



Création d  
Mise à jour

Date de début des faits  
Substitution du domaine juridique



GALERIES THÉMATIQUES  
Découvrir le fa  
A B C

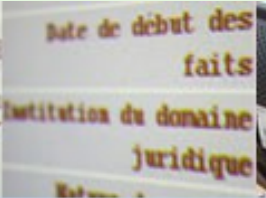
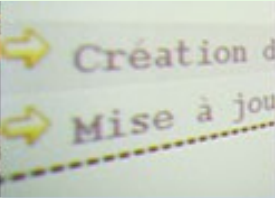
Archives de France

# Archivage numérique et pérennisation : enjeux et risques

Association des Archivistes de l'Église de France

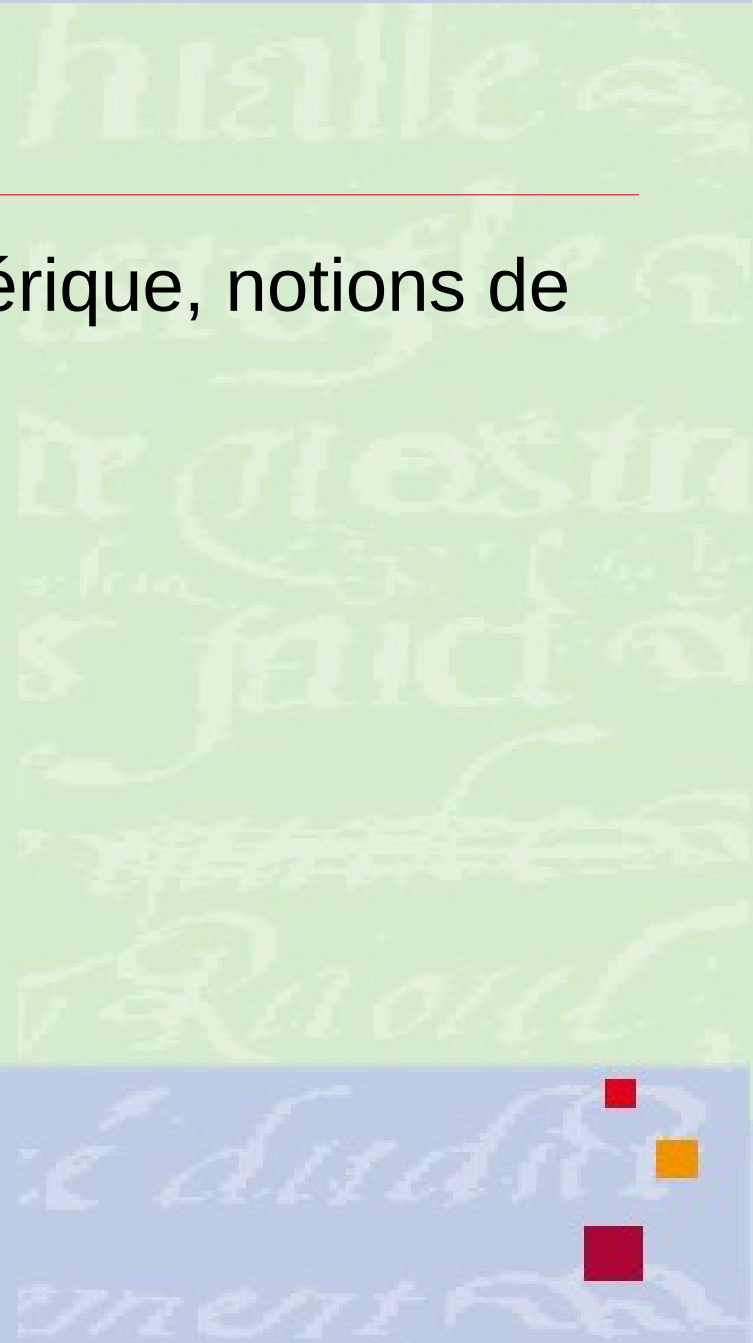
20 mai 2015





# Sommaire

- Caractéristiques de l'information numérique, notions de base
- Les supports
- Les formats
- Les métadonnées
- Les référentiels





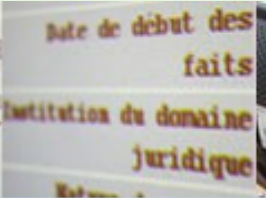
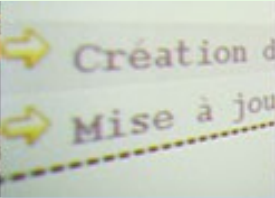
Création d  
Mise à jour

Date de début des faits  
Constitution du domaine juridique



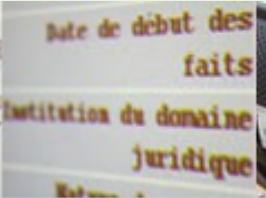
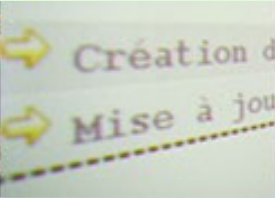
GALERIES THÉMATIQUES  
Découvrir le fa  
A B C

## Caractéristiques, notions de base



## Archives / Information

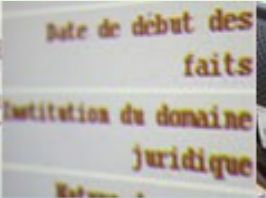
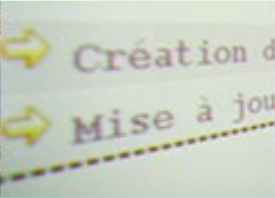
- « *Les **archives** sont l'ensemble des **documents**, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité.* »
  - Code du patrimoine, article L. 211-1.
- *Le modèle de référence pour un OAIS (Système ouvert d'archivage de l'information - ISO 14721) n'utilise que rarement le terme document, au bénéfice du terme **information***



## Le document

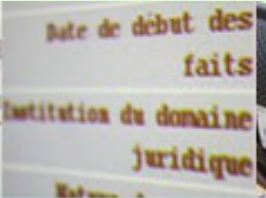
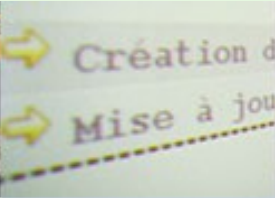
- Les documents sont des informations organisées dans une intention particulière.
  - Par exemple une carte d'identité est un document rassemblant des informations sur une personne (appellation, date de naissance, caractéristiques physiques) et sur une autorité de délivrance du document (l'État).
  - Le document a une fonction : permettre à une personne de prouver son identité.





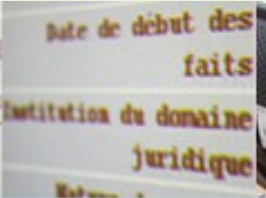
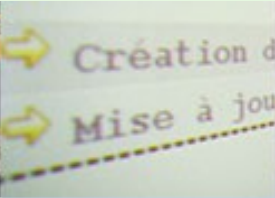
## Information / Données / Métadonnées

- La norme OAIS (ISO 14721) distingue :
  - *l'information* : connaissance que l'on peut échanger ;
  - *la donnée* : représentation formalisée de la connaissance.
- Pour conserver une information, on lui donne une forme (la donnée).
- Cette forme peut être liée à des technologies qui sont éphémères.
- On a donc besoin de conserver aussi des informations sur cette forme (les **métadonnées**)



## Information / Support

- Dans le domaine du numérique, l'association entre le support et l'information qu'il porte perd son sens.
- Le document pourra au cours du temps être stocké sur des supports différents sans pour autant que cela ne change sa nature ni n'altère son intégrité



## Information / Support

- Dans le domaine du numérique, l'association entre le support et l'information qu'il porte perd son sens.
- Le document pourra au cours du temps être stocké sur des supports différents sans pour autant que cela ne change sa nature ni n'altère son intégrité





## Vulnérabilité de l'information numérique

- L'information sous forme numérique est difficile à conserver car :
  - Elle est codée (exemple de problème : les formats propriétaires)
  - Elle est dépendante de la technologie (exemple de risque : l'obsolescence) à tous les étages

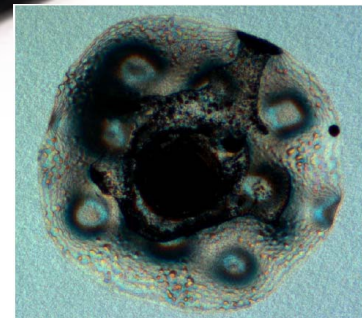
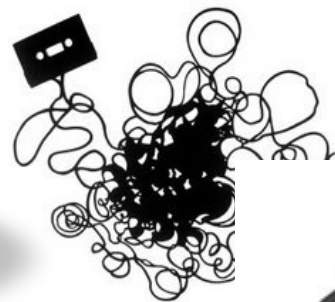




# Vulnérabilité - Couche support

## ■ Agressions

- Magnétiques
- Physiques
- Chimiques
- Biologiques
- Thermiques



9/9

0800 Antan started  
 1000 ■■■ stopped - antan ✓  
 1300 (033) MP - MC 2.130476415  
 (033) PRO 2 2.130476415  
 conv 2.130676415  
 Relays 6-2 in 033 failed special speed test  
 in relay 11.00 test.  
 Relays changed  
 1100 Started Cosine Tape (Sine check)  
 1525 Started Multi Adder Test.  
 1545 Relay #70 Panel F (moth) in relay.  
 First actual case of bug being found.  
 1630 Antan started.  
 1700 closed down.

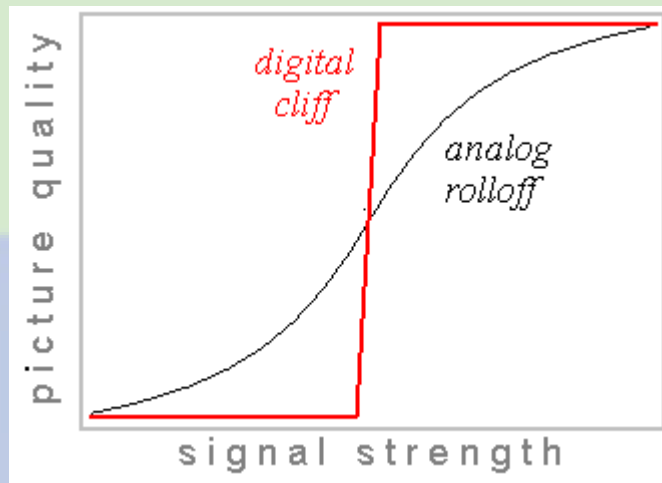
Relay 3145  
 Relay 3370





## Vulnérabilité - Couche support

- La perte d'information peut être brutale et massive
  - L'effet de falaise : au delà d'un seuil, la dégradation du support qui passe complètement inaperçue se traduit par une perte brutale de l'information
  - Zone endommagée (dans le fichier vs. dans la table des matières du support)





Création d  
Mise à jour

Date de début des faits  
Substitution du domaine juridique



GALERIES THÉMATIQUES  
Découvrir le fa  
A B C

## Vieillessement

- Analogie avec un texte sur papier

**Trace physique de l'information** ———→ **Restitution par un lecteur**

**Trace**

« **trace** »

*Trace*

« **trace** »

*Trace*

???





## Vulnérabilité – Obsolescence

- Obsolescences technologiques
  - Matériel de lecture
  - Pilote
  - Câblage
  - Système de fichier
  - Système d'exploitation





## Vulnérabilité – Couche données

### ■ Manque d'information techniques (de représentation)

- Codage/format inconnu

Le « train de bits »

10111011101111111100111010011100  
11001110...

