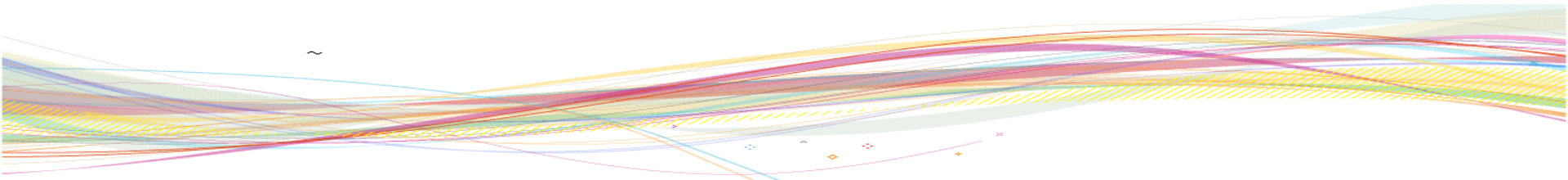


Conserver les documents numériques

Olivier de Solan

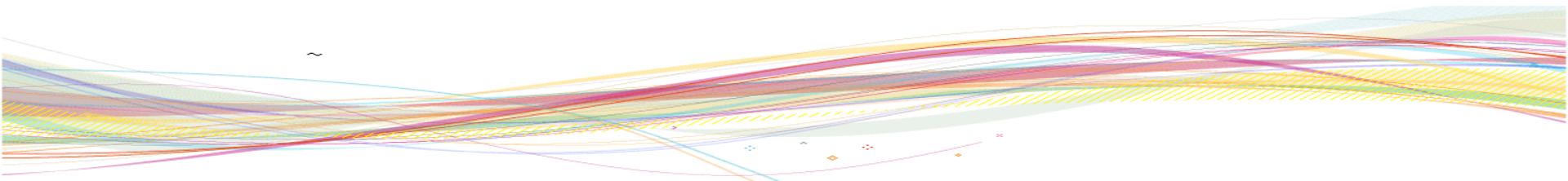
Direction des archives de France

19 mars 2007

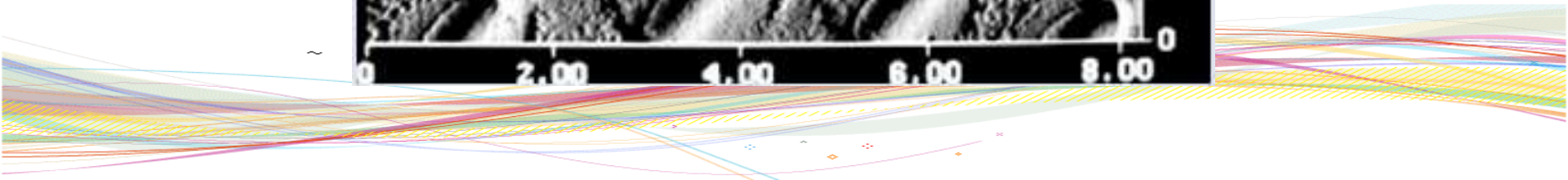
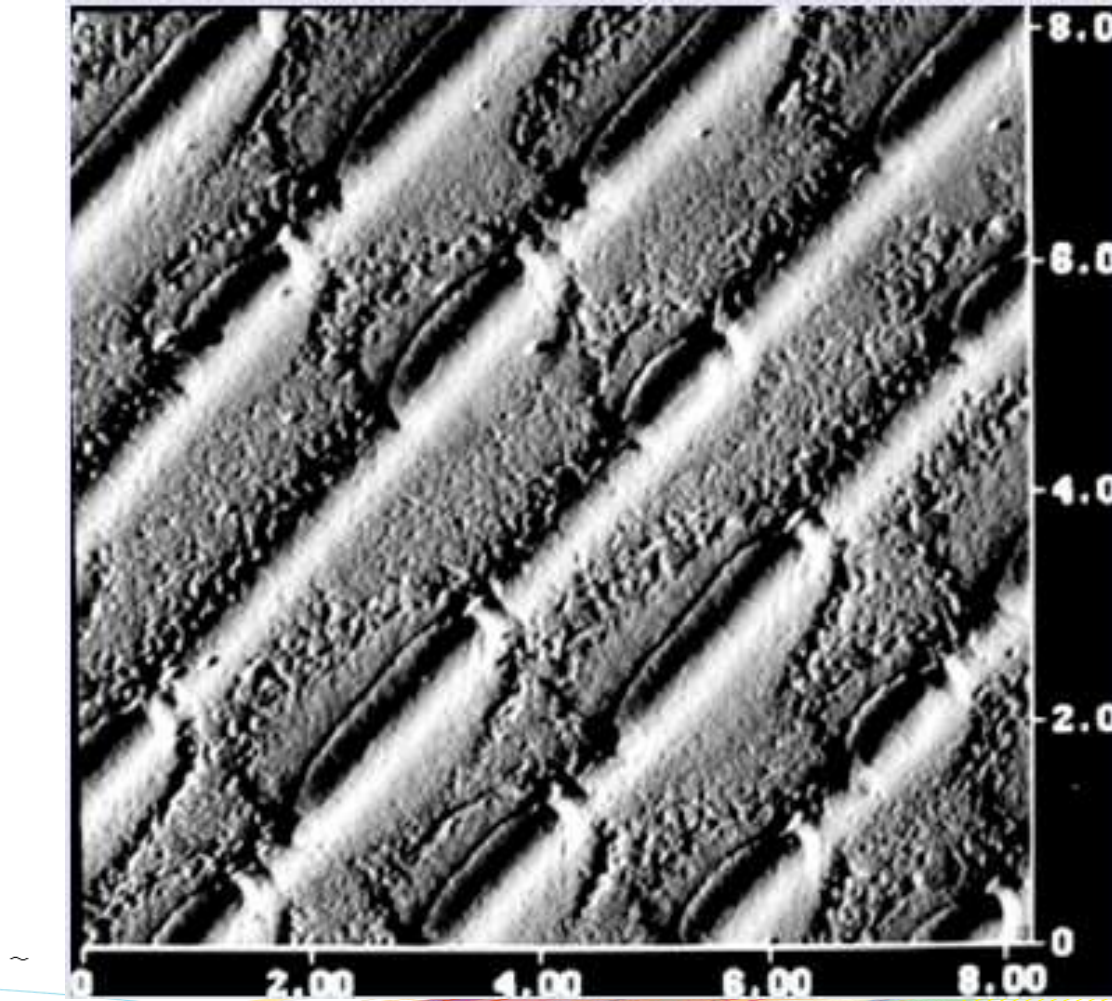


La conservation : une question importante

- La numérisation dans les bibliothèques et les archives a débuté depuis une dizaine d'années
- Un investissement important à sauvegarder
- Certains documents n'existent que sous forme numérique

