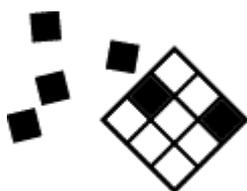




# L'EAD, une DTD

## pour la description archivistique

Claire Sibille  
Archives de France  
Juin 2003



**Un atelier organisé sous l'égide du  
Groupe de travail pour l'archivage de documents électroniques  
de l'AAS**

AAS Association des Archivistes suisses  
VSA Verein Schweizerischer Archivarinnen und Archivare  
AAS Associazione degli archivisti svizzeri  
UAS Uniu da las archivarias e dals archivaris svizzers

# L'EAD, une DTD pour la description archivistique

## 1. Concepts de base : SGML, (HTML), XML, DTD

### 1.1. SGML

### 1.2. HTML

### 1.3. XML

1.3.1. *Qu'est-ce que le XML ?*

1.3.2. *Avantages de XML*

1.3.3. *Les balises en XML*

1.3.4. *Les composants du balisage en XML : les éléments*

1.3.5. *Les composants du balisage en XML : les attributs*

1.3.6. *Les composants du balisage en XML : les entités*

1.3.7. *Analyseur XML*

1.3.8. *Qu'est-ce qu'un document bien formé ?*

1.3.9. *La structure d'un document XML*

### 1.4. DTD

1.4.1. *Qu'est-ce qu'une DTD ?*

1.4.2. *Quelques exemples de DTD*

## 2. Présentation sommaire de l'EAD (version 1.0)

### 2.1. Origine de l'EAD

### 2.2. Caractéristiques essentielles de l'EAD

### 2.3. Structure

2.3.1. *Les trois éléments principaux*

2.3.2. *<eadheader> En-tête EAD*

2.3.3. *<eadheader> : exemple*

2.3.4. *<frontmatter> Préliminaires*

2.3.5. *<archdesc> Description archivistique*

2.3.5.1. *<did> Identification et description de l'unité documentaire*

2.3.5.2. *<add> Données descriptives additionnelles*

2.3.5.3. *<adminfo> Informations de gestion*

2.3.5.4. *<controlaccess> Vedettes et accès contrôlés*

2.3.5.5. *<dsc> Description des sous-composants*

### 2.4. L'EAD et la norme ISAD (G)

2.4.1. *Schéma des niveaux de classement d'un fonds*

2.4.2. *ISAD (G) : 26 éléments de description répartis en 7 zones*

## 3. Exemples d'utilisation de l'EAD

### 3.1. L'EAD en France

### 3.2. L'EAD au Centre historique des Archives nationales

3.2.1. *L'EAD au Centre historique des Archives nationales : premier projet*

3.2.2. *L'EAD au Centre historique des Archives nationales : deuxième projet*

3.2.3. *L'EAD au Centre historique des Archives nationales : la rétroconversion*

### 3.3. Produire (et mettre à jour) les inventaires directement en XML

### 3.4. Diffuser des documents EAD sur Internet

## 4. Pour en savoir plus...

## 1. Concepts de base :

### Les langages de structuration électronique

- Les nouveaux langages informatiques permettent d'encoder des textes saisis sur ordinateur, c'est-à-dire d'y introduire des balises qui en dégagent la structure et guident l'interrogation
- L'encodage le plus élémentaire consiste à introduire des liens hypertextes, tels ceux qu'utilisent les fichiers écrits en HTML et consultables sur les sites web (possibilité de sélectionner sur une table des matières le titre d'un chapitre et d'accéder directement au texte de ce chapitre)
- Mais les langages plus récents (SGML, XML) permettent de concevoir des encodages plus sophistiqués, adaptés à la structure de catégories de documents homogènes : c'est ce que l'on appelle les DTD (définition de types de documents)
- Une DTD spécifique aux instruments de recherche d'archives a été développée en partenariat entre la SAA et les grandes bibliothèques américaines : c'est la DTD EAD (Encoding Archival Description) qui s'appuie sur les éléments descriptifs définis par la norme ISAD (G)
- L'encodage des textes permet la création d'instruments de recherche offrant à la fois :
  - la qualité de présentation matérielle et la possibilité de consultation par feuilletage d'un instrument de recherche édité sur papier
  - la souplesse d'interrogation d'une base de données
- De plus, la mise en ligne d'un instrument de recherche permet de le mettre à jour régulièrement, de le compléter et de l'améliorer

## 1.1. SGML

Standard Generalized Markup Language (=Langage normalisé de balisage généralisé)

ISO 8879 (1986) - AFNOR NF Z 71-010 (1990)

- Successeur du format ML (Markup Language) puis GML (General Markup Language) qui remonte à la seconde moitié des années 1960
- Origine dans le monde de l'édition (années 1980)
- Définit la structure logique d'un document électronique (en-têtes, titres, chapitres, paragraphes, notes,...)
- Applications documentaires dans les bibliothèques ou les centres de documentation
- Très utilisé dans le monde de l'édition et de l'impression
- Lourd à mettre en œuvre et inadapté au Web

## 1.2. HTML

HyperText Markup Language (=Langage de balisage hypertextuel)

- Standard de fait pour les publications sur l'Internet
- Définit plus la structure physique (retraits, marges, justification, couleurs, ...) que la structure logique du document électronique
- Fonction de navigation : liens dans la même page ou vers une autre page du même document, vers d'autres applications Internet, des objets multimédias, des programmes, pointeurs pour les liens (URL : Uniform Resource Locators)
- Peine à gérer les documents de gros volume ou fortement structurés, où la recherche plein texte est insuffisante