Le « train de bits » découpé en octets exprimés en hexadécimal



```

bb bf ce 9c ce b5 20 cf 84 ce b7 ce bd 20 ce ad ce bd ce bd ce bf ce b9 ce b1
ce b4 ce bf ce bc ce b7 ce bc ce ad ce bd ce b1 20 ce b4 ce b5 ce b4 ce bf ce
ce ad ce bd ce b1 20 ce b5 ce bd ce bd ce bf ce bf cf 8d ce bc ce b5 20 ce bc
af ce b1 20 cf 83 cf 85 ce bb ce bb ce bf ce b3 ce ae 20 cf 83 cf 84 ce bf ce
cf 87 ce b5 ce af cf 89 ce bd 20 ce b4 ce b5 ce b4 ce bf ce bc ce ad ce bd cf
ce bd 20 cf 8c cf 80 cf 89 cf 82 20 ce b5 ce af ce bd ce b1 ce b9 20 ce b3 ce
ce b1 20 cf 80 ce b1 cf 81 ce ac ce b4 ce b5 ce b9 ce b3 ce bc ce b1 20 cf 84
b1 20 ce bb ce bf ce b3 ce b9 cf 83 cf 84 ce b9 ce ba ce ac 20 cf 86 cf 8d ce
ce bb ce b1 2c 20 ce bf ce b9 20 ce ba ce b1 cf 84 ce ac ce bb ce bf ce b3 ce
ce b9 20 ce b4 ce b9 ce b5 cf 85 ce b8 cf 8d ce bd cf 83 ce b5 cf 89 ce bd 2c
ce bf ce b9 20 cf 80 ce b1 cf 81 ce ac ce bc ce b5 cf 84 cf 81 ce bf ce b9 20
b4 ce b9 ce b1 ce bc cf 8c cf 81 cf 86 cf 89 cf 83 ce b7 cf 82 2c 20 ce bf ce
20 ce bf ce b9 ce ba ce bf ce bd ce bf ce bc ce b9 ce ba ce ad cf 82 20 cf 83
85 ce bd ce b1 ce bb ce bb ce b1 ce b3 ce ad cf 82 20 ce ba ce b1 ce b9 20 cf
ce b1 20 cf 84 ce b5 cf 87 ce bd ce b9 ce ba ce ac 20 cf 83 cf 87 ce ad ce b4
b9 ce b1 2e

```



## Vulnérabilité – Couche données

### ■ Et si c'était du texte...

- Avec quelques hypothèses : ASCII...

```
10111011101111111100111010011100
11001110...
```

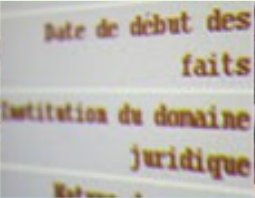
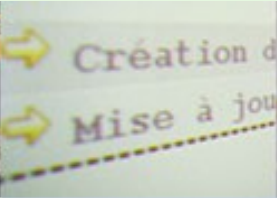


```
ï»¿ïœîµ ï„ï·î½ î-î½î½î½î½î½î½ î´î½î½î·î½î-î½î± î´îµî
´î½î½î-î½î± îµî½î½î½î½î½î½î½î½î½ î½î-î± ïfï...î»¿î»¿î½î³î®
ïfï„î½î³î±îµî-î%î½ î´îµî´î½î½î-î½î%î½ îœïœï
%ï, îµî-î½î±î¹ î³î¹î± ïœî±ï)î-î´îµî¹î³î½î± ï„î±
î»¿î½î³î¹ïfï„î¹î°î- ï+ï½î»¿î»¿î±, î½î¹
î°î±ï„î-î»¿î½î³î½î¹ î´î¹îµï...î,ï½îïfïµï%î½, î½î¹
ïœî±ï)î-î½îµï„ï)î½î¹ î´î¹î±î½îœï)ï+ï%ïfï·ï,, î½î¹
î½î¹î°î½î½î½î½î¹î°î-ï, ïfï...î½î±î»¿î»¿î±î³î-ï, î°î±î¹
ï„î± ï„îµï±î½î¹î°î- ïfï±î-î´î¹î±.
```

- Et si c'était en UTF-8...

Με την έννοια δομημένα δεδομένα εννοούμε μία συλλογή στοιχείων δεδομένων όπως είναι για παράδειγμα τα λογιστικά φύλλα, οι κατάλογοι διευθύνσεων, οι παράμετροι διαμόρφωσης, οι οικονομικές συναλλαγές και τα τεχνικά σχέδια.

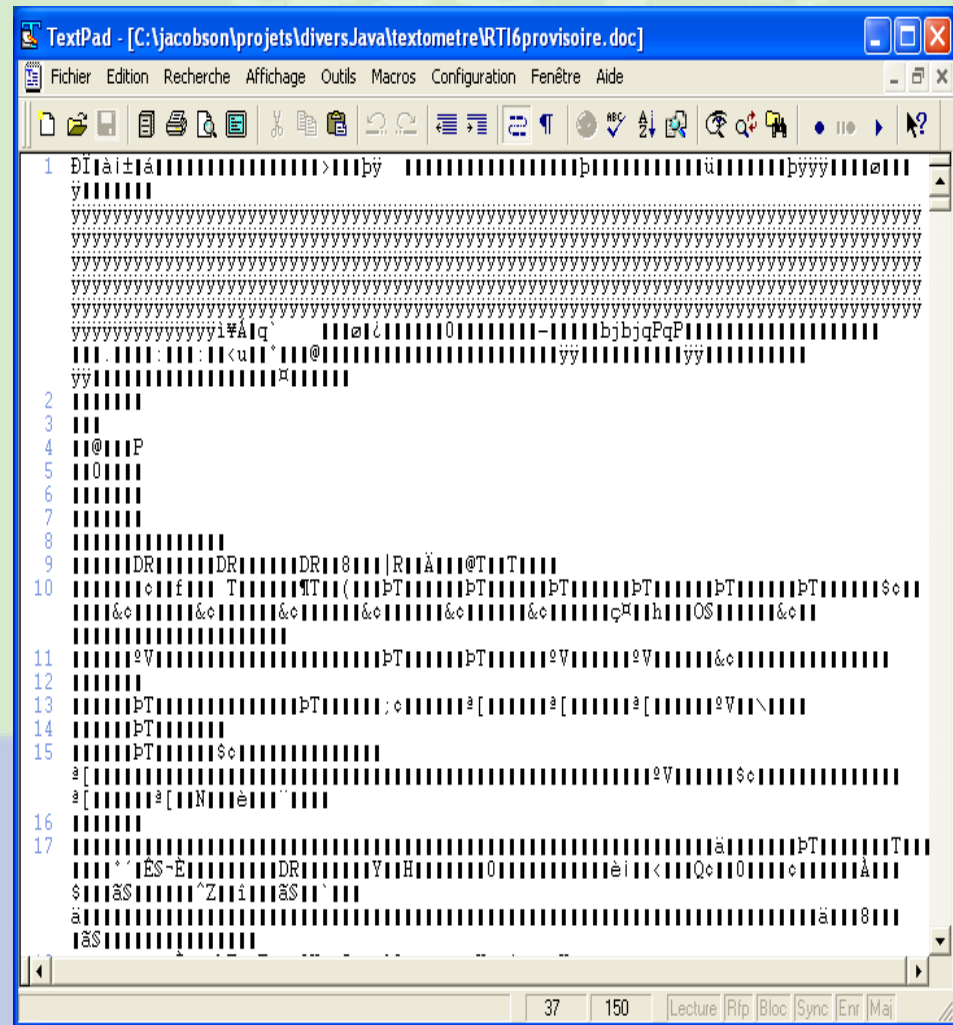




# Vulnérabilité – Couche données

Absence d'information techniques (de représentation)

- Format identifié mais propriétaire







## Vulnérabilité du numérique

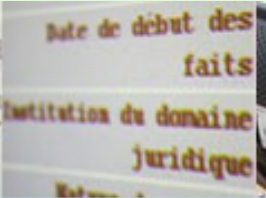
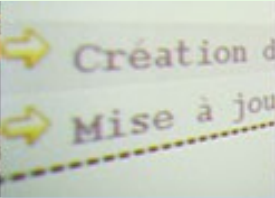
- La perte d'information
  - Erreur de manipulation
    - DEL \*.\*
  - L'information est quelque part... mais où ?



« Où est Charlie ? »







## Importance du contexte de création

- L'information structurée est assez facile à récupérer, car elle est déjà **qualifiée** et **décrite**.
- L'information **non structurée** est beaucoup plus difficile à récupérer si elle n'a pas été encadrée en amont.
- Comment retrouver rapidement les documents importants pour l'archivage ?
- Comment gérer la conservation de ces « vracs » bureautiques ?





## Importance du contexte de création

- Une bonne politique d'archivage électronique commence par **définir des bonnes pratiques sur les conditions de création et d'enregistrement** des informations par les producteurs.
  - *Repérer préalablement les données importantes*
  - *Les classer*
  - *Les documenter*
  - *Gérer leur cycle de vie (durée de conservation nécessaire pour des raisons juridiques)*
  - *Influence du « records management »*
  - *De plus en plus, les producteurs prennent appui sur des outils appelés « Gestion électronique de documents » (GED)*



Création d  
Mise à jour

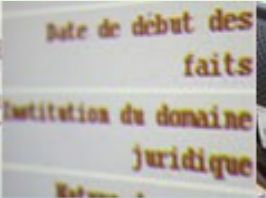
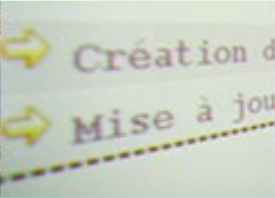
Date de début des faits  
Substitution du domaine juridique



GALERIES THÉMATIQUES  
Découvrir le fa  
A B C

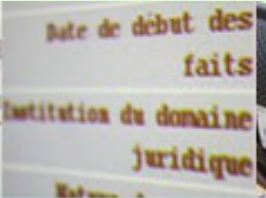
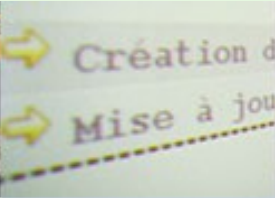
**Quels supports pour la conservation ?**





## Les besoins

- Les opérations que l'on souhaite faire sur les supports
  - Du stockage de données
  - De la conservation des données
  - Des accès aux données



## Les besoins

- Le stockage des données
  - Disposer de supports capacitifs :
- La conservation des données
  - Disposer d'un support fiable
  - Disposer d'un support durable
- L'accès aux données
  - Accès en écriture
  - Accès en lecture
  - Faire en sorte qu'une charge importante d'un type d'accès ne pénalise pas fortement les autres accès



## Quels supports pour la conservation ?

### ■ Que choisir comme support ?

- Des supports amovibles optiques : CD ou DVD-R
- Principales limites : la volumétrie
- Procédures très manuelles à moins d'opter pour des systèmes automatisés de récupération des supports
- Pas d'outils sur le marché qui permettent par exemple de procéder à des migrations massives de données inscrites sur un DVD-R vers d'autres DVD-R, voire d'autres types de supports







## Quels supports pour la conservation ?

- Le stockage sur disques durs / serveurs
- Moins onéreux, plus simples, se prêtent à des volumétries importantes
- Mais sans maîtrise technique
- Stratégies de contournement

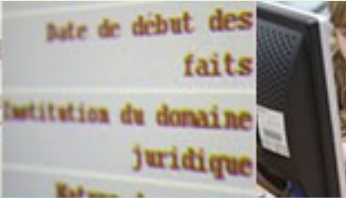
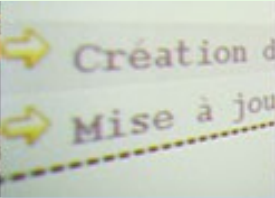




## Quels supports pour la conservation ?

- Le stockage sur bandes magnétiques
  - Volumétries de données importantes
  - Bonne résistance physique au temps
  - Limites

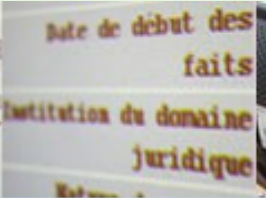
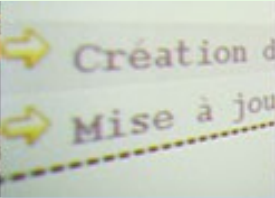




## Quels supports pour la conservation ?

- Migration des supports
- Systèmes de sauvegarde / de redondance / de réplication sur des sites distants
- Contrôle et surveillance des supports : mise en place de cycles de renouvellement des matériels





## Conclusion

- Il n'y a pas de « bon » support
  - Il existe des supports plus ou moins bien adaptés pour l'usage que l'on veut en faire (conservation pérenne, accès rapide...)
  - Il existe des supports plus ou moins abordables pour ses moyens



Création d  
Mise à jour

Date de début des faits  
Constitution du domaine juridique



GALERIES THÉMATIQUES  
Découvrir le fa  
A B C

## Les formats





## De quoi sont fait les fichiers ?

- Un fichier est une suite consécutive de bit (train de bits)
  - Dont le système de fichier connaît des propriétés (nom, localisation, dates, droits)
  - Qui possède une syntaxe (organisation des bits)

```
01010010 01000101 01000110 01000110
00110000 00000000 00000001 00000000
01010111 01000001 01010110 01000101
00010000 00000000 00000000 00000000
00000001 00000000 00000001 00000000
00010001 00101011 00000000 00000000
00010001 00101011 00000000 00000000
10000000 10000000
```

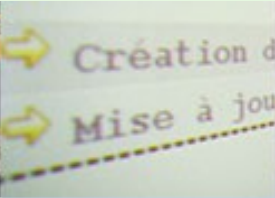




## Conséquences sur la conservation

- Le train de bits n'est pas auto-suffisant
  - Il convient de conserver en même temps que le train de bits, les informations de décodage (= les « spécifications » du format).
  - Pour les formats conteneurs, il faut également connaître les formats et codages utilisés dans les contenus.
  - La description de certains formats ne sont que des sur-couches d'autres formats, il faut donc également connaître les spécifications des formats parents.
  - Si les spécifications sont déjà conservées ailleurs on peut se contenter de conserver des informations d'identification de ce format
  - Importance du choix du format des métadonnées