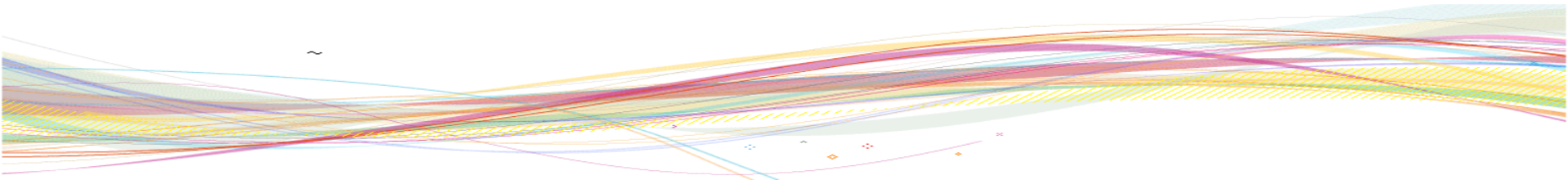


L'information numérique

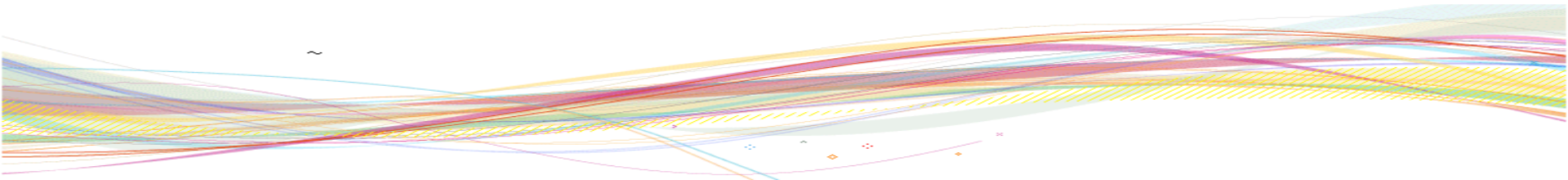


Trois risques majeurs

1. L'obsolescence des applications informatiques et des **formats** de fichiers ;
2. La fragilité et la durée de vie relativement faible des **supports** ;
3. La difficulté de retrouver des fichiers informatiques s'ils n'ont pas été bien **décrits** à l'origine et bien **suivis**.

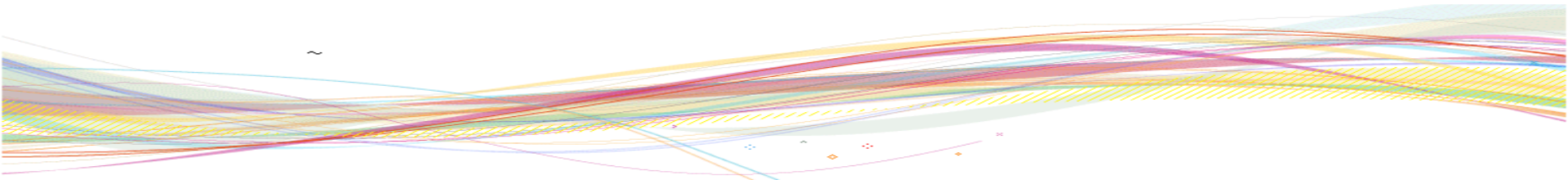
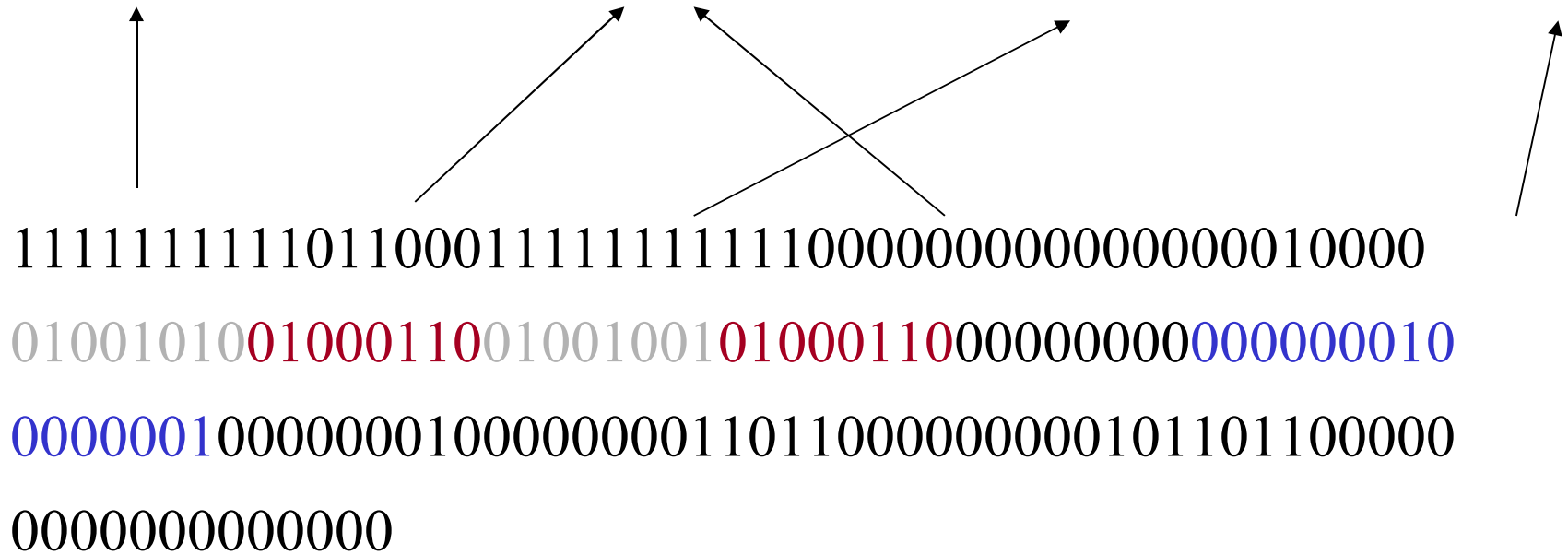