## 1.3. XML

eXtended Markup Language (=Langage de balisage étendu)

### 1.3.1. *Qu'est-ce que le XML ?*

- Créé en 1998, XML (eXtensible Markup Language) est un langage de description et d'échange de documents structurés, version simplifiée du SGML adaptée au Web
- XML sert à décrire la structure logique de documents textuels, à l'aide d'un système de balises permettant de marquer les éléments qui composent cette structure et les relations entre ces éléments
- Ne traite pas de la mise en forme du document, qui sera assurée par une feuille de style externe

#### *Exemple de page codée en HTML*

```
<H2>Bibliographie XML</H2>
```

```
<UL>
```

```
  <LI>Jean-Christophe Bernadac et François Knab,<I>Construire une application XML</I>,
  Paris, Eyrolles, 1999</LI>
```

```
  <LI>Alain Michard,<I>XML, Langage et Applications</I>, Paris, Eyrolles, 1998</LI>
```

```
  <LI>William J. Pardi,<I>XML en Action</I>, Paris, Microsoft Press, 1999, adapté de
  l'anglais par James Guerin</LI>
```

```
</UL>
```

#### *Exemple de page codée en XML*

```
<?xml version="1.0" encoding="ISO-8859-1"?>
```

```
<BIBLIO SUBJECT="XML">
```

```
  <BOOK ISBN="9782212090819" LANG="fr" SUBJECT="applications">
```

```
    <AUTHOR>
```

```
      <FIRSTNAME>Jean-Christophe</FIRSTNAME>
```

```
      <LASTNAME>Bernadac</LASTNAME>
```

```
    </AUTHOR>
```

```
    <AUTHOR>
```

```
      <FIRSTNAME>François</FIRSTNAME>
```

```
      <LASTNAME>Knab</LASTNAME>
```

```
    </AUTHOR>
```

```
    <TITLE>Construire une application XML</TITLE>
```

```
    <PUBLISHER>
```

```
      <NAME>Eyrolles</NAME>
```

```
      <PLACE>Paris</PLACE>
```

```
    </PUBLISHER>
```

```
    <DATEPUB>1999</DATEPUB>
```

```
  </BOOK>
```

### **1.3.2. Avantages de XML**

- En XML, les auteurs inventent librement les balises qui leur paraissent utiles pour marquer les éléments composant le document. En acceptant un ensemble de balises extensible, XML fournit une description plus précise du contenu du document. Ceci oppose XML à HTML, dont les balises sont définies une fois pour toutes
- XML est basé sur le système UNICODE (norme ISO 10646) soit 4 294 967 296 possibilités de caractères codés sur 4 octets
- Si une communauté de travail utilise une même structure type de document (grammaire), les documents qu'elle produit peuvent être échangés sans problème
- XML est indépendant des plates-formes techniques et produit des fichiers autodocumentés
- Les documents produits peuvent donc être archivés au format XML sans risque de devenir illisibles
- A partir d'un fichier structuré en XML et en utilisant diverses feuilles de style XSL (eXtensible Stylesheet Language), on peut produire des présentations différentes du contenu des documents, en fonction des supports d'édition ou des publics visés
- Certains outils informatiques du Web sont capables de faire des recherches dans un document XML en exploitant son marquage structurel, ce qui améliore considérablement la qualité des réponses par rapport à une recherche dans un fichier HTML

### **1.3.3. Les balises en XML**

- Une balise est une chaîne de caractères du type <balise>
- Une balise ouvrante se compose du caractère < suivi du nom de la balise, du ou des attributs utilisés et enfin du caractère > (<balise>)
- Une balise fermante se compose des caractères </ suivis du nom de la balise et du caractère > (</balise>)

### **1.3.4. Les composants du balisage en XML : les éléments**

- Les éléments permettent de décomposer le texte en unités d'information
- Un élément est composé d'une balise ouvrante, d'une balise fermante, et d'un contenu encapsulé entre ces deux balises  
Exemple : <titre>Les composants du balisage en XML : les éléments</titre>
- Un élément peut contenir du texte et/ou d'autres éléments, ou être vide

### 1.3.5. *Les composants du balisage en XML : les attributs*

- Qualifient les éléments  
ex. : code de langue du titre
- Sont saisis à l'intérieur de la balise ouvrante de l'élément
- La valeur d'un attribut est une chaîne de caractère encadrée par des guillemets ou des apostrophes : <balise Nom="Valeur">

```
<LIVRE SUJET="XML"  
<chapitre id ="ch1" titre='Introduction à XML'> </chapitre>
```

### 1.3.6. *Les composants du balisage en XML : les entités*

- Chaînes de caractères qui composent le document ou des documents externes dans un autre codage (image, son...) qui sont attachés à un endroit précis
- Entités internes :  
entités caractères prédéfinies ou non, pour saisir certains caractères que l'on doit obligatoirement coder (>, >, &, “, ’) ou difficiles à composer au clavier  
entités texte, servant à donner un nom à des expressions ou phrases souvent répétées
- Entités externes :  
texte : fragments XML formant chacun un fichier  
non parsées (fichiers non XML, tels qu 'images, enregistrements sonores...)