## Formatage

- Formats codant plusieurs types d'information
  - Par exemple un fichier odt (OpenDocument Text) code du texte mais aussi des images, des informations de mise en page, etc.
- Formats pouvant coder un même type d'information de plusieurs manières
  - WAV autorise une bonne centaine de codages audio dont PCM, MP3, mu-law, a-law, etc.
- Formats codant des métadonnées techniques
  - Les images JPEG peuvent contenir par exemple des métadonnées (exif)
- Formats répartissant l'information dans plusieurs fichiers
  - .xml (.xsd, include, import), plans d'archis ARC+ (.iii, .ddd, .icn)

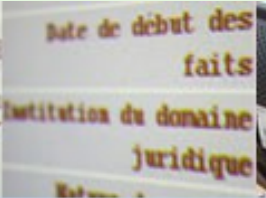
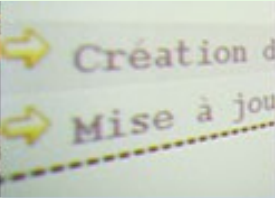




## Les critères d'évaluation

- Spécifié : il existe une description de la convention utilisée pour représenter la donnée et celle-ci est suffisamment décrite pour en développer une implémentation complète.
- Ouvert : la convention est publique (sinon le format est dit fermé). Elle est donc sans restriction d'accès ni de mise en œuvre.
- Normalisé : la convention est adoptée par des organismes de normalisation (ISO, W3C). Exemple : le PDF/A.
- Standardisé : il n'existe pas de norme sur ce format mais son utilisation est tellement répandue qu'il est considéré comme un standard. Exemple : le PDF. ATTENTION : en anglais "standard" signifie "norme".
- Propriétaire : si l'exploitation du format entre dans le cadre du droit privé, il dépend alors de l'existence du propriétaire (Par exemple les formats doc de Microsoft). Cet aspect propriétaire n'empêche par forcément les spécification d'être publiées. Exemple : les spécification du format PDF ont toujours été publiées par Adobe.

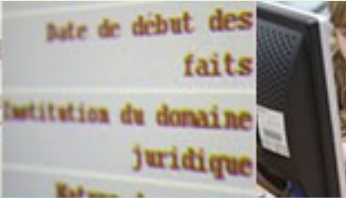
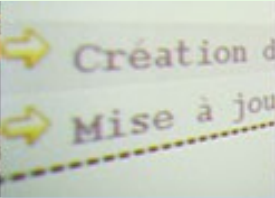
(Source : CNES, *Référentiel normatif, Ingénierie des données. Evaluation des formats de données par rapport à la pérennisation*, 2005)



## Les critères d'évaluation

Ces cinq critères permettent de définir le niveau de pérennité d'un format :

- 1. un format de données doit être entièrement et explicitement spécifié et sa spécification doit être connue du service en charge de l'archivage long terme ;
- 2. un format de données doit être apte à représenter la sémantique et la complexité de l'information à préserver ;
- 3. l'usage de formats normalisés est recommandé. En outre, on évitera l'usage d'éléments propriétaires au sein d'un format normalisé ;
- 4. lorsque le besoin de pouvoir modifier un document est identifié, le choix du format doit prendre en compte cette contrainte ;
- 5. le choix des formats doit pouvoir prendre en compte la disponibilité et le coût des outils d'exploitation ;



## Les critères d'évaluation

(suite)

- 6. il doit être possible de vérifier automatiquement qu'un fichier de données respecte les spécifications du format et respecte également les règles d'utilisation et des restrictions qui auront été éventuellement définis pour la pérennité ;
- 7. la possibilité d'extraire automatiquement tout ou partie des métadonnées à partir des données constitue un avantage certain ;
- 8. l'usage de formats inutilement volumineux sera évité ;
- 9. à fonctionnalités égales un format simple est préférable à un format complexe ;
- 10. les formats largement reconnus et utilisés seront privilégiés ;
- 11. le choix d'un format doit prendre en compte la disponibilité et le coût des outils de transformation des formats et de représentation des données.





## Familles de formats

### ■ Formats bureautiques et non structurés

#### – PDF - PDF/A

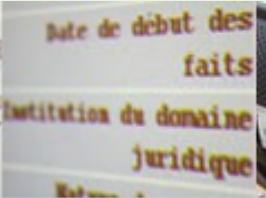
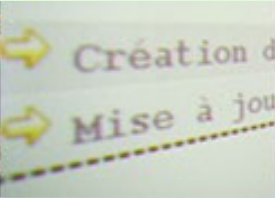
- PDF - format propriétaire (ADOBE)
- PDF/A-1 est devenu la norme ISO 19005-1 en 2005
- PDF 1.7 est devenu la norme ISO 32000-1 en 2008
- PDF/A-2 nouvelle version de PDF-A publiée. (juin 2011): elle s'appuie sur la version 1.7 de PDF lui même normalisé en ISO 32000-1

#### – ODF, Open Document Format, pour les documents bureautiques

- La version 1.0 est standardisé par OASIS en 2005 puis normalisée par l'ISO en 2006 (ISO 26300)
- La version 1.1, est standardisée par Oasis en 2007.puis l'ISO en 2012
- La version 1.2 est actuellement un standard Oasis depuis 2011

#### – OOXML, Office Open XML, pour les documents bureautiques

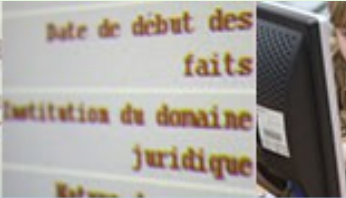
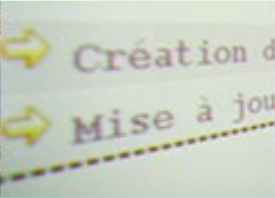
- Standardisé par ECMA en 2006 ; Normalisé par ISO en 2008 (ISO 29500)



## Familles de formats

### ■ Formats image

- PNG, Portable network Graphics,
  - pour les images matricielle (à l'origine crée pour offrir une alternative libre au format GIF qui utilisée une technique de compression sans perte LZW soumise à un brevet). Standardisé par le W3C en 1996 ; Normalisé par l'ISO en 2004 (ISO 15948:2004)
- GIF, Graphics Interchange Format
  - un format ouvert de la société CompuServe pour la représentation d'images matricielle. Il utilise une compression sans perte LZW dont le brevet a maintenant expiré. Version GIF87a ; Version GIF89a permet l'inclusion de plusieurs images dans un même fichier
- JFIF, JPEG File Interchange Format
  - un format pour la représentation d'images matricielles compressées avec l'algorithme JPEG (Joint Photographic Experts Group) ; L'algorithme JPEG est définit par la norme ISO 10918-1 en 1993



## Familles de formats

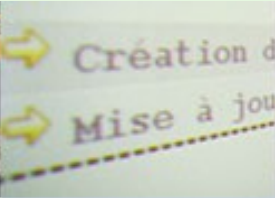
### ■ Formats image (suite)

#### – JPEG2000

- Norme ISO 15444-1 de 2000 utilisant pour la compression (avec ou sans perte) des images un algorithme une transformée en ondelettes permettant des meilleurs taux que l'agorithme de la norme ISO 10918-1.

#### – TIFF, Tagged Image File Format

- format conteneur propriétaire de Adobe pour des images numériques.



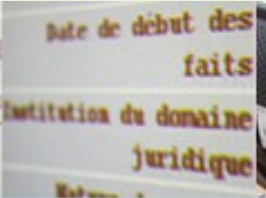
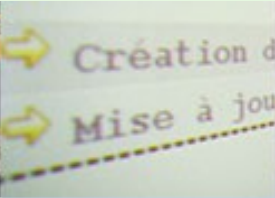
## Familles de formats

### ■ Formats pour les bases de données

- Export de tables en CSV, en XML
- SIARD,
  - Note d'information DGP/SIAF/2010/017 du 21 septembre 2010. Étude du format SIARD pour l'archivage des bases de données relationnelles et au logiciel SIARDSuite mettant en œuvre ce format.

### ■ Formats audio-visuels

- Guide méthodologique pour le choix de formats numériques pérennes dans un contexte de données orales et visuelles : nouvelle version mise à jour en 2011 ; Note d'information DGP/SIAF/2010/010 du 21 mai 2010



## Exemples

- Validation : <http://facile.cines.fr/>
- Identification et validation : <http://projects.iq.harvard.edu/fits/home>
- Base de formats : <http://apps.nationalarchives.gov.uk/PRONOM/BasicSearch/proBasicSearch.aspx?status=new>







Création d  
Mise à jour

Date de début des faits  
Substitution du domaine juridique



GALERIES THÉMATIQUES  
Découvrir le fa  
A B C

## Les métadonnées





Création d  
 Mise à jour

Date de début des faits  
 Constitution du domaine juridique



## Métadonnées sémantiques

A8600200046A36	00900879	2001985363
7183		
P8600200046001000P		
PRE M HABARI		
APARI		BEN
	PEN	1940

*Extraction d'une base de données, lisible mais sans signification*



## Métadonnées sémantiques

**A8600200046A36**

AFFAIRE - Numéro d'affaire ou numéro de parquet : ce numéro est constitué des zones suivantes :

- année sur 2 chiffres
- quantième du jour dans l'année sur 3 chiffres de 001 à 365 (ou 366)
- numéro d'ordre sur 4 chiffres au choix de l'utilisateur
- une clé sur 1 chiffre

**Nomenclatures employées pour coder les informations**

- exemple : nomenclature NATAFF (nature de l'affaire) : B24 Vol par ruse ou usurpation d'identité

### Information sémantique

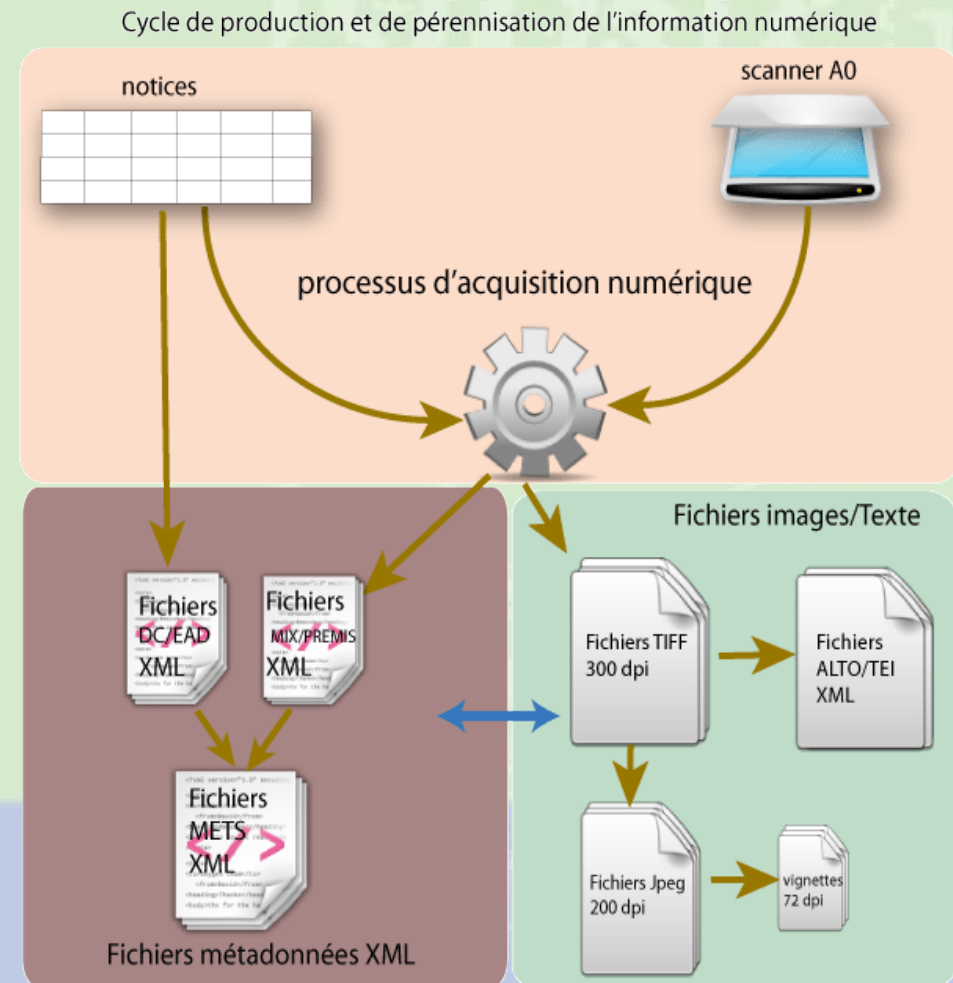
Complément d'information permettant d'interpréter la donnée d'un point de vue utilisateur.

