de codage de l'information, qui permettent de passer d'informations intelligibles à des données binaires, et vice-versa.

Ces opérations de codage et de décodage sont généralement assurées par des logiciels (par ex. Photoshop, Dreamweaver, Mozilla...).

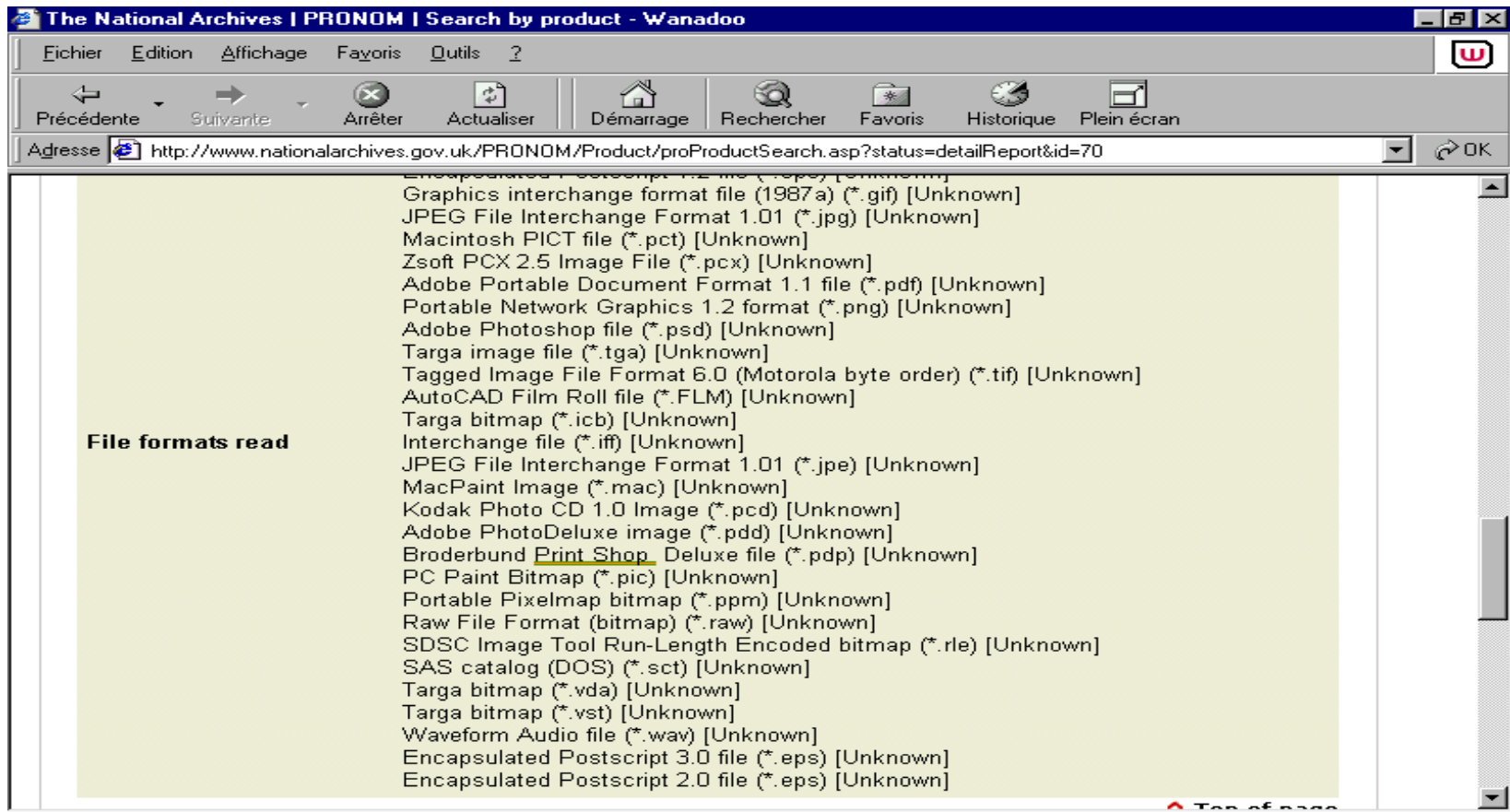


Si les logiciels évoluent ou disparaissent, le sens des données binaires stockées risque de devenir inaccessible.