### 1.3.7. *Analyseur XML*

- Un analyseur (généralement appelé parseur) est un outil logiciel permettant de vérifier que le document XML respecte la syntaxe XML de base
- Si une DTD est déclarée en tête de document XML, le parseur vérifiera également la conformité du document avec la DTD
- Les analyseurs XML sont intégrés dans les navigateurs Netscape 6.0, Internet Explorer 5.5 et dans les éditeurs conçus pour XML, tels que XMetaL ou XML Notepad

### 1.3.8. *Qu'est-ce qu'un document bien formé ?*

- Obéit aux règles syntaxiques de XML
- Sera traité avec succès par un parseur
- Si un document n'est pas bien formé, il provoquera toujours un message d'erreur
- La recommandation XML 1.0 décrit les règles se rapportant aux documents syntaxiquement corrects, telles que : la première ligne du document doit correspondre à la déclaration de document XML, le document doit contenir au moins un élément (ou balise), à chaque balise d'ouverture doit correspondre une balise de fermeture...

### 1.3.9. *La structure d'un document XML*

- Tout document XML se compose d'un prologue, qui contient un certain nombre de déclarations (version du langage XML utilisée, déclaration de DTD), d'un arbre d'éléments, et de commentaires (qui peuvent être inclus dans le document)
- L'arbre d'éléments est formé d'un et d'un seul élément racine qui contient tous les autres éléments ; tout élément fils est inclus dans son père, il ne peut y avoir de recouvrement d'éléments

## 1.4. DTD

Définition de Type de Document (=Document Type Definition)

### 1.4.1. *Qu'est-ce qu'une DTD ?*

- Une DTD est l'ensemble des règles d'encodage que doit suivre un document XML
- Ces règles définissent le nom et le contenu de chaque balise et le contexte dans lequel elles doivent exister
- Une DTD se compose des éléments logiques qui constituent le document, des attributs de qualification de ces éléments et des entités
- La quasi-intégralité des DTD écrites en SGML ont été traduites en XML
- XML n'impose pas l'utilisation d'une DTD pour un document XML mais le respect des règles de base de XML est obligatoire
- Un document bien formé est, pour rappel, un document qui respecte les règles générales de syntaxe du XML
- Un document valide est un document bien formé qui obéit en outre à une structure type définie explicitement dans une DTD

### 1.4.2. *Quelques exemples de DTD*

- ISO 12083 pour les imprimés, <http://www.xmlxperts.com/xmlbookdtd.htm>
- ISO 12083 pour les périodiques, <http://www.xmlxperts.com/xmlserialdtd.htm>
- ISO 12083 pour les articles de périodiques, <http://www.xmlxperts.com/xmlarticledtd.htm>
- ISO 12083 pour les formules mathématiques, <http://www.xmlxperts.com/xmlmathdtd.htm>
- TEI (Text Encoding Initiative) pour le traitement des textes publiés sous forme électronique, notamment en sciences humaines et en linguistique, <http://www.tei-c.org>
- MASTER pour les manuscrits médiévaux, <http://www.cta.dmu.ac.uk/projects/master/>
- EBIND pour lier les métadonnées au document numérique, <http://sunsite.berkeley.edu/Ebind>
- XMLMARC, <http://xmlmarc.stanford.edu/>
- DTD USMARC, <http://lcweb.loc.gov/marc/marcdtd/marcdtdback.htm>
- DTD Unimarc BiblioML (bibliographique et autorités), <http://www.culture.fr/BiblioML>.
- DTD EAD pour le traitement des fonds d'archives et des collections spécialisées, <http://lcweb.loc.gov/ead/>

## 2. Présentation sommaire de l'EAD (version 1.0)

### 2.1. Les origines de l'EAD

- Provient du monde des bibliothèques
- Développée à l'Université de Berkeley pour remplacer le standard MARC-AMC (Archival and Manuscript Control), insuffisant pour les descriptions à plusieurs niveaux
- 1996 : version expérimentale de l'EAD
- Août 1998 : première version officielle de l'EAD
- Septembre 2002 : révision de l'EAD (EAD 2002)
- Propriété de la Society of American Archivists
- Maintenance par la Library of Congress
- Formation par RLG de plus de 400 professionnels américains et canadiens dès 1997
- DAF membre du WG de la SAA depuis septembre 2000

#### *La documentation officielle de l'EAD se compose de :*

- La DTD proprement dite, écrite en anglais (et non modifiable sauf par ses auteurs)
- La Tag Library (dictionnaire des balises)
- Les Application Guidelines (guide d'application)
- Le guide d'application et le dictionnaire des balises décrivent les éléments balisés à utiliser, leur contenu, leurs attributs et les valeurs de ces attributs, ainsi que l'ordre de saisie des éléments et les emboîtements requis
- Cette documentation est disponible sur le site officiel de l'EAD : <http://lcweb.loc.gov/ead/>

#### *La Tag Library en français*

- Un groupe de travail AFNOR (CG 46/CN 357, « Modélisation, produit et accès aux documents ») vient de traduire en français la Tag Library ; la première version officielle de cette traduction est en ligne sur le site de la DAF  
<http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/index.html>
- Le groupe AFNOR/EAD prépare un manuel d'application de l'EAD