Ex: Identification des champs  
Les nomenclatures



## La production de métadonnées

- La production de métadonnées peut s'envisager à différentes étapes du cycle de vie du document porteur d'information :
  - saisie à la création
  - saisies au cours d'une étape de traitement de l'information
  - extraction à posteriori : transformation, diffusion, récupération





## Des formats de représentation de métadonnées



*Des métadonnées descriptives et intellectuelles*



*Des formats d'empaquetage d'ensemble de métadonnées*



*Des métadonnées pour la préservation des objets archivés*



## Encoder et lier les métadonnées

### ■ XML (Extensible Markup Language)

- langage à balises permettant de structurer et de décrire un ensemble d'objets au sein d'un espace de noms
- langage hiérarchique permettant de décrire des relations
- langage de transport d'information permettant de standardiser les échanges de données
- peut être décrit par un schéma ou une DTD
- peut servir au stockage des données mais également à leur interrogation

```
<xml version="1.0" encoding="UTF-8"?>
<ex:collection xml:lang="fr" xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns=http://www.w3.org/1999/xhtml xmlns:ex="http://exemple.org">
  <ex:vehicule>
    <dc:title>Mon super vélo</dc:title>
    <velo id="2">
      <taille>76</taille>
      <marque>cool bike</marque>
      <modele>speeder</modele>
    </ex:velo>
  <ex:velo taille="76" marque="coolbike" modele="speeder" id="2" />
</ex:collection>
```



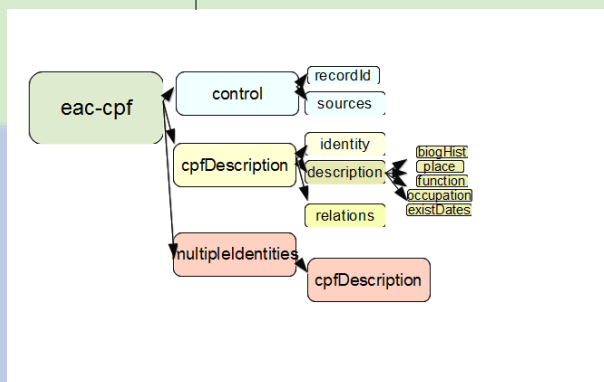
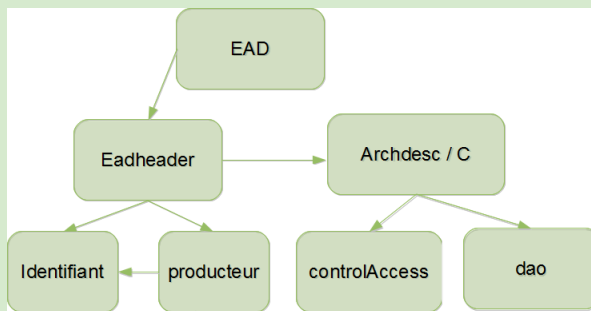
## Encoder et lier les métadonnées

- EAD (Description archivistique encodée) et EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles) pour la description archivistique
  - La DTD EAD est basée sur ISAD(G). L'EAD permet de décrire un fonds d'archives ou d'ouvrages de manière hiérarchique du général au particulier, d'indexer les notices descriptives et de lier ces descriptions à des substituts numériques et à des notices de producteurs EAC
  - Le schéma EAC vise à décrire les producteurs pour retracer l'historique d'une entité administrative ou morale d'archives en se basant sur la norme ISAAR(CPF).
  - Il permet une redéfinition de la notion de contexte de production et d'utilisation des archives. Il se concentre seulement sur la description des collectivités, des personnes et des familles.

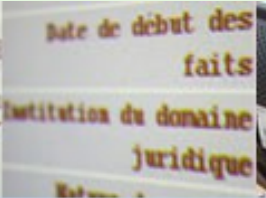
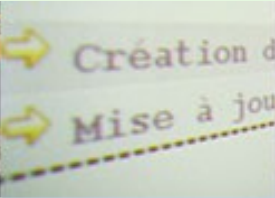


## Encoder et lier les métadonnées

- EAD (Description archivistique encodée) et EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles) pour la description archivistique







## PREMIS : la structuration des objets d'information

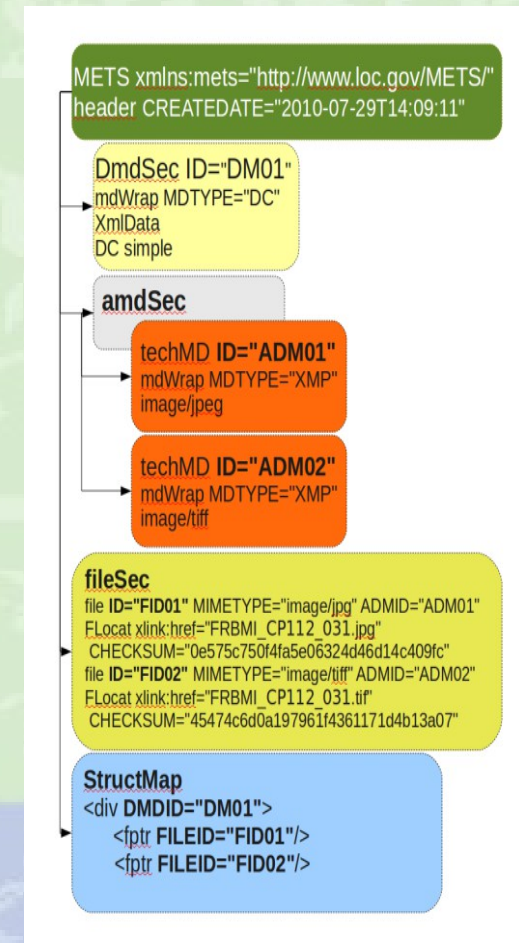
### ■ PREMIS (Preservation Metadata: Implementation Strategies)

- Le modèle de données de PREMIS est composé de 5 entités reliées entre elles par des identifiants: Entité intellectuelle, Objet, Événement, Agent, et Droits.
- Ce schéma répond notamment au besoin de pérennisation de l'information numérique et de conservation du cycle de vie des données (import dans un entrepôt de donnée, migration, gestion d'intégrité, liens entre différentes représentations d'un même document).  
→ [<http://www.loc.gov/standards/premis>]



## METS : la structuration de la description des ressources

- Le schéma METS vise à fournir un méta-modèle pour l'encapsulation d'autres formats de métadonnées.
- Ce schéma XML composé de 7 sections est extrêmement flexible.
- Il permet notamment de représenter l'information sous forme de carte de structure et s'implémenter le modèle conceptuel OAIS.
- Il constitue un format pivot entre la logique de diffusion et celle de conservation de l'information.





## Encoder et lier les métadonnées

### ■ URI (Uniform Resource Identifier)

- Les URI sont des identifiants Web.
- Les URL sont des URI qui identifient la ressource par le moyen d'y accéder.
- Les URL peuvent être pérennes.

Syntaxe :

<scheme:chaîne/de.caractère>

<http://monsite.com/dossier/fichier.html#ancree>

<http://mondomaine.org/ressource/1234>

- un préfixe qui indique le contexte dans lequel l'identifiant est attribué (par ex. http:, ftp:, urn:, etc.)
- un élément qui permet de désigner l'autorité nommante qui a attribué l'identifiant au sein de ce système
- le « nom » lui-même, c'est-à-dire une chaîne de caractères qui identifie la ressource de manière unique, au sein de ce système et pour cette autorité.



Création d  
Mise à jour

Date de début des faits  
Substitution du domaine juridique



GALERIES THÉMATIQUES  
Découvrir le fa  
A B C

## Les référentiels

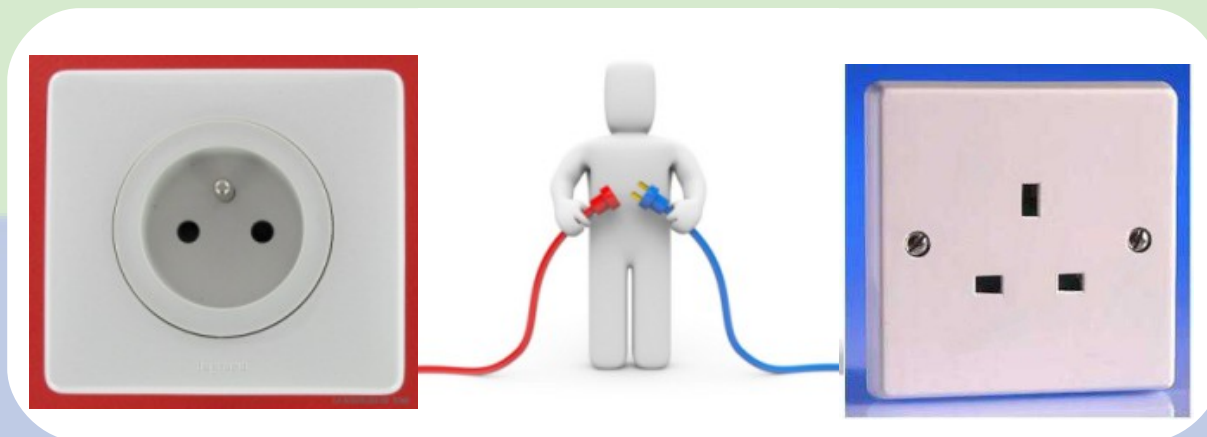
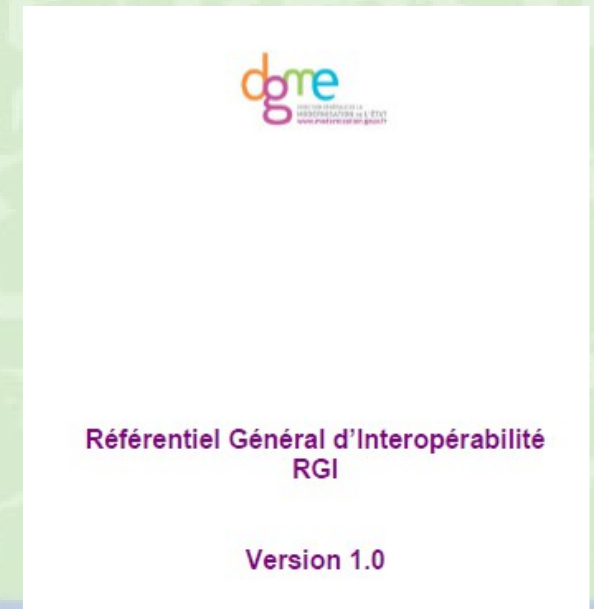




## Référentiel général d'interopérabilité

- Normes et standards que les autorités administratives doivent utiliser afin de garantir l'interopérabilité des informations échangées et de leurs systèmes d'information.
- Publié en novembre 2009.

[references.modernisation.gouv.fr/sites/default/files/RGI\\_Version1%200.pdf](http://references.modernisation.gouv.fr/sites/default/files/RGI_Version1%200.pdf)





# Importance des formats ouverts

## Recommandé Formats d'image

Pour les images numériques, il est RECOMMANDÉ de choisir parmi les formats GIF, PNG, JPEG, TIFF ou DNG.

## Recommandé XML / CSV W3C



Il est RECOMMANDÉ d'utiliser XML ou CSV pour l'exportation de tout ou partie d'une base de données.

## En observation ODF OASIS



Des améliorations de la norme concernant l'accessibilité ont été apportées avec la version 1.1, approuvée par l'OASIS en octobre 2006. La majorité des implémentations du marché reprennent cette dernière version.

La version ODF 1.2 est en cours d'élaboration et devrait être soumise à l'approbation OASIS et ISO en 2009.

## En observation Office Open XML ISO



Le format Office Open XML est un format bureautique basé sur XML. Il supporte nativement une partie des formats binaires bureautiques existants. Il n'existe pas à ce jour d'implémentation de cette norme.

## Recommandé PDF/A ISO



Il est RECOMMANDÉ d'utiliser le format PDF/A, pour l'archivage des documents bureautiques statiques non révisables.