1. Les formats



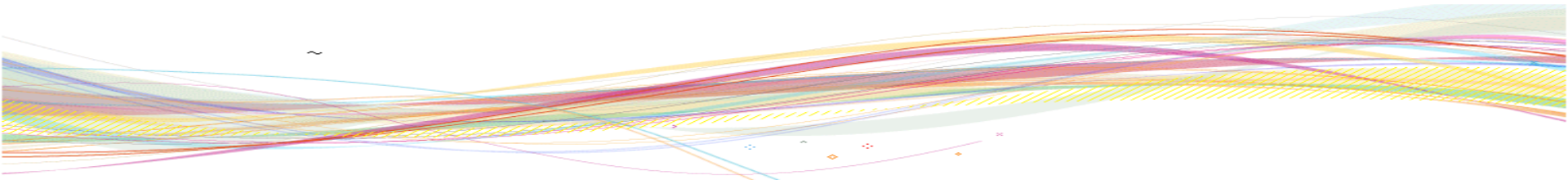
Le codage JFIF

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



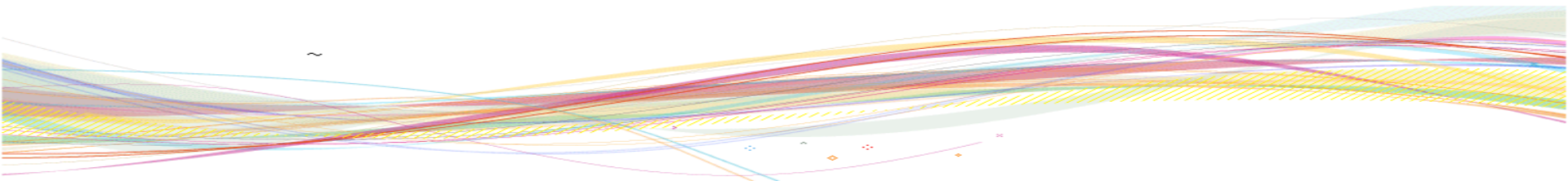
Les formats sont des systèmes de codage de l'information, qui permettent de passer d'informations intelligibles à des données binaires, et vice-versa.

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

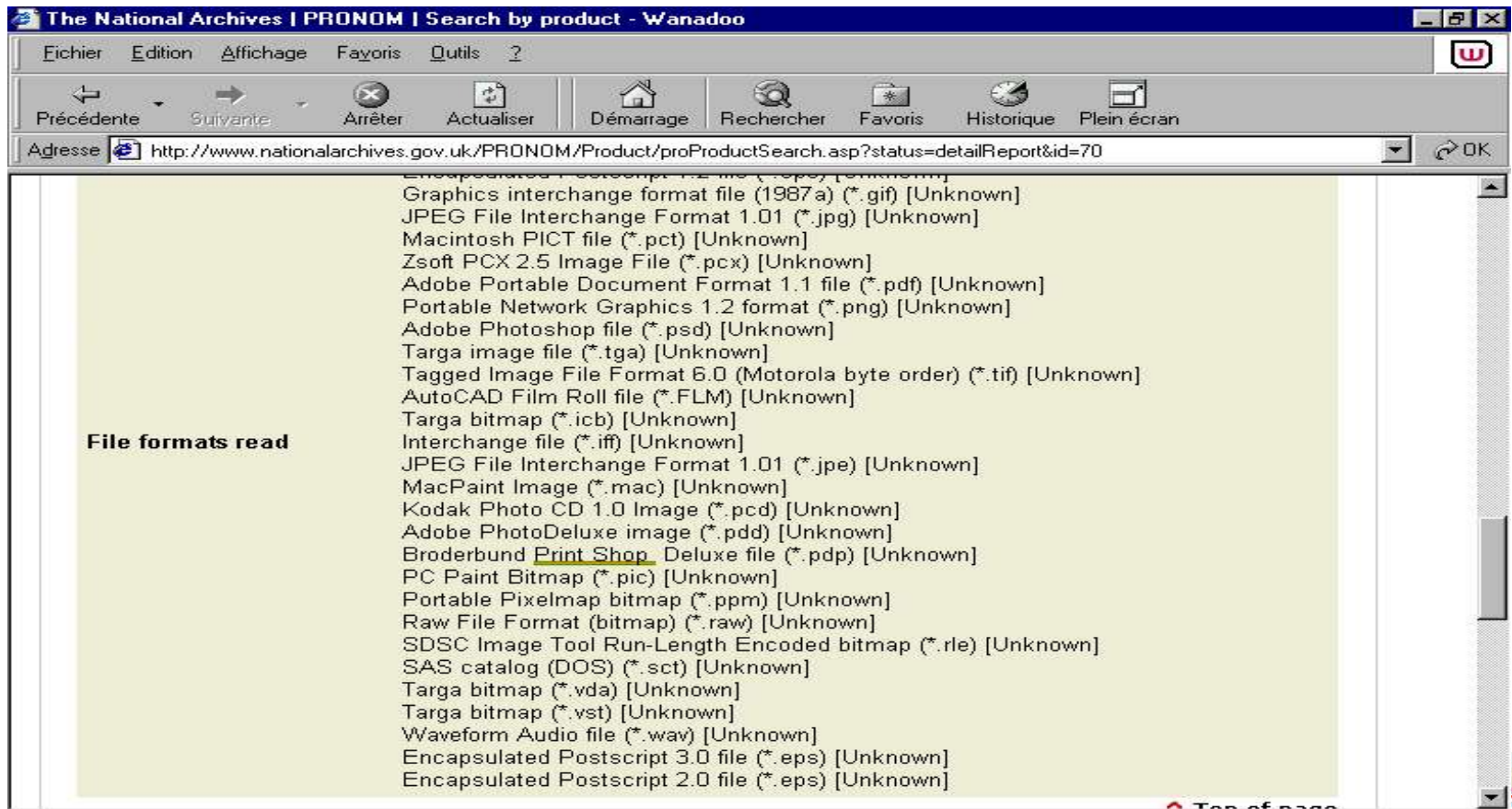


Si les logiciels évoluent ou disparaissent, le sens des données binaires stockées risque de devenir inaccessible.

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

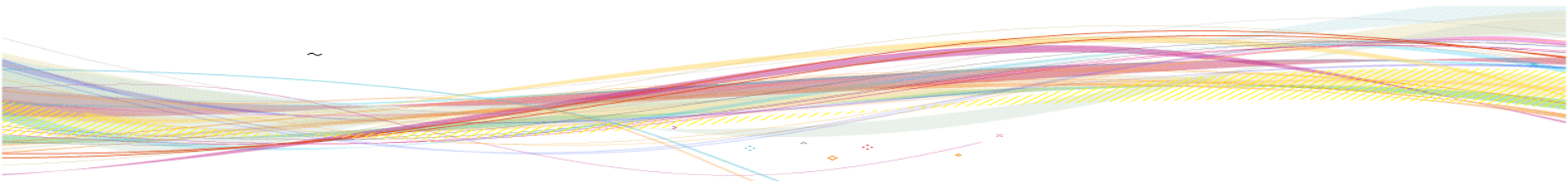


Que faire ?