## 2.2. Caractéristiques essentielles de l'EAD

- L'EAD permet d'encoder les inventaires et répertoires, s'applique bien aux catalogues, états sommaires et états des fonds, ne s'applique pas aux guides ni aux bordereaux de versement (critiques australiennes)
- Pas de limite de taille pour les instruments de recherche encodés
- Application possible pour tous les types de fonds (papiers, objets, archives numériques...), dont les différents éléments sont liés entre eux par une structure hiérarchique selon le principe de la description à plusieurs niveaux
- L'EAD est conçue pour être utilisée avec des normes de description, en particulier ISAD (G)
- L'EAD facilite l'échange et la portabilité des instruments de recherche : il est possible d'utiliser le même document encodé pour des restitutions multiples
- L'EAD utilise largement des standards apparentés ou complémentaires : TEI, ISAAR(CPF), MARC (possibilité de faire des liens depuis une notice MARC décrivant un fonds de manière globale vers l'inventaire détaillé encodé en EAD ; inversement, possibilité de générer l'ossature d'une notice MARC à partir des éléments de haut niveau d'un document EAD)
- Les noms des éléments sont "neutres", afin de permettre de transcender les divergences de vocabulaire entre les différents types d'établissements
- L'EAD permet de créer de nouveaux instruments de recherche, mais elle est aussi très bien adaptée à des opérations de rétroconversion, notamment par la souplesse de l'ordre de ses éléments
- Il est possible de choisir la finesse de l'encodage en fonction des besoins et des moyens ; seuls quelques éléments sont obligatoires (en-tête EAD, description du fonds, niveaux hiérarchiques)
- L'EAD permet d'intégrer aux instruments de recherche des liens vers des représentations numériques des documents décrits (grâce à la balise <dao>, Objet archivistique numérique), possibilité particulièrement intéressante dans le cas de fonds essentiellement visuels ou numériques

## 2.3. Structure de l'EAD

- Deux segments importants : métadonnées (informations sur l'instrument de recherche lui-même) et information sur le contenu documentaire
- 145 éléments, dont un tout petit nombre obligatoires
- Les niveaux hiérarchiques de description ont les mêmes balises possibles

### 2.3.1. Les trois éléments principaux

- Sous l'élément racine <ead> :

<eadheader> en-tête EAD (décrit l'instrument de recherche)

<frontmatter> préliminaires (page de titre, préface)

<archdesc> description archivistique avec attribut de niveau de description

### 2.3.2. <eadheader>En-tête EAD

- Fournit des informations bibliographiques et descriptives sur l'instrument de recherche, et non sur les documents d'archives que celui-ci décrit
- Prend modèle sur l'élément en-tête de la TEI afin d'encourager la fourniture de métadonnées uniformes entre des documents de type différent)
- Quatre sous-éléments sont disponibles et doivent être agencés dans l'ordre suivant : Identifiant EAD <eadid> (obligatoire), Description du fichier <filedesc> (obligatoire), Description de profil <profiledesc> (facultatif) et Description des révisions <revisiondesc> (facultatif)
- <eadheader> possède sept attributs (findaidstatus : degré d'exhaustivité/de finesse des informations)

- <eadheader>
  - <eadid>**Identifiant EAD** (du fichier électronique)
  - <filedesc>**Description du fichier**
    - <titlestmt>*Mention de titre*
      - <titleproper>*Titre propre de l'inventaire*
      - <subtitle>*sous-titre de l'inventaire*
    - <author>*auteur*
  - <editionstmt>*Mention d'édition*
  - <publicationstmt>*Mention de publication*
  - <seriestmt>*Mention de collection*
  - <notestmt>*Mention de note*
  - <profiledesc>**Description du profil**
    - <creation>*Informations sur la création de l'inventaire*
  - <revisiondesc>**Descriptions des révisions ; permet de gérer les versions successives de l'instrument de recherche encodé**

### 2.3.3. <eadheader> : *exemple*

```
<eadheader findaidstatus= "edited-partial-draft" audience "external">
  <eadid>DAFANCH00AP_0000512AP</eadid>
  <filedesc>
    <titlestmt>
      <titleproper>Fonds Miromesnil</titleproper>
      <subtitle>Répertoire numérique détaillé </subtitle>
      <author>Guénola Lévesque</author>
    </titlestmt>
    <publicationstmt>
      <address><addressline>Paris</addressline></address>
      <publisher>Centre historique des Archives nationales</publisher>
      <date>2002</date>
    </publicationstmt>
  </filedesc>
  <profiledesc>
    <creation>Cet instrument de recherche a été converti en XML conformément à l'EAD par
    Claire Sibille, conservateur à la Section des archives privées.</creation>
    <langusage>Instrument de recherche rédigé en français</langusage>
  </profiledesc>
</eadheader>
```

### 2.3.4. <frontmatter>

- Sert à créer une page de titre ou une première de couverture
- Comprend les éléments <titlepage> (page de titre), <author> (nom du ou des auteurs), <edition> (mentions d'édition)

exemple :

```
<titlepage><titleproper>Inventaire des archives communales de Toulouse antérieures à
1790<subtitle>Répertoire numérique détaillé de la série AA. Numéros 1 à
60</subtitle></titleproper><author>par E. Roschach et FR. Bordes</author>
<publisher>Archives communales de Toulouse</publisher><date>2002</date> </titlepage>
```