C 'est le cas lorsque le processus de codage est secret et que seul le logiciel d 'origine est capable de le réaliser.



Certains formats sont rares

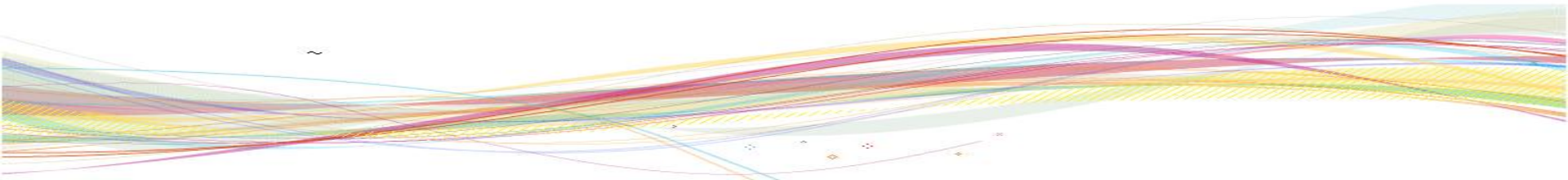


Que faire ?

Il est indispensable d'utiliser des formats dont les spécifications sont publiques, et de préférence reconnus comme normes internationales.

Il est également conseillé d'utiliser des formats courants.

Des migrations seront nécessaires si les documents sont à conserver longtemps.



Le référentiel général d'interopérabilité (RGI)
préparé par la DGME avec tous les ministères.

Pour la conservation

- Images :

JPEG, TIFF

- Sons :

MP3, WAVE

- Vidéo :


MPEG-2, MPEG-4

- Documents :

Open Document Format, PDF/A



La base PRONOM



The screenshot shows a web browser window titled "The National Archives | PRONOM | Search by format - Wanadoo". The address bar contains the URL: <http://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.asp?status=new>. The browser's menu bar includes "Fichier", "Edition", "Affichage", "Favoris", and "Outils". The toolbar contains icons for "Précédente", "Suivante", "Arrêter", "Actualiser", "Démarrage", "Rechercher", "Favoris", "Historique", and "Plein écran".

The website's navigation bar is red and contains the following links: Home, About us, Visit us, Exhibitions & Learning online, Getting started, Search our collections, Search other archives, Services for professionals, News, Shop, and Business services.

The breadcrumb trail reads: "You are here: [Home](#) > [Services for professionals](#) > [Preservation](#) > [PRONOM](#) > Search: Search by format".

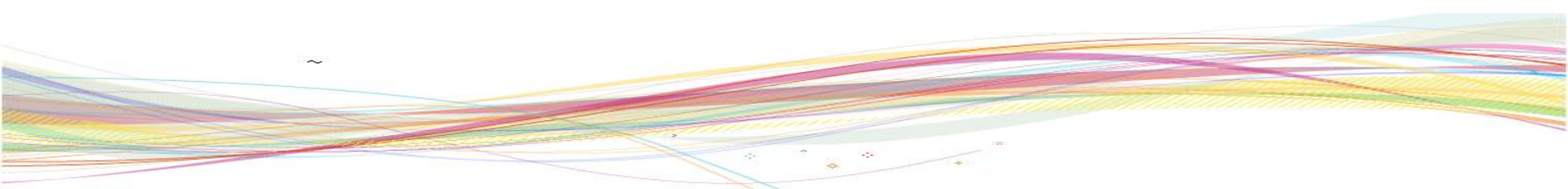
The main content area features a large heading "PRONOM" with the subtitle "The file format registry" and an image of a CD and a folder. To the right, there are links for "Welcome", "Search", "About", and "Help". A "Help : search by format" link is also present.