- *Ex. : les formats de données et les recommandations du Référentiel Général d'Interopérabilité à destination des systèmes d'information des administrations*



# Agence nationale de la sécurité des systèmes d'information

[Que faire en cas d'incident ?](#)

[Le site du CERTA](#)

[Portail de la sécurité informatique/Documentation](#)

Vous êtes ici : [Accueil](#) > [Réglementation SSI](#) > [Référentiel général de sécurité](#)

## Référentiel général de sécurité



### Présentation

Le Référentiel général de sécurité (RGS) est créé par l'article 9 de l'ordonnance [n° 2005-1516 du 8 décembre 2005](#) relative aux échanges électroniques entre les usagers et les autorités administratives et entre les autorités administratives. Ses conditions d'élaboration, d'approbation, de modification et de publication sont fixées par le décret [n° 2010-112 du 2 février 2010](#) pris pour l'application des articles 9, 10 et 12 de l'ordonnance citée relatif à la sécurité des informations échangées par voie électronique.

La présente version du RGS (1.0), rendue officielle par arrêté du Premier ministre en date du 6 mai 2010 et publiée ci-dessous, est le résultat d'un travail conjoint entre la Direction générale de la modernisation de l'État (DGME) et l'Agence nationale de la sécurité des systèmes d'information (ANSSI), qui a pris en compte les nombreuses contributions et remarques reçues à la suite d'appels à commentaires adressés aux ministères et au public en 2008, et de présentations faites aux diverses autorités administratives, parmi lesquelles notamment des services de l'État et des représentants des collectivités territoriales.

### Grandes dates

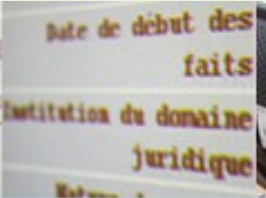
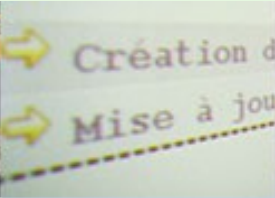
- ✦ notification à la Commission Européenne, en application de la directive 98/34/CE, le 25 mars 2009 ;
- ✦ [avis favorable](#) de la [Commission consultative d'évaluation des normes](#) (CCEN) le 7 mai 2009 ;
- ✦ publication du décret [n° 2010-112 du 2 février 2010](#), dit décret RGS, le 4 février 2010 au *Journal Officiel* ;
- ✦ publication le 18 mai 2010 au *Journal Officiel* de l'[arrêté du 6 mai 2010](#), dit arrêté RGS, portant approbation de la première version du RGS.

### Errata

Dans l'attente de la seconde version du RGS, l'ANSSI souhaite d'ores et déjà assouplir la règle relative au [double usage des clés privées au sein des certificats d'authentification serveur](#), publiée le 6 mai 2010.

- ✦ [Liste des documents constitutifs du RGS v1.0](#)
- ✦ [Documents référencés par le RGS v1.0](#)
- ✦ [Processus de mise à jour du RGS](#)
- ✦ [Appel à commentaires : RGS 2.0](#)

Près de deux ans après la publication de la première version du Référentiel général de sécurité, l'ANSSI souhaite faire évoluer certains documents constitutifs du RGS et y ajouter d'autres annexes.



## La politique d'archivage (P2A)

- Étude en partenariat avec la DCSSI (cabinet Caprioli et Associés)
  - Publication du référentiel : juillet 2006 (<http://www.ssi.gouv.fr/IMG/pdf/ArchivageSecurise-P2A-2006-07-24.pdf>)
- Les livrables :
  - une politique d'archivage type pour le secteur public -à décliner ensuite sur place pour une déclaration des politiques d'archivage (le comment)
  - une prise en compte non seulement de la sécurité (réducteur) mais également du métier (fonctionnalités)





## Archivage électronique sécurisé



9 juillet 2009

# Documentation ANSSI pour l'archivage électronique sécurisé




### LES ANALYSES PREALABLES

 <b>La synthèse des enjeux juridiques</b> PDF - 763.7 ko V.2006-05-16	 <b>L'état de l'art juridique, technique, organisationnel et des offres</b> PDF - 1 Mo V.2006-11-29
---	---

### LES DOCUMENTS INTRODUCTIFS

 <b>La plaquette</b> PDF - 74.6 ko V.2006-08-18	 <b>Le mémento</b> PDF - 278.4 ko V.2006-05-16
---	--

### LES DOCUMENTS D'AIDES A L'ELABORATION DU REFERENTIEL

 <b>Le cahier des charges pour un système d'archivage électronique</b> PDF - 609.4 ko V.2006-05-16	 <b>La politique et les pratiques d'archivage</b> PDF - 545.7 ko V.2006-07-24	 <b>La grille d'audit Excel</b> - 35.5 ko V.2006-07-24
--	---	---

<http://>

[www.ssi.gouv.fr/fr/bonnes-pratiques/outils-methodologiques/archivage-electronique-securise.html](http://www.ssi.gouv.fr/fr/bonnes-pratiques/outils-methodologiques/archivage-electronique-securise.html)



## DISIC

- Création de la **direction interministérielle des systèmes d'information et de communication de l'Etat (DISIC)** par décret en date du **21 février 2011** (décret n° 2011-193) – A rejoint en octobre 2012 le Secrétariat général pour la modernisation de l'action publique (SGMAP)
  - *Chargée notamment d'organiser toutes les opérations de mutualisation des systèmes d'information entre les différentes administrations de l'État. Sorte de DSI interministérielle.*
  - *Objectifs : anticiper les évolutions technologiques, les besoins, fixer les orientations en matière d'externalisation, de sous-traitance, etc. Pour une simplification des relations de l'utilisateur avec l'administration.*
  - *Missions : cartographie des SI de l'Etat, consultation sur les projets d'un certain coût.*
  - **Elle organise et anime le cadre de concertation nécessaire à l'évolution des référentiels généraux d'interopérabilité et d'accessibilité, des modèles de données de référence et des modèles d'échange et, en liaison avec l'Agence nationale de la sécurité des systèmes d'information, du référentiel général de sécurité ;**
  - **A un mandat archivage électronique** dont le but est à la fois de sensibiliser les DSI des ministères à la question de l'archivage électronique et de diffuser un guide de bonnes pratiques sur ce sujet.



## Livrables DISIC sur le mandat archivage électronique

- Accessibles sur le site [references.gouv.fr](http://references.gouv.fr)
  - <http://references.modernisation.gouv.fr/rgi-interoperabilite>

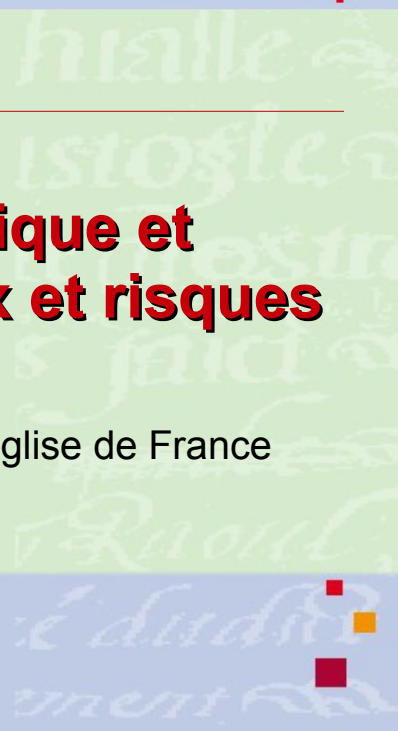
The screenshot shows the website interface for 'RGI - Interopérabilité'. At the top left, there are logos for the French Republic and the Ministry of Modernization. A blue banner reads 'AVEC VOUS l'administration SE MODERNISE'. A search bar with a 'RECHERCHER' button is on the right. Below is a navigation menu with categories: 'URBANISATION SI ETAT', 'RGI - INTEROPÉRABILITÉ', 'RGS ET RÉFÉRENCEMENT - SÉCURITÉ', 'RGAA - ACCESSIBILITÉ', and 'CHARTRE INTERNET DE L'ETAT'. Under 'RGI', there are sub-links: 'Volet technique', 'Volet sémantique', 'Volet organisationnel', and 'Archivage numérique'. The main content area shows 'Accueil » RGI - Interopérabilité' and a large heading 'RGI - interopérabilité'. On the right, there is a button 'TÉLÉCHARGER LA VERSION PDF' and a download icon for 'RGI\_Version1 0.pdf'.



# **Archivage numérique et pérennisation : enjeux et risques**

Association des Archivistes de l'Église de France

20 mai 2015





## Sommaire

---

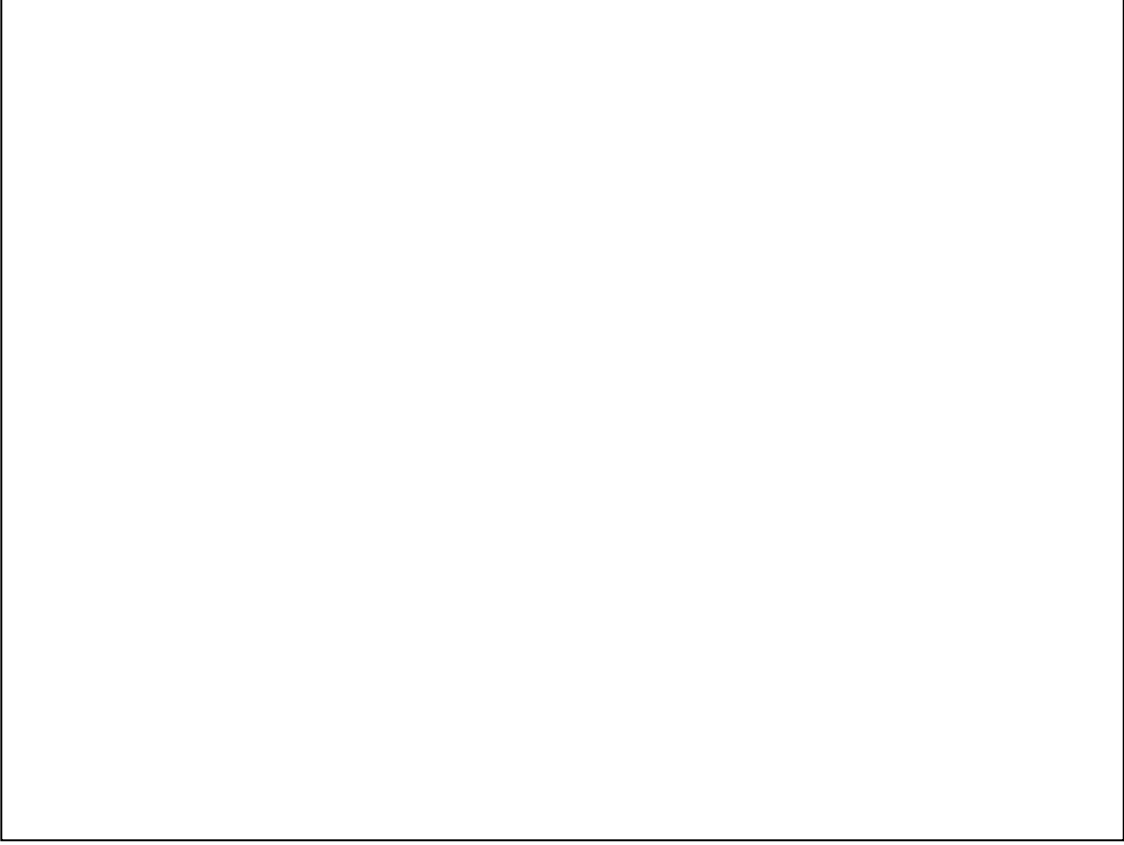
- Caractéristiques de l'information numérique, notions de base
- Les supports
- Les formats
- Les métadonnées
- Les référentiels