## VOIR ANNEXE 1

### 2.3.5. *<archdesc>Description archivistique*

- Englobe toutes les informations décrivant un corpus documentaire dans son ensemble et recouvre les différents niveaux hiérarchiques de cette description
- Comprend des informations de haut niveau : identification du fonds (élément <did>), biographie/histoire institutionnelle (élément <bioghist>) ; puis on décrit les parties du fonds dans l'élément <dsc>Description des sous-composants, en commençant par le niveau hiérarchique le plus élevé et en allant si on le souhaite jusqu'au niveau de la pièce
- L'élément <archdesc> a plusieurs attributs spécifiques
- Par exemple, l'attribut LEVEL, obligatoire, permet de définir le niveau de l'ensemble documentaire : “collection” (collection), “fonds” (fonds), “subgroup” (sous-fonds), “series” (série organique), “subseries” (sous-série organique) ou “otherlevel” (autre niveau)
- Cet attribut est comparable à l'élément de données 3.1.4 d'ISAG (G) et au sous-champ 351\$*c* de MARC21.

#### *<archdesc>Description archivistique*

<archdesc>

<did> identification et description

<add> Données descriptives additionnelles

<adminfo> informations de gestion

<bioghist> biographie/histoire institutionnelle

<scopecontent> présentation du contenu

<controlaccess> point d'accès contrôlé

<dsc> description des composants

<c> niveaux de description définis grâce à un attribut

</archdesc>

### 2.3.5.1. <did> Identification et description de l'unité documentaire

```
<did>
  <repository>organisme responsable de l'accès intellectuel
  <origination>origine
  <unittitle>intitulé de l'UD
  <unitid>identifiant de l'UD
  <physdesc>description physique
  <abstract>résumé
  <physloc>localisation physique
</did>
```

#### Exemple d'utilisation de <did>

```
<archdesc level="fonds"
  <did>
    <unitid>320AP</unitid>
    <unittitle>Fonds Bazaine</unittitle>
    <unitdate normal="1801/1900">XIX<emph render="super">e</emph>siècle</unitdate>
    <origination>Bazaine, Achille</origination>
    <physdesc>7 cartons (320AP/1 à 320AP/7), 0,75 mètres linéaires</physdesc>
  </did>

  <!-- Etc. : les autres éléments de archdesc-->
```

### 2.3.5.2. *<add>Données descriptives additionnelles*

- *<p>* Paragraphe OU
- *<bibliography>* Bibliographie
- *<chronlist>* Liste chronologique
- *<fileplan>* Plan de classement
- *<index>* Index
- *<note<* Note
- *<otherfindaid>* Autre instrument de recherche
- *<relatedmaterial>* Documents en relation
- *<separatedmaterial>* Documents séparés

### 2.3.5.3. *<adminfo>Informations de gestion*

- *<accessrestrict>* Restrictions d'accès
- *<accruals>* Accroissements
- *<acqinfo>* Informations sur les modalités d'entrée
- *<altformavail>* Documents de substitution
- *<appraisal>* Informations sur l'évaluation
- *<custodhist>* Historique de la conservation
- *<note>* Note
- *<processinfo>* Informations sur le traitement
- *<userrestrict>* Restrictions d'utilisation
- OU *<p>* Paragraphe

#### 2.3.5.4. L'élément <controlaccess> Vedettes et Accès contrôlés

- Permet de saisir sous une forme normalisée des points d'accès pour les unités documentaires décrites, en citant le cas échéant le vocabulaire ou la notice utilisés
- Sous-éléments : <corpname> *Collectivité*, <famname> *Nom de famille*, <function> *Activité*, <genreform> *Genre et caractéristiques physiques*, <geogname> *Nom géographique*, <name> *Nom*, <occupation> *Fonction*, <persname> *Nom de personne*, <subject> *Mot-matière*, <title> *Titre*

#### Exemple d'utilisation de <controlaccess>

```
<controlaccess>
  <subject>XVIe siècle</subject>
  <subject>guerre de 1914-1918 </subject>
  <geogname>Bouville (Eure-et-Loir)</geogname>
  <geogname>Eure-et-Loir (Centre - département)</geogname>
  <famname>Argent de Deux-Fontaines (famille d')</famname>
</controlaccess>
Utilisation de <dsc>
```

### 2.3.5.5. <dsc>Description des sous-composants

- <dsc> ouvre la liste décomposant la description.
- Dans <dsc>, on saisit un élément <c> de premier niveau, ou <c01>, pour y mettre les informations décrivant un ensemble de documents d'archives formant le deuxième niveau de description archivistique (après celui de <archdesc>)
- Il est déconseillé d'utiliser les composants numérotés (<c01>, <c02>, etc) : il est difficile de déplacer et modifier les niveaux si on utilise ces composants numérotés, et on est contraint par le nombre de niveaux (12) alors que <c> peut être emboîté indéfiniment dans <c>

#### <dsc>

- On ouvre ensuite <did> dans lequel on emboîtera toujours : les éléments <unittitle>, <unitid>, <unitdate> et éventuellement <origination> si on connaît le nom du sous-producteur des documents, <physdesc>, <container>, <note>
- On pourra avoir des structures très simples du type :
 

```
<archdesc> fonds
  <dsc>
    <c> niveau article
    <c> niveau article
    <c> niveau article
  </dsc>
```

Ou des structures plus complexes du type :

```
<archdesc> fonds
  <dsc>
    <c> sous-fonds
      <c> série organique
        <c> article
    <c> sous-fonds
      <c> série organique
        <c> sous- série organique
          <c> article
```