The search interface is titled "Search : By format" and includes a search icon. Below this, there are tabs for "File format", "Product", "Vendor", "Support Period", and "Release date".

1. File formats
Enter a file extension and click 'search' to find all file formats with that extension. Leave the file extension blank to find all file formats in the database.
*

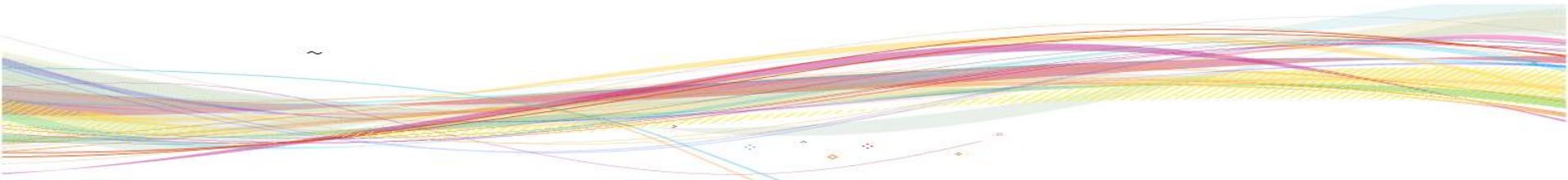
2. Compatible products
Enter a file extension and click 'search' to find products which can read or write files with that extension.
*

3. L'intégrité



Des altérations intentionnelles

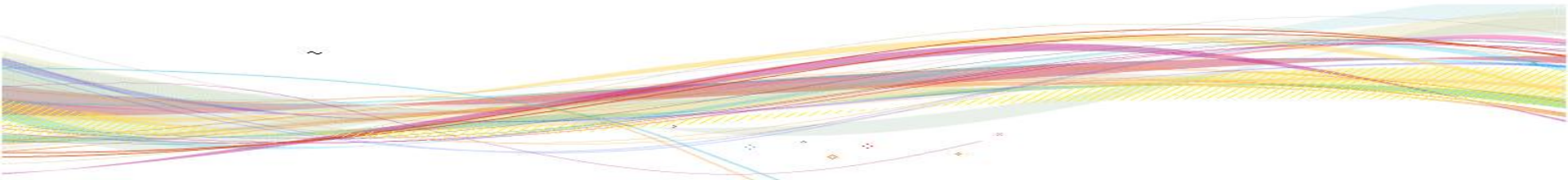
- Modification volontaire d'une donnée dans une base de données
 - Modification d'un message électronique transféré pour cacher certaines informations
 - Modification de la date ou de l'auteur d'un document
- Etc.



Des altérations non intentionnelles

- Erreur sur la date d'un document bureautique par l'utilisation de la fonction «Date du jour»
- Erreur sur la date d'un message si l'heure du serveur de messagerie est fausse
- Conversion automatique d'un document d'une version ancienne d'un format de fichier dans une version plus récente

Etc.

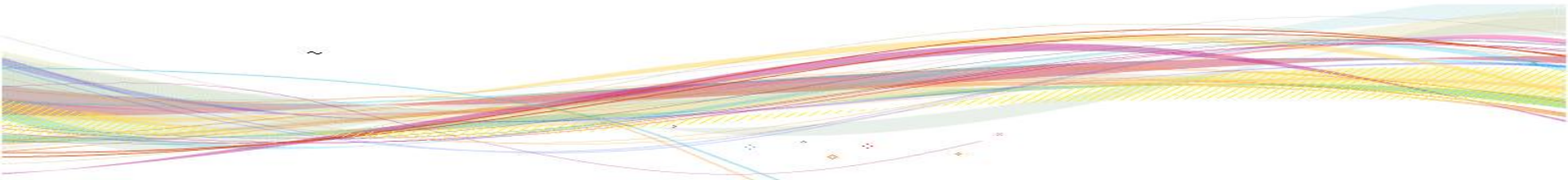


Que faire ?