---

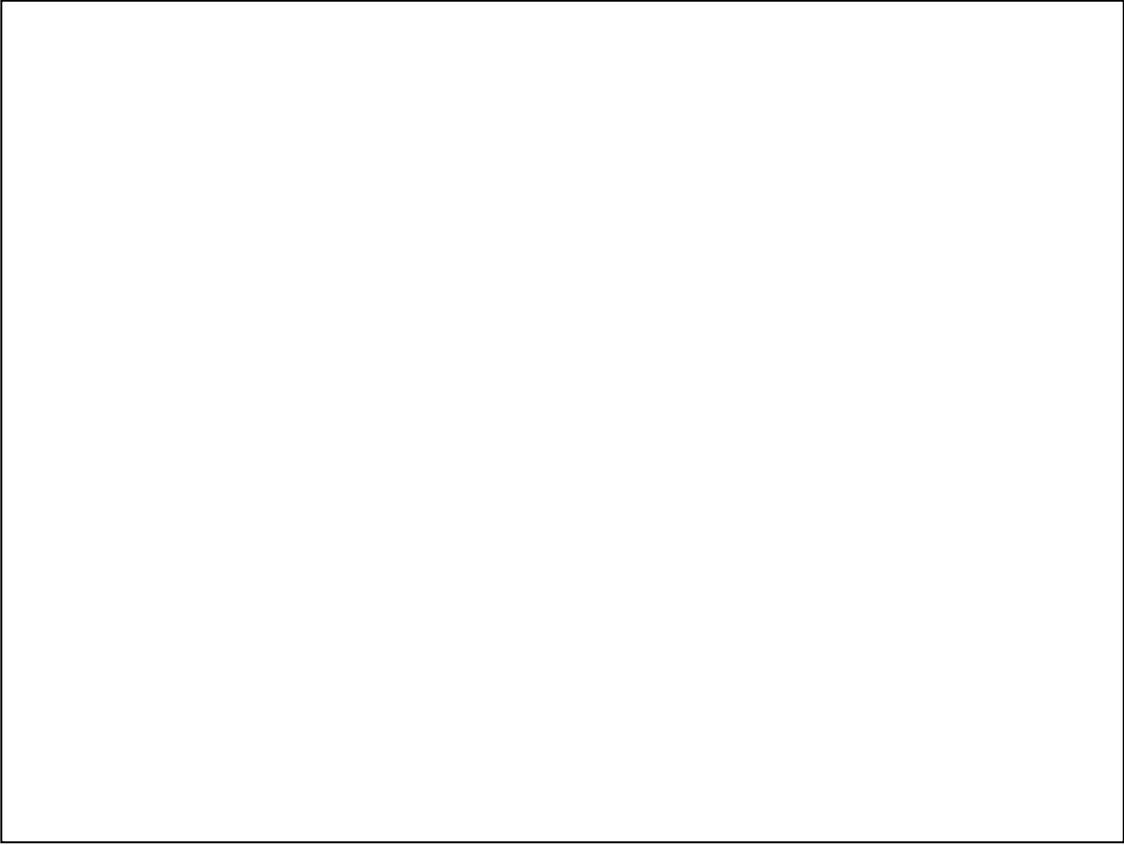
## Caractéristiques, notions de base

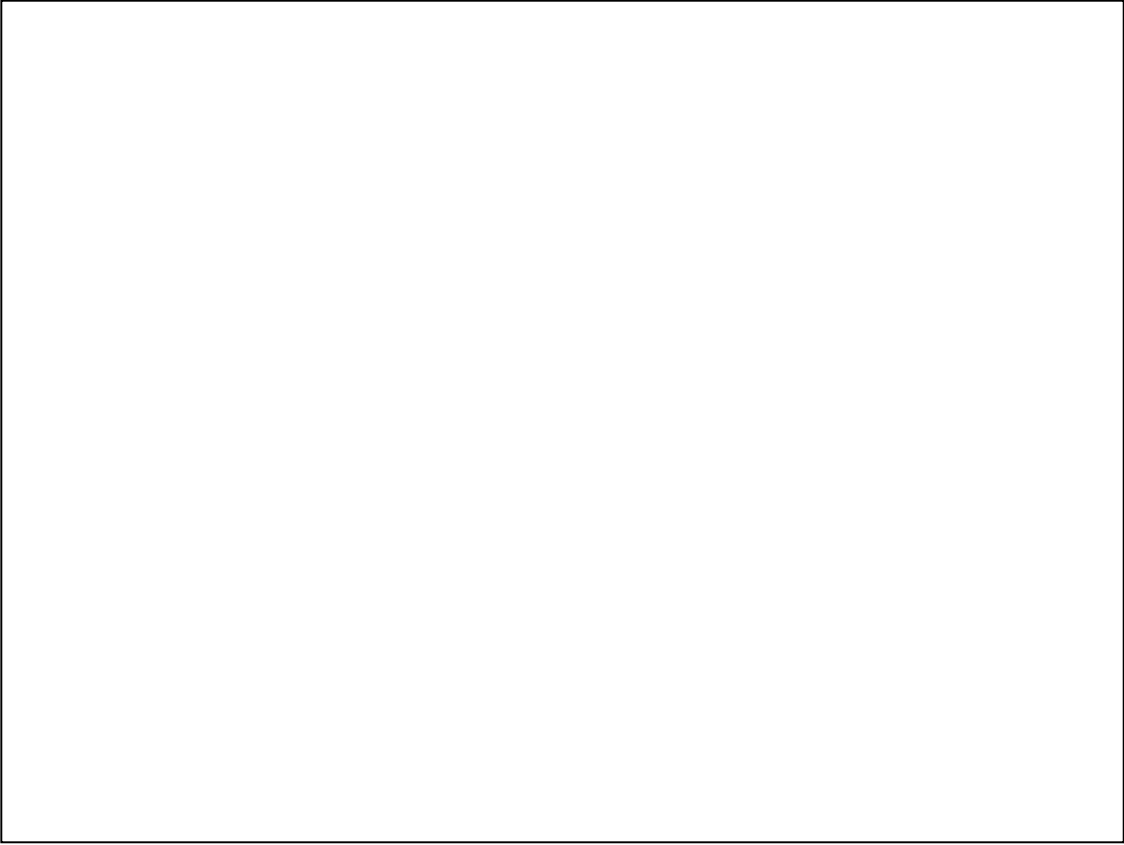






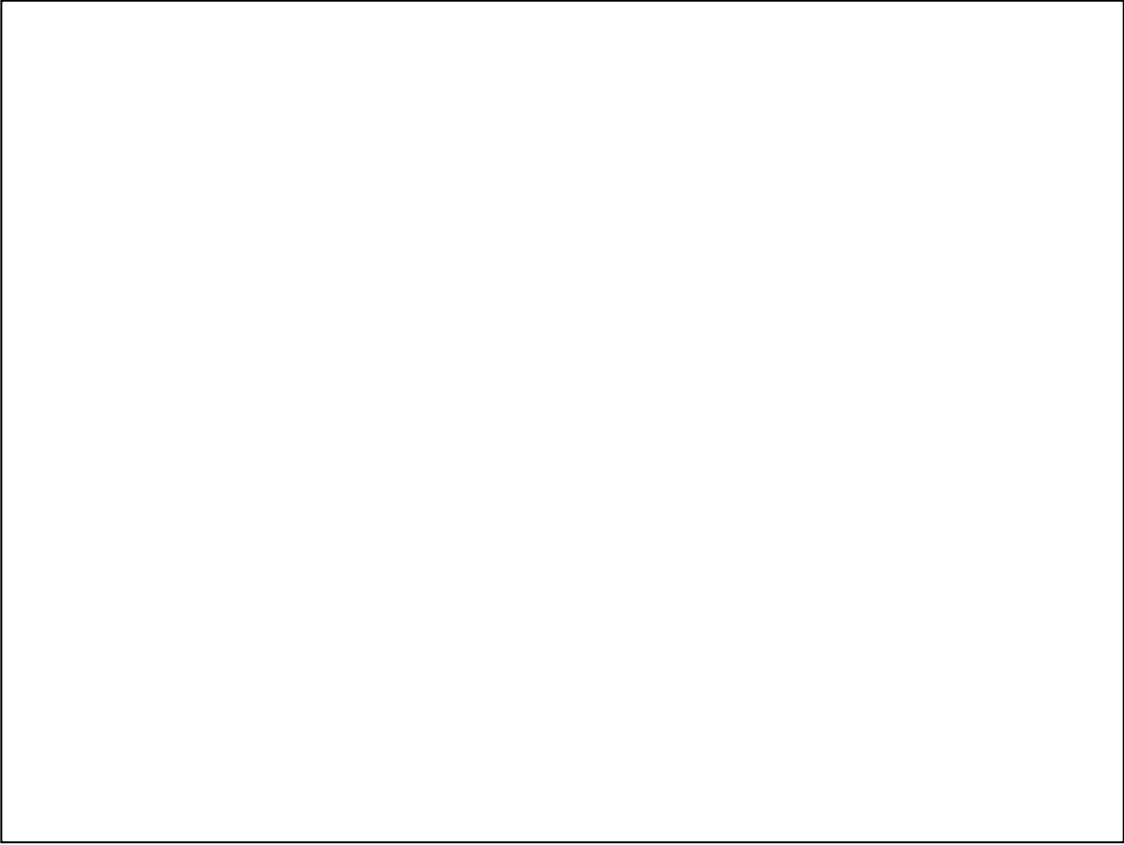














## Vieillessement

### ■ Analogie avec un texte sur papier

**Trace physique de l'information** —→ **Restitution par un lecteur**

**Trace**

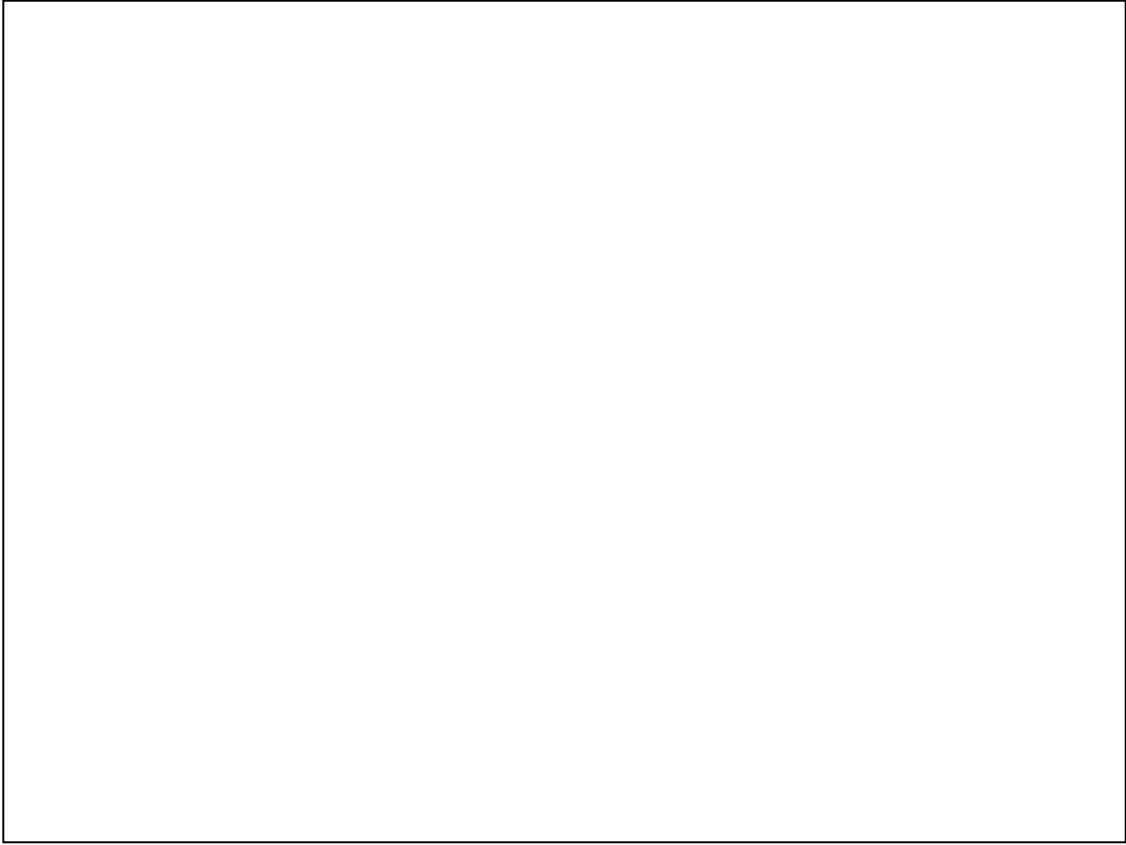
« trace »

*Trace*

« trace »

*Trace*

???







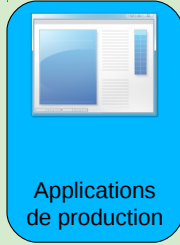






## Deux grands types de données

Contenus mixtes et non structurés



Bureautique      Images      Son      Vidéos...



Contenus structurés

Bases de données      Applications métier      Dispositifs de l'administration électronique





## Importance du contexte de création

- L'information structurée est assez facile à récupérer, car elle est déjà **qualifiée** et **décrite**.
- L'information **non structurée** est beaucoup plus difficile à récupérer si elle n'a pas été encadrée en amont.
- Comment retrouver rapidement les documents importants pour l'archivage ?
- Comment gérer la conservation de ces « vracs » bureautiques ?