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

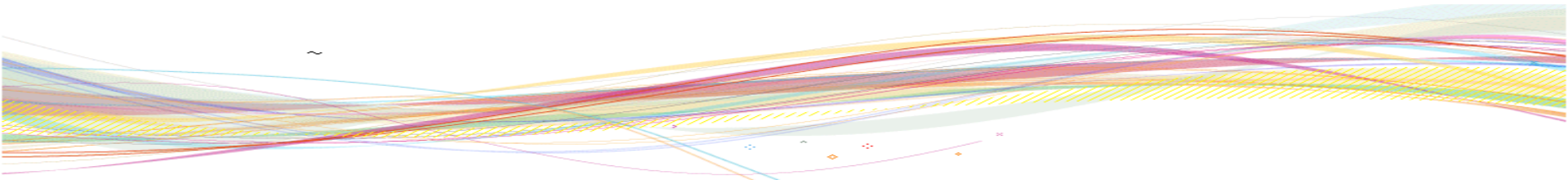
Il est également conseillé d'utiliser des formats courants.

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



Le format TIFF

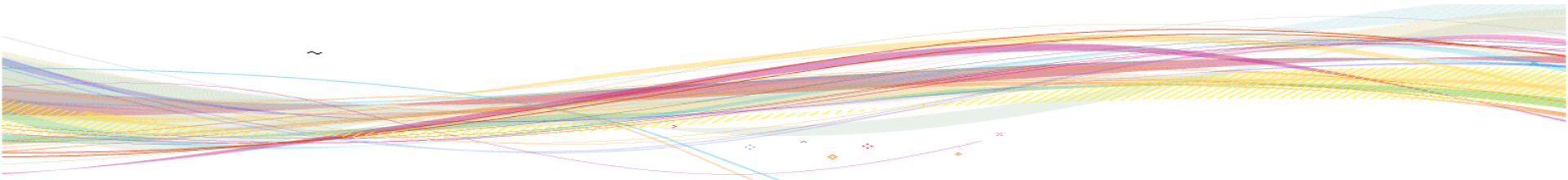
- TIFF : Tagged Image File Format
- Format matriciel pour images en noir et blanc ou couleur (multipage possible)
- Supporte divers algorithmes de compression (ex : LZW, CCITT Groupe 4...)
- Créé par la société Aldus, appartient actuellement à Adobe mais spécifications publiques (<http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf>)
- Très utilisé



Le format JPEG

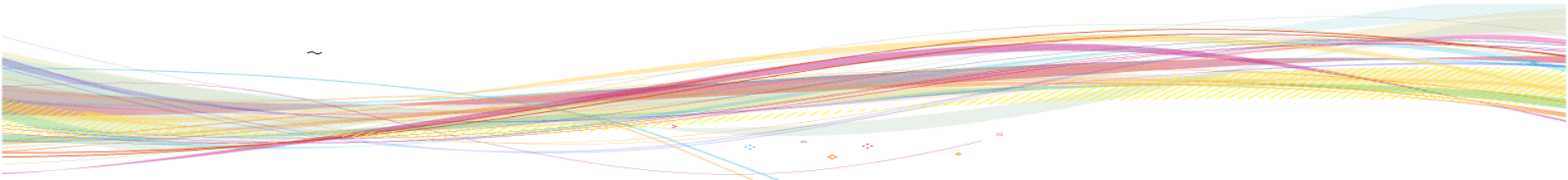
- JFIF : JPEG File Interchange Format
- Algorithme de compression JPEG créé par le Joint Photographic Experts Group ISO/CCITT en 1990 (ISO 10918-1)
- Non défini dans la norme JPEG mais spécifications publiques (<http://www.jpeg.org/public/jfif.pdf>)
- Très utilisé

~



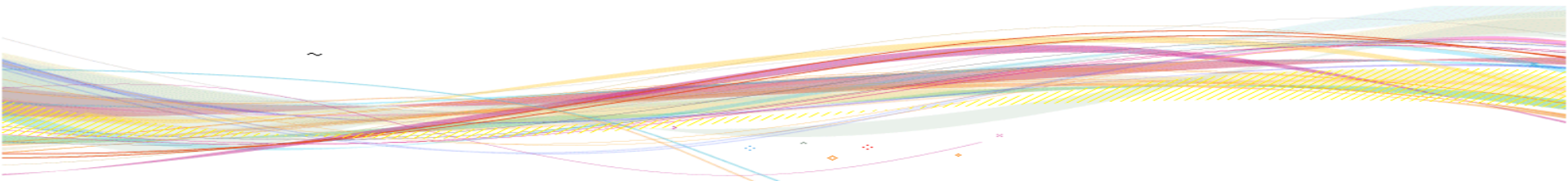
Le format JPEG 2000

- Permet la compression sans pertes, l'intégration de métadonnées XML
- Format normalisé ISO 15444 (2004)
- Encore peu utilisé



Le format PNG

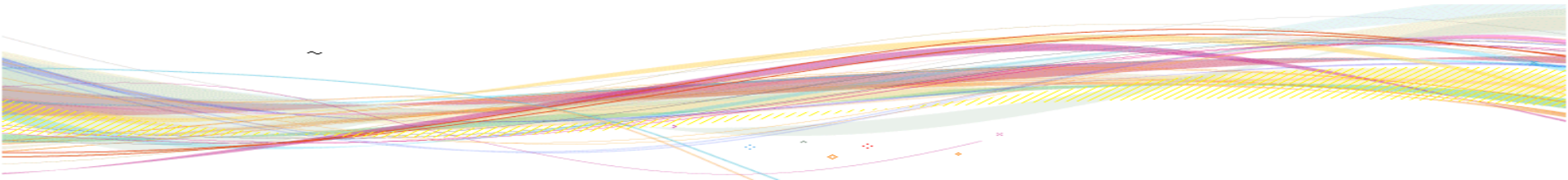
- PNG : Portable Network Graphics
- Format matriciel noir et blanc ou couleur, supportant un algorithme de compression sans perte
- Format recommandé par le W3C et spécifications publiques (<http://www.w3.org/TR/REC-png.pdf>)
- Encore assez peu utilisé, mais en progression rapide



Le format PDF

- PDF : Portable Document Format
- Format destiné à maintenir la représentation de documents sur n'importe quelle plate-forme, pouvant contenir de nombreux éléments (texte, images, son, vidéo...)
- Format développé par la société Adobe mais spécifications publiques
- Très utilisé

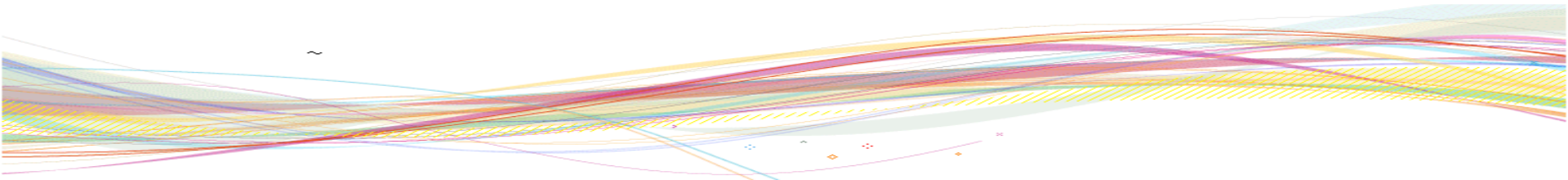
~



Le format PDF/A

- PDF/A : Portable Document Format / Archival
- Sous-ensemble du format PDF 1.4 destiné à une conservation à long terme (certaines fonctionnalités avancées ont été exclues)
- Norme ISO 19005-1 (2005)
- Encore peu utilisé, mais progression rapide. Des outils commencent à exister (ex. Acrobat 8)