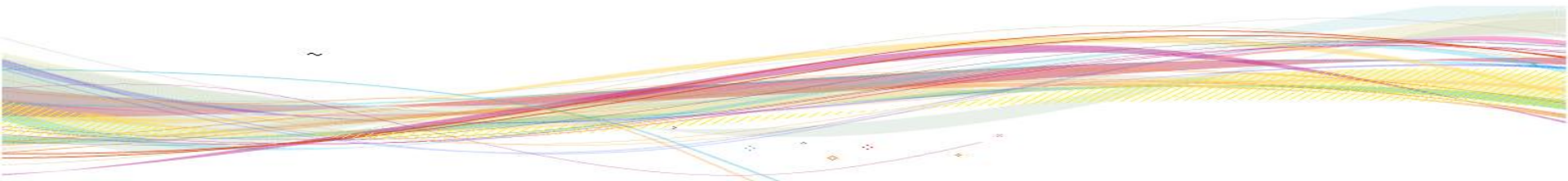
Utiliser des supports permettant d'enregistrer les informations de manière irréversible.

Utiliser des méthodes permettant de vérifier l'intégrité (empreintes), l'horodatage (jetons), l'identité de l'auteur (certificats) d'un document.

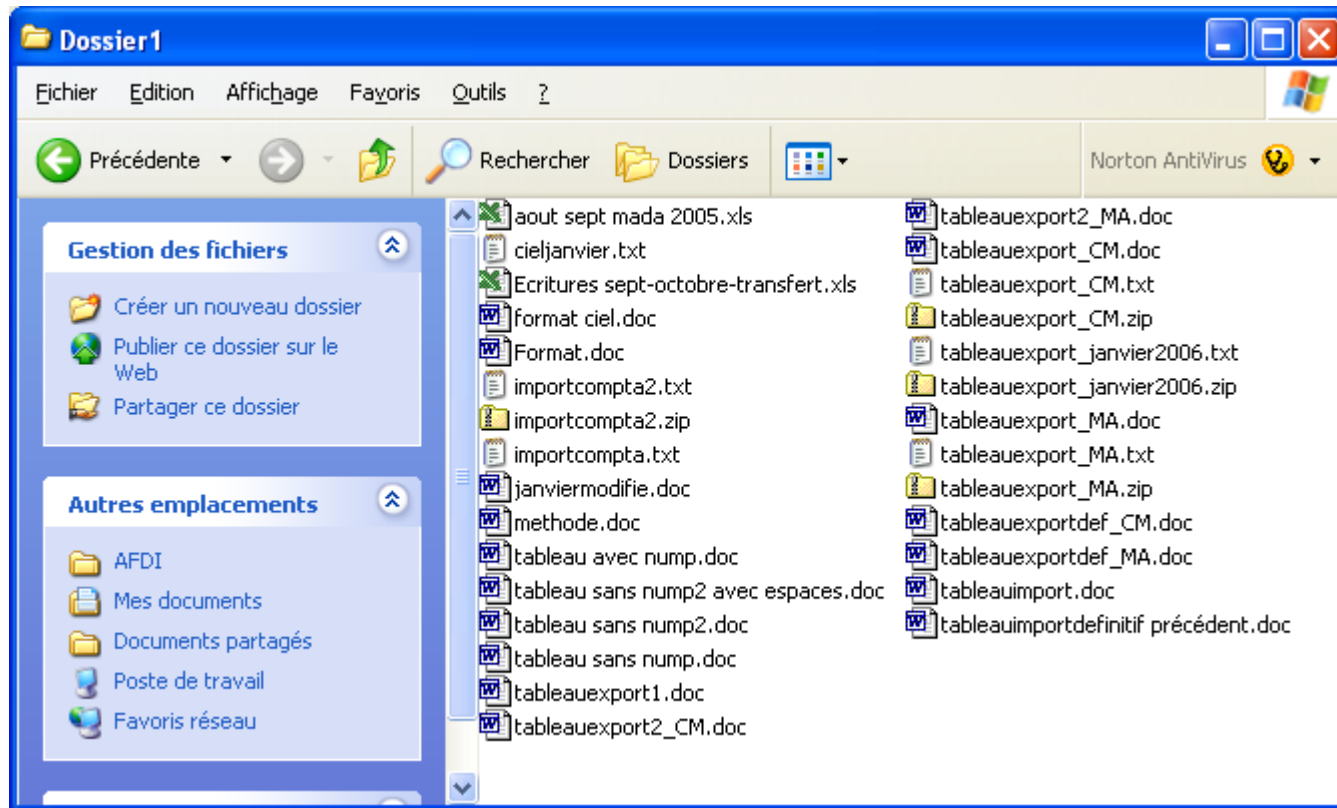
Fixer des règles internes de gestion des documents : bonnes pratiques, droits d'accès...