## Importance du contexte de création

- Une bonne politique d'archivage électronique commence par **définir des bonnes pratiques sur les conditions de création et d'enregistrement** des informations par les producteurs.
  - *Repérer préalablement les données importantes*
  - *Les classer*
  - *Les documenter*
  - *Gérer leur cycle de vie (durée de conservation nécessaire pour des raisons juridiques)*
  - *Influence du « records management »*
  - *De plus en plus, les producteurs prennent appui sur des outils appelés « Gestion électronique de documents » (GED)*

Création de  
Mise à jour  
date de début des faits  
transmission de domaine juridique  
GALERIES THÉMATIQUES  
Découvrir le fa  
Archives de France

---

**Quels supports pour la conservation ?**





## Les besoins

---

- Les opérations que l'on souhaite faire sur les supports
  - Du stockage de données
  - De la conservation des données
  - Des accès aux données



## Les besoins

- Le stockage des données
  - Disposer de supports capacitifs :
- La conservation des données
  - Disposer d'un support fiable
  - Disposer d'un support durable
- L'accès aux données
  - Accès en écriture
  - Accès en lecture
  - Faire en sorte qu'une charge importante d'un type d'accès ne pénalise pas fortement les autres accès

## Quels supports pour la conservation ?

- Que choisir comme support ?
  - Des supports amovibles optiques : CD ou DVD-R
  - Principales limites : la volumétrie
  - Procédures très manuelles à moins d'opter pour des systèmes automatisés de récupération des supports
  - Pas d'outils sur le marché qui permettent par exemple de procéder à des migrations massives de données inscrites sur un DVD-R vers d'autres DVD-R, voire d'autres types de supports



## Quels supports pour la conservation ?

- Le stockage sur disques durs / serveurs
- Moins onéreux, plus simples, se prêtent à des volumétries importantes
- Mais sans maîtrise technique
- Stratégies de contournement



## Quels supports pour la conservation ?

- Le stockage sur bandes magnétiques
  - Volumétries de données importantes
  - Bonne résistance physique au temps
  - Limites





## Quels supports pour la conservation ?

- Migration des supports
- Systèmes de sauvegarde / de redondance / de réplication sur des sites distants
- Contrôle et surveillance des supports : mise en place de cycles de renouvellement des matériels



## Conclusion

- Il n'y a pas de « bon » support
  - Il existe des supports plus ou moins bien adaptés pour l'usage que l'on veut en faire (conservation pérenne, accès rapide...)
  - Il existe des supports plus ou moins abordables pour ses moyens



Archives de France

## Les formats







## De quoi sont fait les fichiers ?

- Un fichier est une suite consécutive de bit (train de bits)
  - Dont le système de fichier connaît des propriétés (nom, localisation, dates, droits)
  - Qui possède une syntaxe (organisation des bits)

```
01010010 01000101 01000110 01000110
00110000 00000000 00000001 00000000
01010111 01000001 01010110 01000101
00010000 00000000 00000000 00000000
00000001 00000000 00000001 00000000
00010001 00101011 00000000 00000000
00010001 00101011 00000000 00000000
10000000 10000000
```



## Conséquences sur la conservation

- Le train de bits n'est pas auto-suffisant
  - Il convient de conserver en même temps que le train de bits, les informations de décodage (= les « spécifications » du format).
  - Pour les formats conteneurs, il faut également connaître les formats et codages utilisés dans les contenus.
  - La description de certains formats ne sont que des sur-couches d'autres formats, il faut donc également connaître les spécifications des formats parents.
  - Si les spécifications sont déjà conservées ailleurs on peut se contenter de conserver des informations d'identification de ce format
  - Importance du choix du format des métadonnées



## Formatage

- Formats codant plusieurs types d'information
  - Par exemple un fichier odt (OpenDocument Text) code du texte mais aussi des images, des informations de mise en page, etc.
- Formats pouvant coder un même type d'information de plusieurs manières
  - WAV autorise une bonne centaine de codages audio dont PCM, MP3, mu-law, a-law, etc.
- Formats codant des métadonnées techniques
  - Les images JPEG peuvent contenir par exemple des métadonnées (exif)
- Formats répartissant l'information dans plusieurs fichiers
  - .xml (.xsd, include, import), plans d'archis ARC+ (.iii, .ddd, .icn)



## Les critères d'évaluation

- Spécifié : il existe une description de la convention utilisée pour représenter la donnée et celle-ci est suffisamment décrite pour en développer une implémentation complète.
- Ouvert : la convention est publique (sinon le format est dit fermé). Elle est donc sans restriction d'accès ni de mise en œuvre.
- Normalisé : la convention est adoptée par des organismes de normalisation (ISO, W3C). Exemple : le PDF/A.
- Standardisé : il n'existe pas de norme sur ce format mais son utilisation est tellement répandue qu'il est considéré comme un standard. Exemple : le PDF. ATTENTION : en anglais " standard " signifie " norme ".
- Propriétaire : si l'exploitation du format entre dans le cadre du droit privé, il dépend alors de l'existence du propriétaire (Par exemple les formats doc de Microsoft). Cet aspect propriétaire n'empêche par forcément les spécification d'être publiées. Exemple : les spécification du format PDF ont toujours été publiées par Adobe.

(Source : CNES, *Référentiel normatif, Ingénierie des données. Evaluation des formats de données par rapport à la pérennisation, 2005*)



## Les critères d'évaluation

Ces cinq critères permettent de définir le niveau de pérennité d'un format :

- 1. un format de données doit être entièrement et explicitement spécifié et sa spécification doit être connue du service en charge de l'archivage long terme ;
- 2. un format de données doit être apte à représenter la sémantique et la complexité de l'information à préserver ;
- 3. l'usage de formats normalisés est recommandé. En outre, on évitera l'usage d'éléments propriétaires au sein d'un format normalisé ;
- 4. lorsque le besoin de pouvoir modifier un document est identifié, le choix du format doit prendre en compte cette contrainte ;
- 5. le choix des formats doit pouvoir prendre en compte la disponibilité et le coût des outils d'exploitation ;



## Les critères d'évaluation

(suite)

- 6. il doit être possible de vérifier automatiquement qu'un fichier de données respecte les spécifications du format et respecte également les règles d'utilisation et des restrictions qui auront été éventuellement définis pour la pérennité ;
- 7. la possibilité d'extraire automatiquement tout ou partie des métadonnées à partir des données constitue un avantage certain ;
- 8. l'usage de formats inutilement volumineux sera évité ;
- 9. à fonctionnalités égales un format simple est préférable à un format complexe ;
- 10. les formats largement reconnus et utilisés seront privilégiés ;
- 11. le choix d'un format doit prendre en compte la disponibilité et le coût des outils de transformation des formats et de représentation des données.



## Familles de formats

### ■ Formats bureautiques et non structurés

#### – PDF - PDF/A

- PDF - format propriétaire (ADOBE)
- PDF/A-1 est devenu la norme ISO 19005-1 en 2005
- PDF 1.7 est devenu la norme ISO 32000-1 en 2008
- PDF/A-2 nouvelle version de PDF-A publiée. (juin 2011): elle s'appuie sur la version 1.7 de PDF lui même normalisé en ISO 32000-1

#### – ODF, Open Document Format, pour les documents bureautiques

- La version 1.0 est standardisé par OASIS en 2005 puis normalisée par l'ISO en 2006 (ISO 26300)
- La version 1.1, est standardisée par Oasis en 2007, puis l'ISO en 2012
- La version 1.2 est actuellement un standard Oasis depuis 2011

#### – OOXML, Office Open XML, pour les documents bureautiques

- Standardisé par ECMA en 2006 ; Normalisé par ISO en 2008 (ISO 29500)



## Familles de formats

### ■ Formats image

- PNG, Portable network Graphics,
  - pour les images matricielle (à l'origine crée pour offrir une alternative libre au format GIF qui utilisée une technique de compression sans perte LZW soumise à un brevet). Standardisé par le W3C en 1996 ; Normalisé par l'ISO en 2004 (ISO 15948:2004)
- GIF, Graphics Interchange Format
  - un format ouvert de la société CompuServe pour la représentation d'images matricielle. Il utilise une compression sans perte LZW dont le brevet a maintenant expiré. Version GIF87a ; Version GIF89a permet l'inclusion de plusieurs images dans un même fichier
- JFIF, JPEG File Interchange Format
  - un format pour la représentation d'images matricielles compressées avec l'algorithme JPEG (Joint Photographic Experts Group) ; L'algorithme JPEG est définit par la norme ISO 10918-1 en 1993





## Familles de formats

### ■ Formats image (suite)

- JPEG2000
  - Norme ISO 15444-1 de 2000 utilisant pour la compression (avec ou sans perte) des images un algorithme une transformée en ondelettes permettant des meilleurs taux que l'agorithme de la norme ISO 10918-1.
- TIFF, Tagged Image File Format
  - format conteneur propriétaire de Adobe pour des images numériques.



## Familles de formats

### ■ Formats pour les bases de données

- Export de tables en CSV, en XML
- SIARD,
  - Note d'information DGP/SIAF/2010/017 du 21 septembre 2010. Étude du format SIARD pour l'archivage des bases de données relationnelles et au logiciel SIARDSuite mettant en œuvre ce format.

### ■ Formats audio-visuels

- Guide méthodologique pour le choix de formats numériques pérennes dans un contexte de données orales et visuelles : nouvelle version mise à jour en 2011 ; Note d'information DGP/SIAF/2010/010 du 21 mai 2010



## Exemples

- Validation : <http://facile.cines.fr/>
- Identification et validation :  
<http://projects.iq.harvard.edu/fits/home>
- Base de formats :  
<http://apps.nationalarchives.gov.uk/PRONOM/BasicSearch/proBasicSearch.aspx?status=new>



## Les métadonnées



## Métadonnées sémantiques

A8600200046A36	00900879	2001985363
7183		
P8600200046001000P		
PRE M	HABARI	
APARI		BEN
	PEN	1940

*Extraction d'une base de données, lisible mais sans signification*



## Métadonnées sémantiques

**A8600200046A36**

AFFAIRE - Numéro d'affaire ou numéro de parquet : ce numéro est constitué des zones suivantes :

- année sur 2 chiffres
- quantième du jour dans l'année sur 3 chiffres de 001 à 365 (ou 366)
- numéro d'ordre sur 4 chiffres au choix de l'utilisateur
- une clé sur 1 chiffre

### Nomenclatures employées pour coder les informations

- exemple : nomenclature NATAFF (nature de l'affaire) : B24 Vol par ruse ou usurpation d'identité

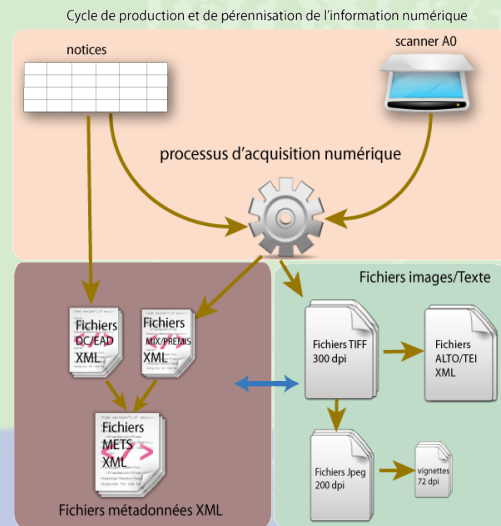
### Information sémantique

Complément d'information permettant d'interpréter la donnée d'un point de vue utilisateur.

Ex: Identification des champs  
Les nomenclatures

## La production de métadonnées

- La production de métadonnées peut s'envisager à différentes étapes du cycle de vie du document porteur d'information :
  - saisie à la création
  - saisies au cours d'une étape de traitement de l'information
  - extraction à posteriori : transformation, diffusion, récupération



## Des formats de représentation de métadonnées



*Des métadonnées descriptives et intellectuelles*



*Des formats d'emballage d'ensemble de métadonnées*



*Des métadonnées pour la préservation des objets archivés*





## Encoder et lier les métadonnées

- XML (Extensible Markup Language)
  - langage à balises permettant de structurer et de décrire un ensemble d'objet au sein d'espace de noms
  - langage hiérarchique permettant de décrire des relations
  - langage de transport d'information permettant de standardiser les échanges de données
  - peut être décrit par un schéma ou une DTD
  - peut servir au stockage des données mais également à leur interrogation

```
<xml version="1.0" encoding="UTF-8"?>
<ex:collection xml:lang="fr" xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns=http://www.w3.org/1999/xhtml xmlns:ex="http://exemple.org">
<ex:vehicule>
<dc:title>Mon super vélo</dc:title>
<velo id="2">
<taille>76</taille>
<marque>cool bike</marque>
<modele>speeder</modele>
</ex:velo>
<ex:velo taille="76" marque="coolbike" modele="speeder" id="2" />
```

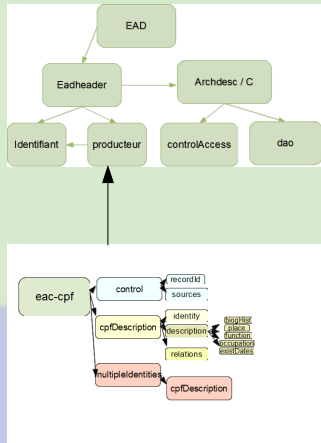


## Encoder et lier les métadonnées

- EAD (Description archivistique encodée) et EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles) pour la description archivistique
  - La DTD EAD est basée sur ISAD(G). L'EAD permet de décrire un fonds d'archives ou d'ouvrages de manière hiérarchique du général au particulier, d'indexer les notices descriptives et de lier ces descriptions à des substituts numériques et à des notices de producteurs EAC
  - Le schéma EAC vise à décrire les producteurs pour retracer l'historique d'une entité administrative ou morale d'archives en se basant sur la norme ISAAR(CPF).
  - Il permet une redéfinition de la notion de contexte de production et d'utilisation des archives. Il se concentre seulement sur la description des collectivités, des personnes et des familles.