## VOIR ANNEXE 2

### 2.3. L'EAD et la norme ISAD (G)

- Développée ultérieurement, au départ sans lien avec le groupe d'experts du CIA puis en collaboration avec lui, l'EAD est totalement compatible avec ISAD (G)
- L'EAD est fondée sur le même principe de description par niveaux
- A chacun des éléments définis dans ISAD (G) correspond un élément (éventuellement subdivisé en sous-éléments) ou un attribut dans l'EAD
- Mais l'EAD contient des éléments de description inexistant dans l'ISAD (G), comme les métadonnées de l'instrument de recherche

#### ISAD (G) repose sur 3 préalables

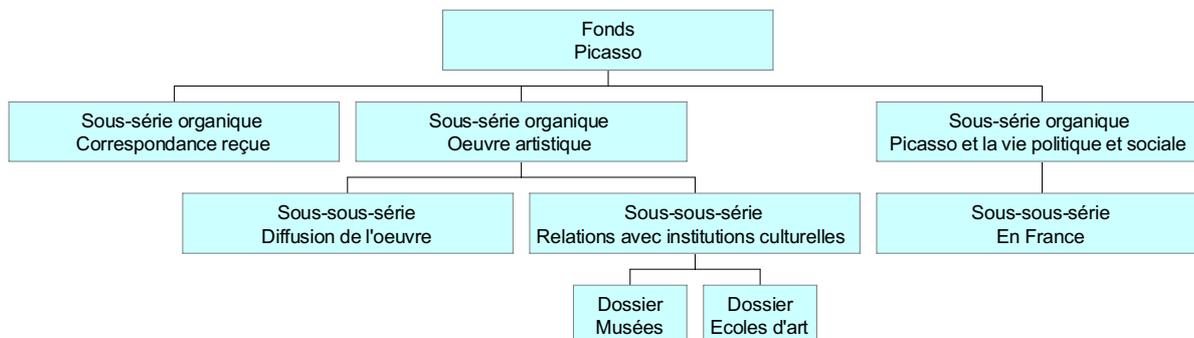
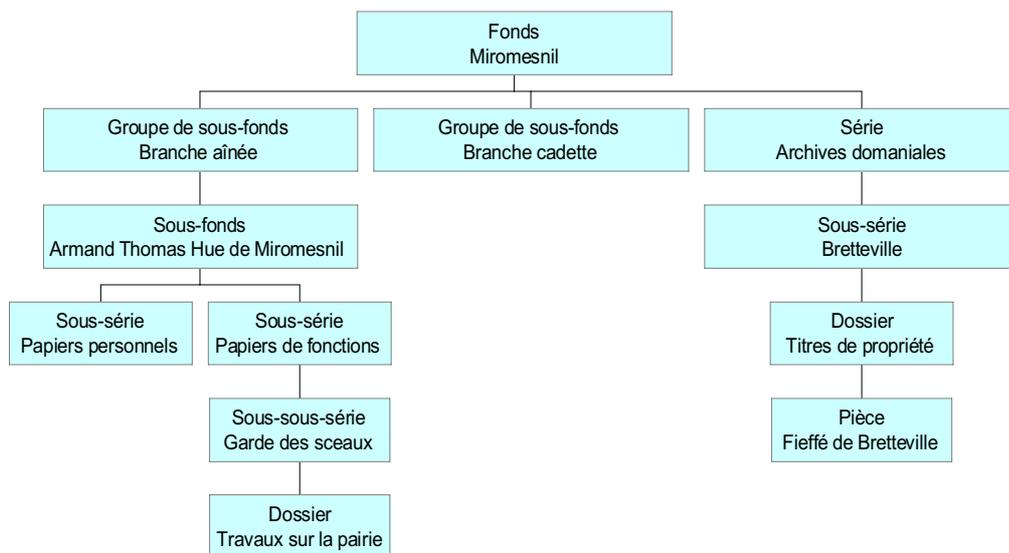
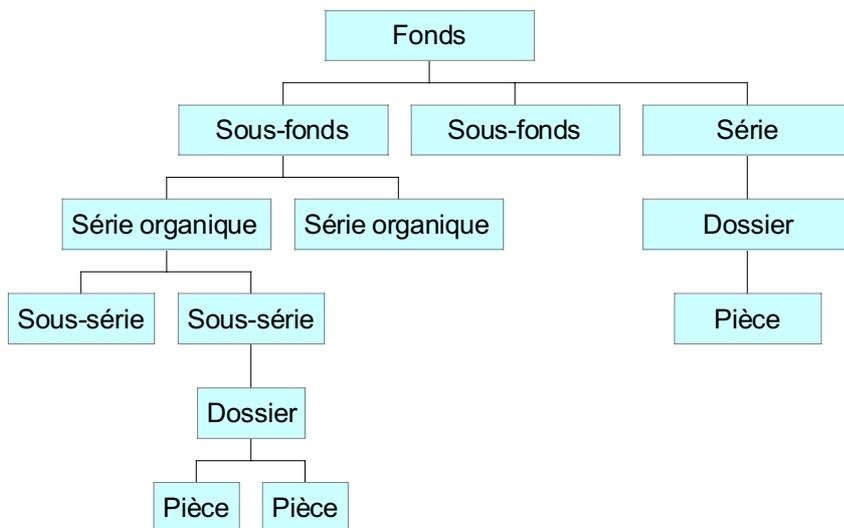
- validité du concept de fonds et du principe de respect des fonds
- correspondance entre niveaux de classement et niveaux de description
- nécessité de la description à plusieurs niveaux et non redondance des informations entre ces différents niveaux de description