Accompagner les documents
pendant tout leur cycle de vie



Une prise en charge précoce

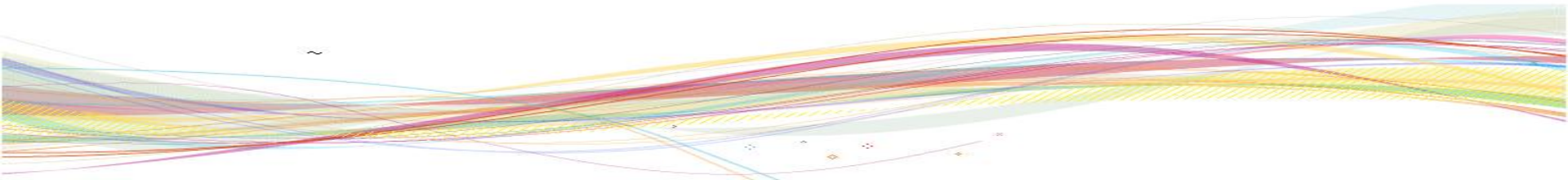


Une prise en charge précoce

Il est difficile de classer ou trier des documents électroniques a posteriori.

Le plan de classement, les règles de nommage, les règles de gestion (durée de conservation, confidentialité...) doivent être définis dès le départ.

Notion de « records management ».

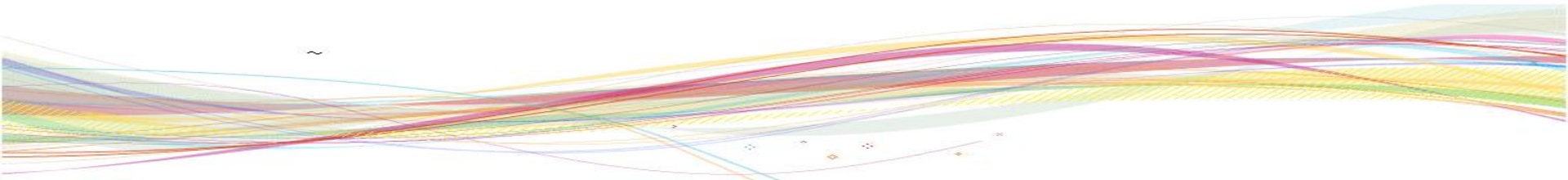


Le rôle central des métadonnées

Les métadonnées sont des données sur les données.

Elles permettent de :

- retrouver les documents (métadonnées de contenu) ;
- gérer le cycle de vie et les droits d'accès aux documents (métadonnées administratives) ;
- connaître les caractéristiques techniques des documents en vue de migrations (métadonnées techniques) ;
- enregistrer les traitements effectués sur les documents.

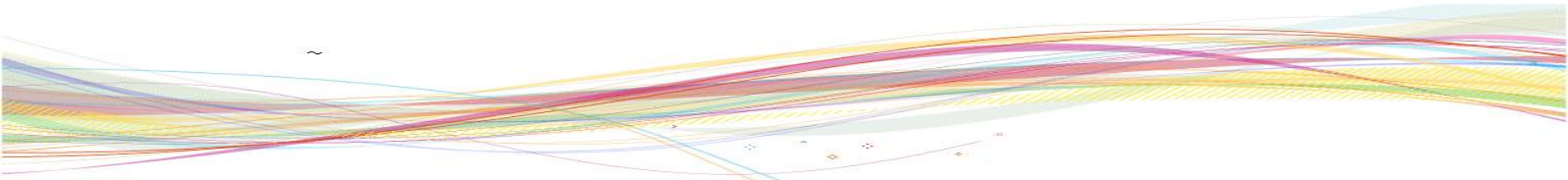


Une veille permanente

- Surveillance des supports
- Surveillance des formats
- Surveillance des procédures de gestion

Un document électronique qui dort court le risque d'être rapidement illisible.