## Encoder et lier les métadonnées

- EAD (Description archivistique encodée) et EAC-CPF (Contexte archivistique encodé – Collectivités, personnes, familles) pour la description archivistique



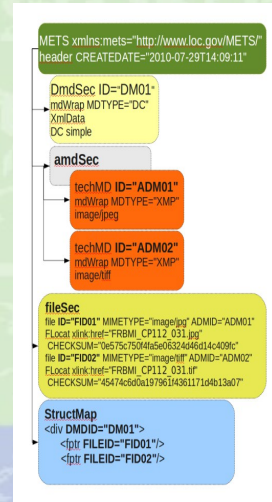


## PREMIS : la structuration des objets d'information

- PREMIS (Preservation Metadata: Implementation Strategies)
  - Le modèle de données de PREMIS est composé de 5 entités reliées entre elles par des identifiants: Entité intellectuelle, Objet, Événement, Agent, et Droits.
  - Ce schéma répond notamment au besoin de pérennisation de l'information numérique et de conservation du cycle de vie des données (import dans un entrepôt de donnée, migration, gestion d'intégrité, liens entre différentes représentations d'un même document).
    - [<http://www.loc.gov/standards/premis>]

## METS : la structuration de la description des ressources

- Le schéma METS vise à fournir un méta-modèle pour l'encapsulation d'autres formats de métadonnées.
- Ce schéma XML composé de 7 sections est extrêmement flexible.
- Il permet notamment de représenter l'information sous forme de carte de structure et s'implémenter le modèle conceptuel OAIS.
- Il constitue un format pivot entre la logique de diffusion et celle de conservation de l'information.





## Encoder et lier les métadonnées

- URI (Uniform Resource Identifier)
  - Les URI sont des identifiants Web.
  - Les URL sont des URI qui identifient la ressource par le moyen d'y accéder.
  - Les URL peuvent être pérennes.

Syntaxe :

<scheme:chaîne/de.caractère>

<http://monsite.com/dossier/fichier.html#ancree>

<http://mondomaine.org/ressource/1234>

- un préfixe qui indique le contexte dans lequel l'identifiant est attribué (par ex. http:,ftp:, urn:, etc.)
- un élément qui permet de désigner l'autorité nommante qui a attribué l'identifiant au sein de ce système
- le « nom » lui-même, c'est-à-dire une chaîne de caractères qui identifie la ressource de manière unique, au sein de ce système et pour cette autorité.

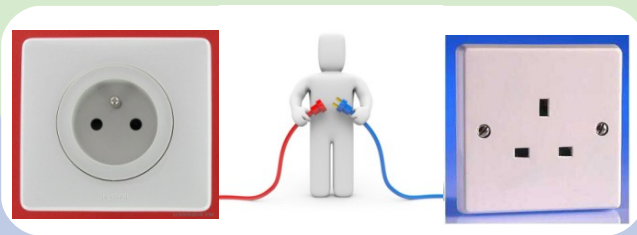
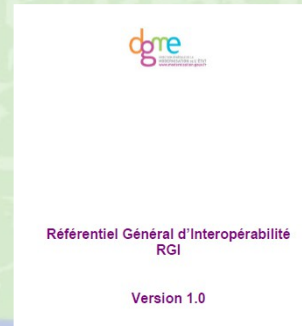


## Référentiel général d'interopérabilité

- Normes et standards que les autorités administratives doivent utiliser afin de garantir l'interopérabilité des informations échangées et de leurs systèmes d'information.

- Publié en novembre 2009.

[references.modernisation.gouv.fr/sites/default/files/RGI\\_Version1%200.pdf](http://references.modernisation.gouv.fr/sites/default/files/RGI_Version1%200.pdf)







## Importance des formats ouverts

- Ex. : les formats de données et les recommandations du Référentiel Général d'Interopérabilité à destination des systèmes d'information des administrations

<b>Recommandé</b>	<b>Formats d'image</b>
Pour les images numériques, il est RECOMMANDÉ de choisir parmi les formats GIF, PNG, JPEG, TIFF ou DNG.	

<b>Recommandé</b>	<b>XML / CSV</b>	<b>W3C</b>	
Il est RECOMMANDÉ d'utiliser XML ou CSV pour l'exportation de tout ou partie d'une base de données.			

<b>En observation</b>	<b>ODF</b>	<b>OASIS</b>	
Des améliorations de la norme concernant l'accessibilité ont été apportées avec la version 1.1, approuvée par l'OASIS en octobre 2006. La majorité des implémentations du marché reprennent cette dernière version.			
La version ODF 1.2 est en cours d'élaboration et devrait être soumise à l'approbation OASIS et ISO en 2009.			

<b>En observation</b>	<b>Office Open XML</b>	<b>ISO</b>	
Le format Office Open XML est un format bureautique basé sur XML. Il supporte nativement une partie des formats binaires bureautiques existants. Il n'existe pas à ce jour d'implémentation de cette norme.			

<b>Recommandé</b>	<b>PDF/A</b>	<b>ISO</b>	
Il est RECOMMANDÉ d'utiliser le format PDF/A, pour l'archivage des documents bureautiques statiques non révisables.			



# Agence nationale de la sécurité des systèmes d'information

Que faire en cas d'incident ? | Le site du CERTA | Portail de la sécurité informatique/Documentation

Vous êtes ici : [Accueil](#) > [Réglementation SSI](#) > Référentiel général de sécurité

## Référentiel général de sécurité



### Présentation

Le Référentiel général de sécurité (RGS) est créé par l'article 9 de l'ordonnance n° 2005-1516 du 8 décembre 2005 relative aux échanges électroniques entre les usagers et les autorités administratives et entre les autorités administratives. Ses conditions d'élaboration, d'approbation, de modification et de publication sont fixées par le décret n° 2010-112 du 2 février 2010 pris pour l'application des articles 9, 10 et 12 de l'ordonnance citée relatif à la sécurité des informations échangées par voie électronique.

La présente version du RGS (1.0), rendue officielle par arrêté du Premier ministre en date du 6 mai 2010 et publiée ci-dessous, est le résultat d'un travail conjoint entre la Direction générale de la modernisation de l'État (DGME) et l'Agence nationale de la sécurité des systèmes d'information (ANSSI), qui a pris en compte les nombreuses contributions et remarques reçues à la suite d'appels à commentaires adressés aux ministères et au public en 2008, et de présentations faites aux diverses autorités administratives, parmi lesquelles notamment des services de l'État et des représentants des collectivités territoriales.

### Grandes dates

- ◊ notification à la Commission Européenne, en application de la directive 98/34/CE, le 25 mars 2009 ;
- ◊ avis favorable de la Commission consultative d'évaluation des normes (CCEN) le 7 mai 2009 ;
- ◊ publication du décret n° 2010-112 du 2 février 2010, dit décret RGS, le 4 février 2010 au Journal Officiel ;
- ◊ publication le 18 mai 2010 au Journal Officiel de l'arrêté du 6 mai 2010, dit arrêté RGS, portant approbation de la première version du RGS.

### Errata

Dans l'attente de la seconde version du RGS, l'ANSSI souhaite d'ores et déjà assouplir la règle relative au double usage des clés privées au sein des certificats d'authentification serveur, publiée le 6 mai 2010.

- ❖ [Liste des documents constitutifs du RGS v1.0](#)
- ❖ [Documents référencés par le RGS v1.0](#)
- ❖ [Processus de mise à jour du RGS](#)
- ❖ [Appel à commentaires : RGS 2.0](#)

Près de deux ans après la publication de la première version du Référentiel général de sécurité, l'ANSSI souhaite faire évoluer certains documents constitutifs du RGS et y ajouter d'autres annexes.



## La politique d'archivage (P2A)

- Étude en partenariat avec la DCSSI (cabinet Caprioli et Associés)
  - Publication du référentiel : juillet 2006 (<http://www.ssi.gouv.fr/IMG/pdf/ArchivageSecurise-P2A-2006-07-24.pdf>)
- Les livrables :
  - une politique d'archivage type pour le secteur public -à décliner ensuite sur place pour une déclaration des politiques d'archivage (le comment)
  - une prise en compte non seulement de la sécurité (réducteur) mais également du métier (fonctionnalités)

# Documentation ANSSI pour l'archivage électronique sécurisé



[http://  
www.ssi.gouv.fr/fr/bonnes-pratiques/outils-methodologiques/archivage-electronique-secure.html](http://www.ssi.gouv.fr/fr/bonnes-pratiques/outils-methodologiques/archivage-electronique-secure.html)

Vous êtes ici : [Accueil](#) > [Bonnes pratiques](#) > [Outils méthodologiques](#) > Archivage électronique sécurisé



## Archivage électronique sécurisé

9 juillet 2009




### LES ANALYSES PRÉALABLES

 <b>La synthèse des enjeux juridiques</b> PDF - 763.7 ko V.2006-05-16	 <b>L'état de l'art juridique, technique, organisationnel et des offres</b> PDF - 1 Mo V.2006-11-29
---	---

### LES DOCUMENTS INTRODUCTIFS

 <b>La plaquette</b> PDF - 74.6 ko V.2006-08-18	 <b>Le mémento</b> PDF - 278.4 ko V.2006-05-16
--	---

### LES DOCUMENTS D'AIDES À L'ÉLABORATION DU RÉFÉRENTIEL

 <b>Le cahier des charges pour un système d'archivage électronique</b> PDF - 609.4 ko V.2006-05-16	 <b>La politique et les pratiques d'archivage</b> PDF - 545.7 ko V.2006-07-24	 <b>La grille d'audit Excel</b> Excel - 35.5 ko V.2006-07-24
--	---	---



## DISIC

- Création de la **direction interministérielle des systèmes d'information et de communication de l'Etat (DISIC)** par décret en date du **21 février 2011** (décret n° 2011-193) – A rejoint en octobre 2012 le Secrétariat général pour la modernisation de l'action publique (SGMAP)
  - *Chargée notamment d'organiser toutes les opérations de mutualisation des systèmes d'information entre les différentes administrations de l'État. Sorte de DSI interministérielle.*
  - *Objectifs : anticiper les évolutions technologiques, les besoins, fixer les orientations en matière d'externalisation, de sous-traitance, etc. Pour une simplification des relations de l'utilisateur avec l'administration.*
  - *Missions : cartographie des SI de l'Etat, consultation sur les projets d'un certain coût.*
  - **Elle organise et anime le cadre de concertation nécessaire à l'évolution des référentiels généraux d'interopérabilité et d'accessibilité, des modèles de données de référence et des modèles d'échange et, en liaison avec l'Agence nationale de la sécurité des systèmes d'information, du référentiel général de sécurité ;**
  - **A un mandat d'archivage électronique** dont le but est à la fois de sensibiliser les DSI des ministères à la question de l'archivage électronique et de diffuser un guide de bonnes pratiques sur ce sujet.



## Livrables DISIC sur le mandat archivage électronique

- Accessibles sur le site [references.gov.fr](http://references.gov.fr)
  - <http://references.modernisation.gov.fr/rgi-interoperabilite>

The screenshot shows the website interface for 'RGI - Interopérabilité'. At the top left, there are logos for 'Ministère de l'Économie, des Finances et du Développement Durable' and 'Ministère de l'Énergie, du Développement Durable et de la Mer'. A blue banner reads 'AVEC VOUS l'administration SE MODERNISE'. A search bar with a 'RECHERCHER' button is on the right. Below the search bar is a navigation menu with categories: 'URBANISATION SI ETAT', 'RGI - INTEROPÉRABILITÉ', 'RGS ET RÉFÉRENCIEMENT - SÉCURITÉ', 'RGAA - ACCESSIBILITÉ', and 'CHARTRE INTERNET DE L'ÉTAT'. Under the 'RGI' tab, there are sub-menus: 'Volet technique', 'Volet sémantique', 'Volet organisationnel', and 'Archivage numérique'. The main content area shows 'Accueil » RGI - Interopérabilité' and the title 'RGI - interopérabilité'. A button on the right says 'TÉLÉCHARGER LA VERSION PDF', and below it, a file icon and the text 'RGI\_Version1 0.pdf' are visible.