#### Correspondance entre niveaux de classement/de description

- Un fonds se décompose en sous-fonds, série organique, sous-série organique, dossier et pièce
- Ces niveaux peuvent être eux-mêmes subdivisés en un nombre indéfini de niveaux intermédiaires suivant la complexité de la structure interne du fonds

### 2.4.1. Schéma des niveaux de classement d'un fonds

#### Exemples



### ***Nécessité de la description à plusieurs niveaux***

- regrouper les informations communes à plusieurs unités de description dans un ou plusieurs niveaux supérieurs
- établir des liens entre ces descriptions, afin de rapprocher toutes les informations intéressant une même unité

### ***Règles de la description à plusieurs niveaux***

- la description doit être effectuée du général au particulier
- les informations doivent être adaptées au niveau de description
- les descriptions doivent être liées entre elles
- une même information ne doit pas être répétée

### ***2.4.2. ISAD (G) : 26 éléments de description répartis en 7 zones***

- zone d'identification
- zone du contexte
- zone du contenu
- zone des conditions d'accès et d'utilisation
- zone des sources complémentaires
- zone des notes
- zone du contrôle de la description

#### ***Zone d'identification***

- 3.1.1. Référence  
EAD <unitid>
- 3.1.2. Intitulé/analyse  
EAD <unittitle>
- 3.1.3. Dates  
EAD <unitdate>
- 3.1.4. Niveau de description  
EAD attribut level dans <archdesc> et <c>
- 3.1.5. Importance matérielle et support de l'ud  
EAD <physdesc>

#### ***Zone du contexte***

- 3.2.1. Nom du producteur  
EAD <origination>
- 3.2.2. Histoire administrative/notice biographique  
EAD <bioghist>
- 3.2.3. Historique de la conservation  
EAD <custodhist>
- 3.2.4. Modalités d'entrée  
EAD <acqinfo>

***Zone du contenu et de la structure***

- 3.3.1. Présentation du contenu  
EAD <scopecontent>
- 3.3.2. Evaluation, tris et éliminations, sort final  
EAD <appraisal>
- 3.3.3. Accroissement  
EAD <accruals>
- 3.3.4. Mode de classement  
EAD <organization> et <arrangement>

***Zone des conditions d'accès et d'utilisation***

- 3.4.1. Conditions d'accès  
EAD <accessrestrict>
- 3.4.2. Conditions de reproduction  
EAD <userrestrict>
- 3.4.3. Langue et écriture des documents  
EAD attribut langmaterial dans <archdesc> et <c>
- 3.4.4. Caractéristiques matérielles et contraintes techniques
- 3.4.5. Instruments de recherche  
EAD <otherfindaid>

***Zone des sources complémentaires***

- 3.5.1. Existence et lieu de conservation des originaux  
EAD <odd>
- 3.5.2. Existence et lieu de conservation de copies  
EAD <altformavail>
- 3.5.3. Sources complémentaires  
EAD <relatedmaterial> et <separatedmaterial>
- 3.5.4. Bibliographie  
EAD <bibliography>

***Zone des notes***

- 3.6.1. Note  
EAD <odd>

***Zone du contrôle de la description***

- 3.7.1. Notes de l'archiviste  
EAD <processinfo>
- 3.7.2. Règles ou convention
- 3.7.3. Date(s) de la description  
EAD <processinfo>

### 3. Exemples d'utilisation de l'EAD

- Utilisée par les archives, les bibliothèques et les musées
  - pour la description,
  - pour la recherche dans des applications Web
- Adoptée par :
  - Library of Congress (nombreux inventaires en ligne, American Memory),
  - New York Public Library (128 inventaires en ligne),
  - Universités de Harvard, Berkeley, Stanford, Yale, Cornell, etc.,
  - Metropolitan Museum, Fondation Guggenheim,
  - Université de Toronto,
  - British Library,
  - Public Record Office,
  - Bibliothèque nationale d'Australie,
  - Musei Civici Veneziani,
  - Université de Göttingen,
  - Universités de Vienne, Salzburg, etc.

#### *Quelques exemples de sites donnant accès à des instruments de recherche encodés en EAD*

- Tous sites EAD du monde sur <http://jefferson.village.virginia.edu/ead/sitesann.html>
- Research Library Group sur <http://eureka.rlg.org>
- Library of Congress sur <http://lcweb2.loc.gov/faid/>
- Duke university sur <http://scriptorium.lib.duke.edu/findaids/> avec outil Dynaweb

#### *Quelques exemples de sites donnant accès à des instruments de recherche encodés en EAD*

- Cornell University sur <http://cidc.library.cornell.edu/xml/>
- Museums and the Online Archive of California sur <http://www.bampfa.berkeley.edu/moac/>
- PRO sur [http://catalogue.pro.gov.uk/ListInt/browse\\_lettercode\\_frameset.asp](http://catalogue.pro.gov.uk/ListInt/browse_lettercode_frameset.asp)