~



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

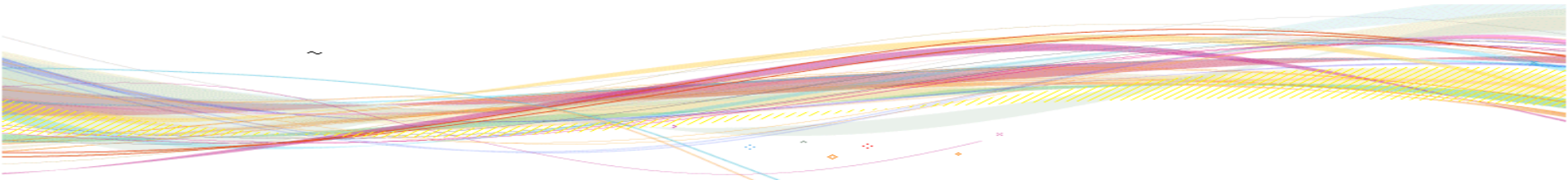
- Documents :

Open Document Format, PDF/A

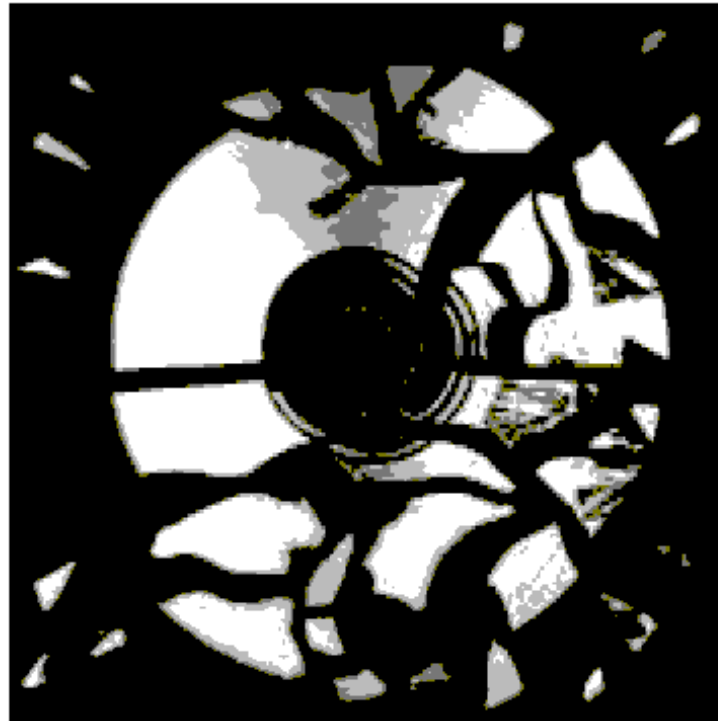
~



2. Les supports



Les supports sont fragiles



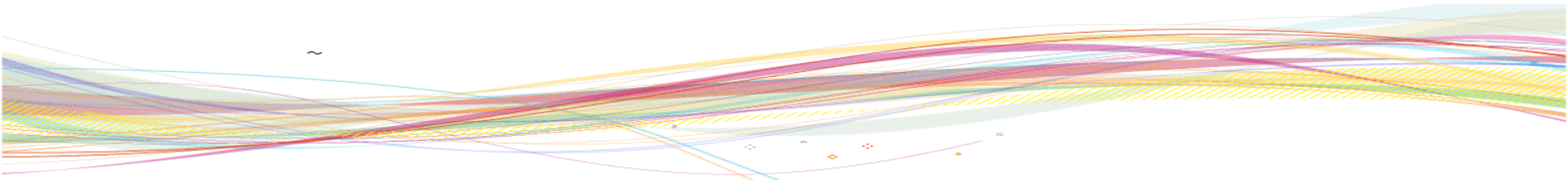
© LAM - JMF

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

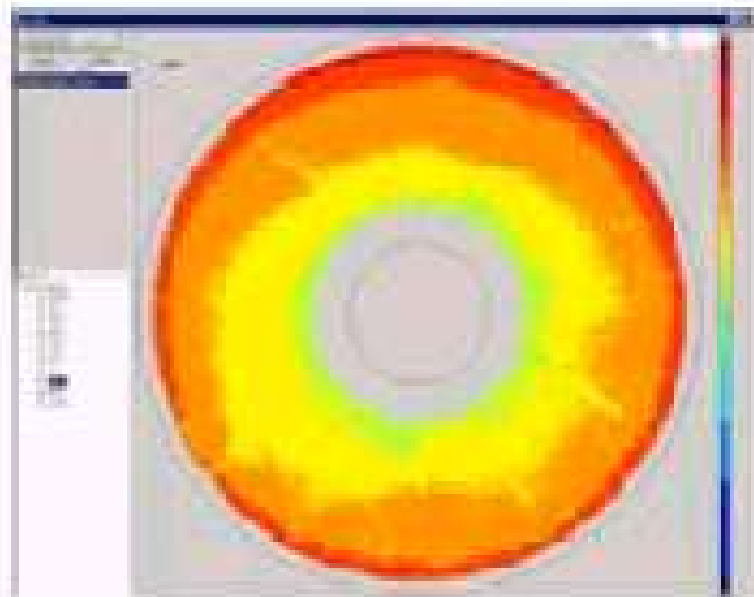
~

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)