La conservation est une démarche active.

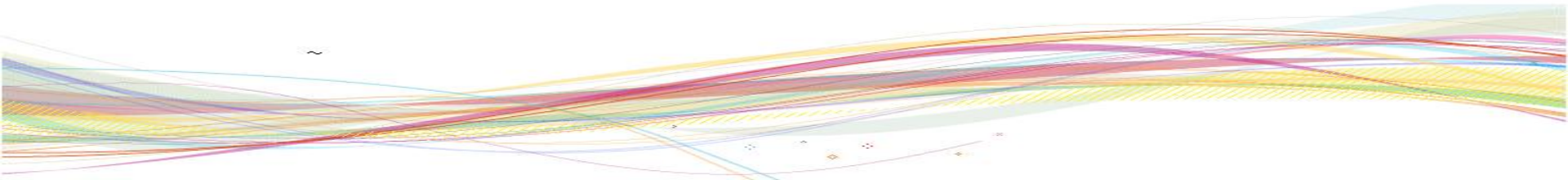


L'importance de l'organisation

Une bibliothèque / un service d'archives doit définir :

- les fonctions à remplir par les différents acteurs : entrées, gestion de données descriptives, stockage, administration, accès, planification de la pérennisation.
- les règles organisationnelles et techniques à suivre.

Ces éléments constituent une **politique d'archivage ou de conservation numérique.**



Politique de conservation numérique

Dans le cas de la numérisation, une politique de numérisation abordera par exemple :

- le nommage,
- les formats (avec le maximum de précision),
- le stockage (distinction support de livraison, support de stockage),
- la migration,
- l'outil de suivi.

Elle précisera clairement les responsabilités des acteurs (bibliothèque ou archives, service informatique...).

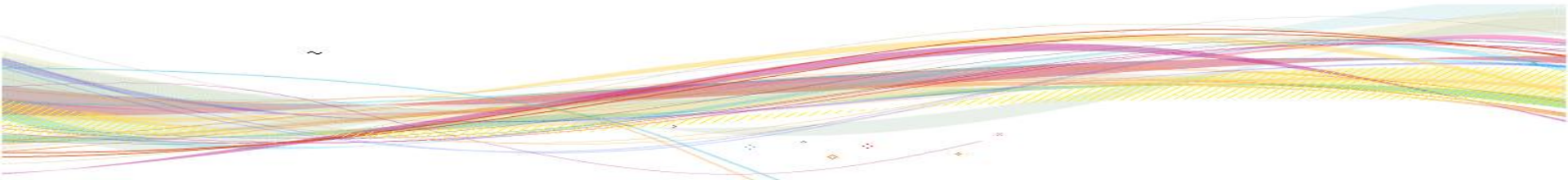


Des concepts utiles : le modèle OAIS

Le modèle OAIS (Open Archival Information System) –
norme ISO 14721

Un ensemble de concepts sur :

- l'information numérique et les métadonnées nécessaires à sa préservation ;
- les différentes fonctions à remplir pour assurer la conservation.



Conclusion

La sensibilisation sur le thème de la conservation numérique progresse (relevé dans l'enquête DLL).

Aspect pris en compte dans le plan national de numérisation.

Intérêt de la mutualisation.

- Au niveau de l'Etat (PILAE, SPAR, CINES...)
 - Au niveau d'une collectivité (cf. enquête DLL)
 - Entre plusieurs collectivités (ex. AD Yvelines / BM Versailles / autres communes).
- 

olivier.de-solan@culture.gouv.fr