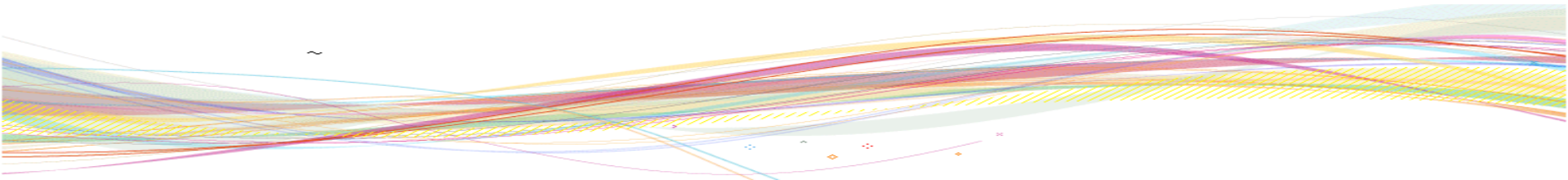
Les supports se dégradent



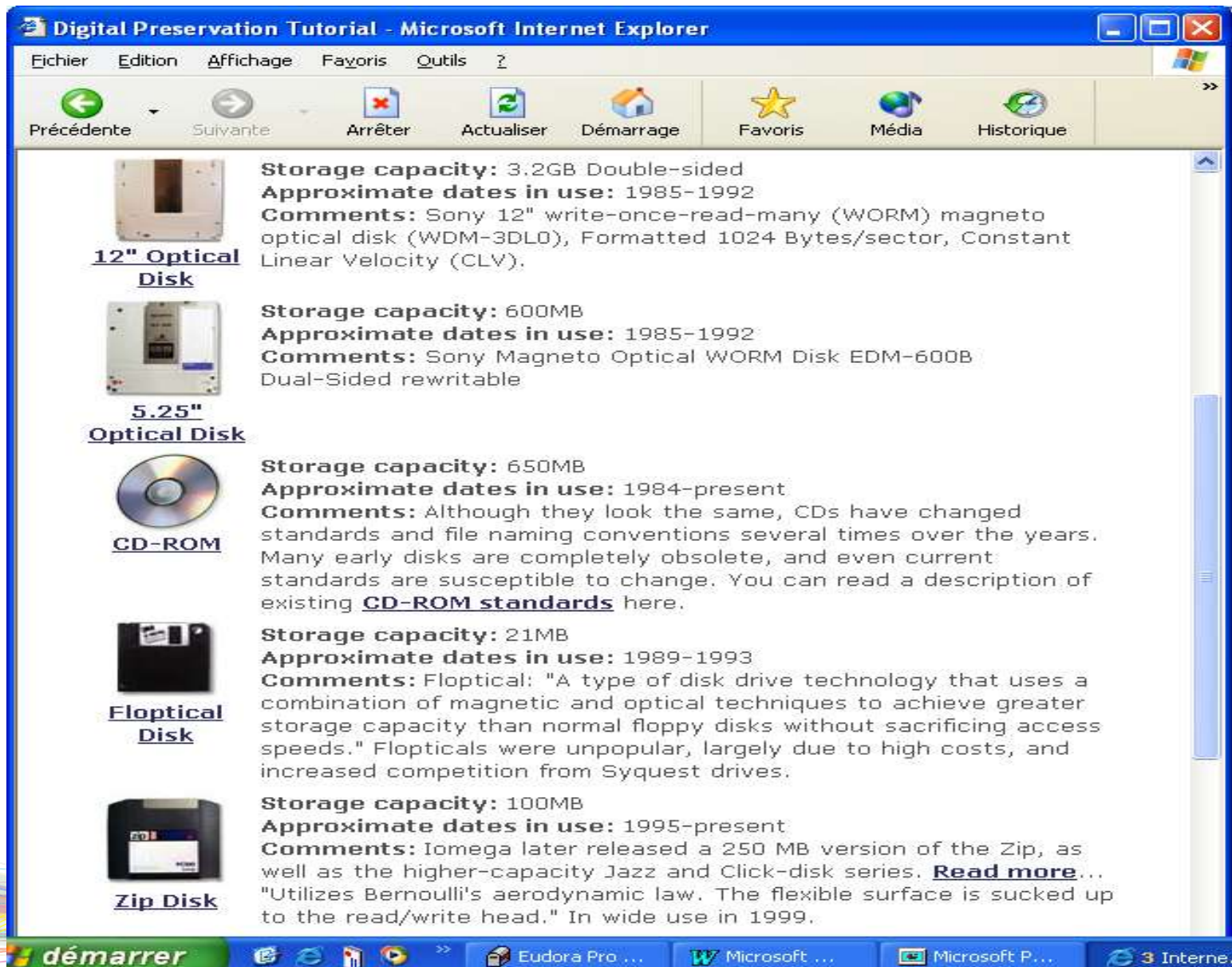
~

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...





Digital Preservation Tutorial - Microsoft Internet Explorer


Fichier Edition Affichage Favoris Outils ?


Précédente Suivante Arrêter Actualiser Démarrage Favoris Média Historique


12" Optical Disk
Storage capacity: 3.2GB Double-sided
Approximate dates in use: 1985-1992
Comments: Sony 12" write-once-read-many (WORM) magneto optical disk (WDM-3DL0), Formatted 1024 Bytes/sector, Constant Linear Velocity (CLV).


5.25" Optical Disk
Storage capacity: 600MB
Approximate dates in use: 1985-1992
Comments: Sony Magneto Optical WORM Disk EDM-600B Dual-Sided rewritable


CD-ROM
Storage capacity: 650MB
Approximate dates in use: 1984-present
Comments: Although they look the same, CDs have changed standards and file naming conventions several times over the years. Many early disks are completely obsolete, and even current standards are susceptible to change. You can read a description of existing [CD-ROM standards](#) here.

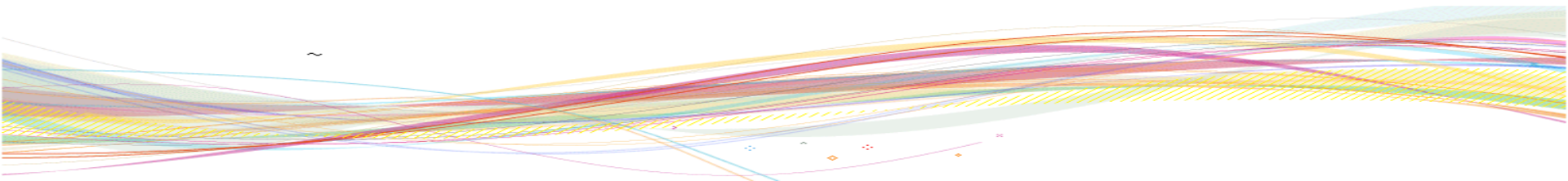

Floptical Disk
Storage capacity: 21MB
Approximate dates in use: 1989-1993
Comments: Floptical: "A type of disk drive technology that uses a combination of magnetic and optical techniques to achieve greater storage capacity than normal floppy disks without sacrificing access speeds." Flopticals were unpopular, largely due to high costs, and increased competition from Syquest drives.


Zip Disk
Storage capacity: 100MB
Approximate dates in use: 1995-present
Comments: Iomega later released a 250 MB version of the Zip, as well as the higher-capacity Jazz and Click-disk series. [Read more...](#) "Utilizes Bernoulli's aerodynamic law. The flexible surface is sucked up to the read/write head." In wide use in 1999.

démarrer Eudora Pro ... Microsoft ... Microsoft P... 3 Interne...

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 ?

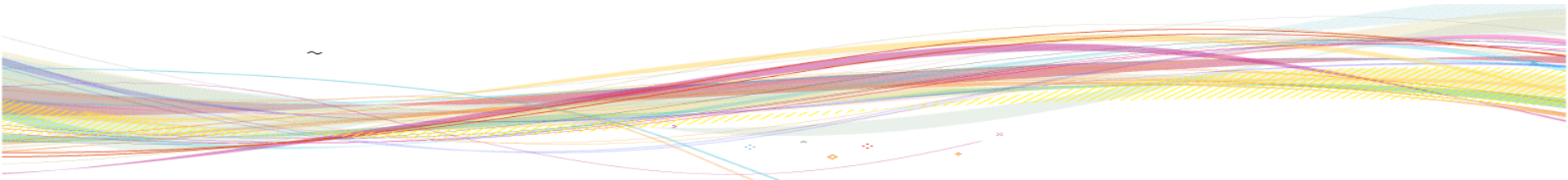
- Choisir des supports robustes, normalisés, produits par plusieurs fabricants
- Créer plusieurs copies des informations, sur des supports différents
- Surveiller les conditions de stockage
- Surveiller régulièrement l'état des supports
- Surveiller l'état du marché

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)

Les grands types de supports

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



Les plates-formes

- Manuelles

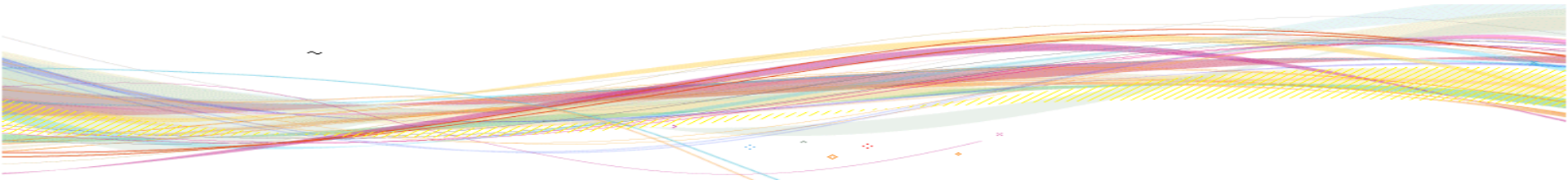
- Automatisées :

Juke-box de CD ou de DVD

Librairie de bandes

Baies de disques

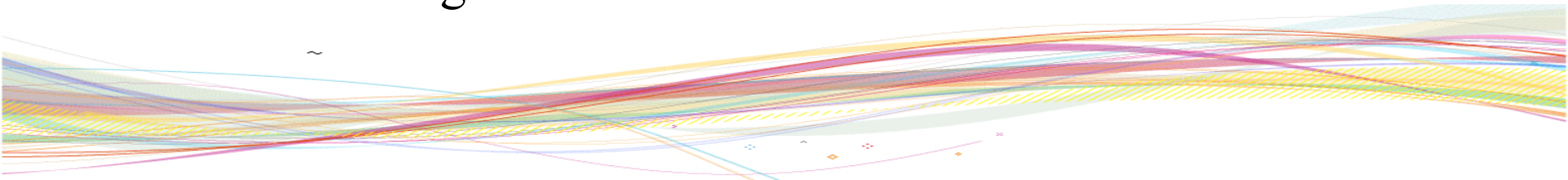
~



Le CD-R

- 700 Mo
- Support grand public, de qualité inégale
- Support très utilisé, risque technologique faible
- Coût du support
 - 50 Go : 120 euros
 - 1 To : 2 500 euros
- Lecteur/enregistreur : 300 euros

~



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.2005.004.rec>

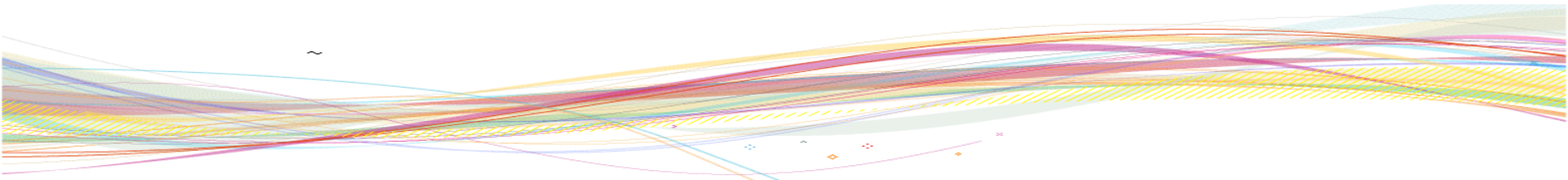
Choix du CD-R (métal, dye, capacité...)

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

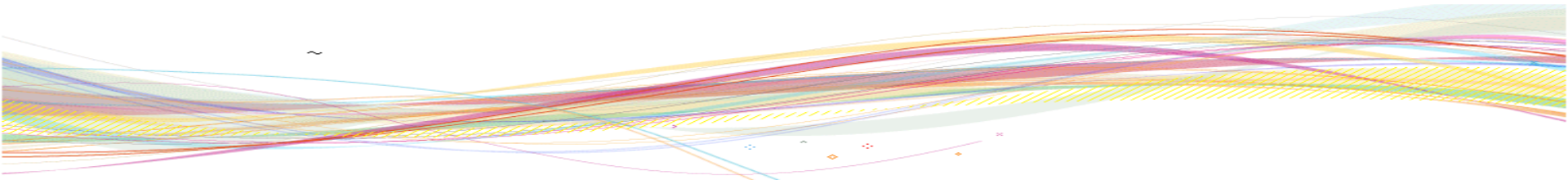
Résultats mitigés pour les ateliers de numérisation internes

=> 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

~



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

- Depuis 2006 : **Observatoire des CD-R du marché** – DAF et LNE

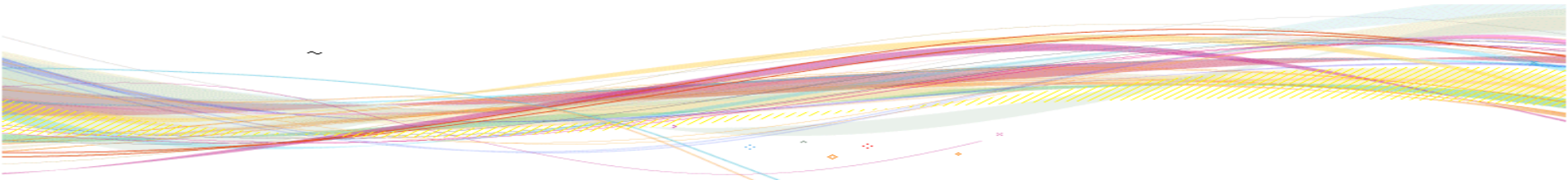
<http://www.archivesdefrance.culture.gouv.fr/fr/circAD/DITN.2006.008.pdf>

Identifier des couples CD-R / graveurs qui donnent des bons résultats (bonne gravure et bonne tenue dans le temps).

7 références de CD-R et 11 références de graveurs testées

Les résultats sont favorables aux disques MPO Gold, Fujifilm CDR 52x et Sony CDQ80NSLD, aux graveurs Plextor PX-755, LG GCE-8527B, HP Dvd840.

~



Le DVD-R

- 4,7 Go
- Support grand public, de qualité très inégale
- Support très utilisé, risque technologique faible

- Coût du support
- 50 Go : 30 euros
- 1 To : 650 euros
- Lecteur/enregistreur : 3 000 euros
- Juke-box pour 1 To : 25 000 euros

2007 : étude DAF / LNE sur la qualité des DVD-R



La bande LTO

- 400 Go
- Support professionnel, de qualité contrôlée
- Support assez utilisé, risque technologique élevé

- Coût du support
- 50 Go : 130 euros
- 1 To : 390 euros
- Lecteur/enregistreur : 3 000 euros
- Librairie pour 1 Go : 8 750 euros

Ex. AD Gironde, AD Haute-Garonne...



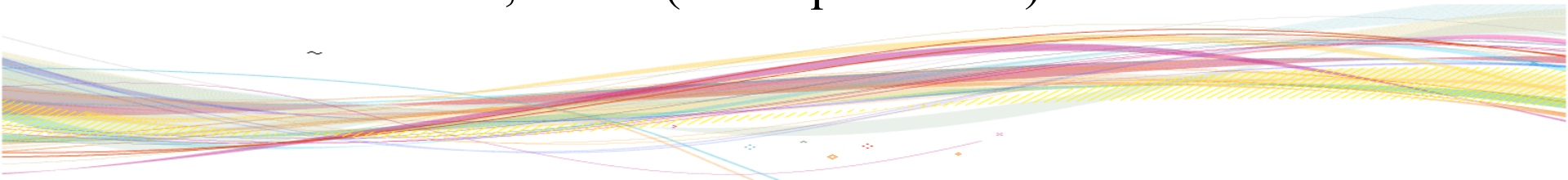
Le disque dur

- 400 Go
- Support professionnel, de qualité contrôlée
- Support très utilisé, risque technologique faible

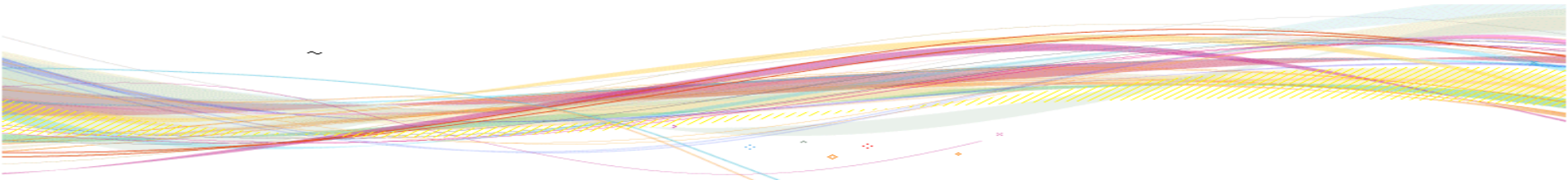
- Coût du support
 - 50 Go : 500 euros
 - 1 To : 1 500 euros
- Baie pour 1 To : 7 500 euros

Ex. AD Yvelines, 7 BM (cf. enquête DLL)

~



3. L'organisation



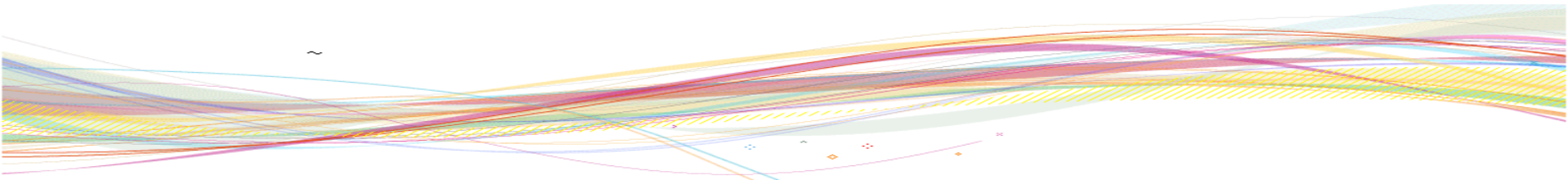
Importance de l'organisation

Il n'y a pas de format ni de support idéal :

- des migrations sont inévitables

- le choix dépend du contexte (en particulier, implication du service informatique, volume de données)

=> L'essentiel est de savoir ce qu'on a et de s'organiser pour le surveiller.



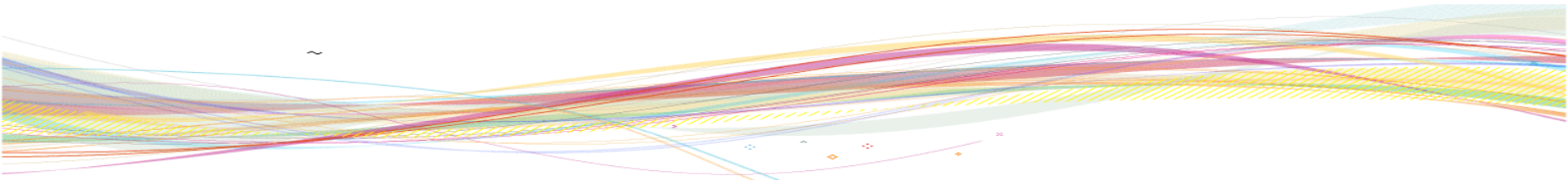
Politiques de conservation numérique

Politique de conservation numérique :

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

Il est indispensable de répartir clairement les responsabilités (bibliothèque ou archives, service informatique).

~



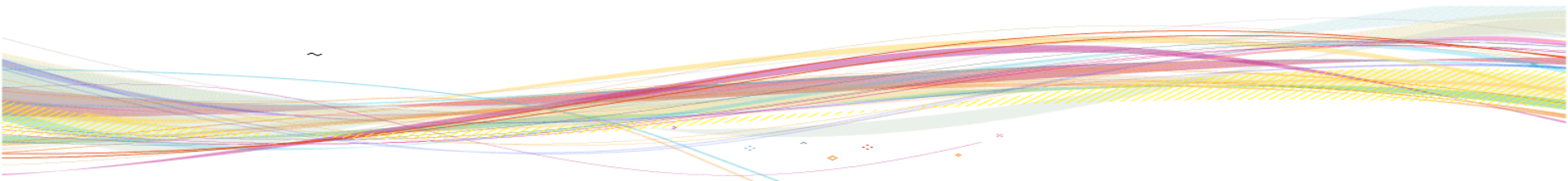
Conclusion

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

Aspect pris en compte dans le plan national de numérisation.

Intérêt de la mutualisation.

- Au niveau d'une collectivité (cf. enquête DLL)
- Entre plusieurs collectivités (ex. AD Yvelines / BM Versailles).



Pour tout renseignement complémentaire :

olivier.de-solan@culture.gouv.fr