#### *Research Library Group (RLG)*

- Association à but non lucratif, fondée en 1974
- 164 adhérents dans 13 pays
- Bureaux à Mountain View (Californie)
- Buts : organiser la collaboration entre bibliothèques, archives et musées, améliorer l'accès à l'information pour l'enseignement et la recherche, rechercher des économies d'échelle dans toute opération d'organisation et de sauvegarde de la documentation culturelle mondiale, quelle que soit sa forme ou son support
- Plusieurs bases accessibles sur Internet, dont « Archival Resources » (510 000 descriptions de fonds ou collections et 21000 inventaires encodés)
- recherche en plein texte dans les documents
- liens avec les documents originaux
- l'utilisateur peut récupérer ses résultats en ligne ou les envoyer à une imprimante ou les télécharger sur son disque dur
- voie d'accès : Eureka (interface web pour professionnels et chercheurs)

### 3.1. L'EAD en France

- Réseau des Archives de France
- 2000 : la Mission de la recherche et de la technologie fait développer la DTD BiblioML à partir des formats Unimarc bibliographique et d'autorités,
  - à la BNF, conversion des données en EAD et structuration en XML du dernier volume paru des Nouvelles acquisitions françaises (NAF) et ajout des images et textes libres de droits disponibles sur Gallica (achèvement en avril 2001), DLL (catalogue des manuscrits)
  - études menées par l'INHA sur une éventuelle utilisation de EAD et XML
- Projet européen MALVINE (Manuscripts and Letters Via Integrated Networks in Europe) : <http://helmer.hit.uib.no/malvine/EAD/page.html>
- Napoleonica, encodage de l'inventaire des documents de travail du Conseil d'Etat
- fonds Cuvier (Institut et Museum) ...

#### *Dans le réseau des Archives de France*

- Test de conversion en EAD réalisé en 1999 à partir du répertoire numérique de la série M des AD des Pyrénées-Orientales
- Groupe de travail francophone
  - groupe AFNOR : traduction du dictionnaire des éléments (en ligne depuis octobre 2002) et manuel d'application
  - bulletin francophone
- Journées de présentation de l'EAD (2000, 2002)
- Recommandations de la DAF sur XML et EAD (juin 2002), BORA en ligne depuis octobre 2002
- Les sociétés auteurs des principaux logiciels documentaires utilisés en archivistique commencent à étudier et à mettre en place des fonctions d'export de données en XML (ex. EVER avec CLARA ; ARKHEIA)
- Divers projets pilotes de conversion sont en cours aux Archives nationales et dans des services territoriaux (Archives municipales de Toulouse) ; mais la saisie directe reste peu pratiquée
- La formation continue s'organise depuis peu

#### *BORA (=Base d'orientation et de recherche dans les archives)*

- La Direction des Archives de France met progressivement en œuvre une Base d'orientation et de recherche qui deviendra un portail.
- Tous les fonds d'archives conservés dans les services publics français y seront à terme décrits selon le standard international ISAD (G).
- Dans la version actuelle, BORA est composée de deux sous-éléments : archives privées et fonds photographiques
- Ces deux sous-éléments seront fusionnés dans la prochaine version de l'application

### 3.2. L'EAD au Centre historique des Archives nationales

- L'intérêt porté par le CHAN à l'EAD est né dès 2000 et s'inscrit dans un contexte général (le ministère de la Culture souhaite que toute sa documentation scientifique soit convertie ou produite en XML, le XML est le format d'archivage préconisé par l'Agence pour le développement de l'administration électronique et par la direction des Archives de France dans son manuel)
- Le Service des Nouvelles Technologies de l'Information et de la Communication (florence.clavaud@culture.gouv.fr) est chargé d'élaborer et de mettre en œuvre la politique informatique du CHAN

#### 3.2.1. L'EAD au Centre historique des Archives nationales : premier projet

- Encodage en EAD de deux inventaires avec un financement de la Gladys Kriebble Delmas Foundation américaine :  
*Etat des fonds d'archives privées* (317 à 630 AP) *Etat sommaire du fonds Napoléon* (400 AP)  
[ainsi que l'état des fonds d'archives d'entreprises du Centre des archives du monde du travail et l'inventaire du fonds de l'abbaye de Clairvaux aux Archives départementales de l'Aube]
- Encodage commandé à la société AJLSM (Martin Sévigny), qui a produit par programme la première version des fichiers XML et HTML
- L'Etat sommaire du fonds Napoléon en XML/EAD a été relu et corrigé par le SNTIC, une première version HTML diffusée en mars 2002, une 2e en octobre
- L'Etat des fonds AP (316 à 630 AP) en XML/EAD a été relu, corrigé et diffusé en HTML ; par la suite les notices des fonds 1 à 315 AP et des fonds d'associations et de presse seront intégrées à cet état sommaire.
- Ces travaux ont été riches d'enseignements :
  - pour l'état des fonds AP conforme à ISAD (G) : les seules difficultés de conversion sont liées à quelques règles d'irrégularité ou ambiguïtés de présentation qui pourront être évitées ; l'index a été entièrement récupéré
  - pour l'état sommaire du fonds 400 AP : la relecture comparée des versions HTML et XML a révélé des irrégularités très fréquentes de structure (l'inventaire n'utilise pas la description à plusieurs niveaux ; nombreuses notes de bas de page) ; toute conversion doit être précédée d'une étude très poussée de l'instrument de recherche à convertir, voire de sa restructuration préalable
- AJLSM a livré fin mai puis en août 2002 en toute propriété une version préliminaire des programmes de transformation XML vers HTML (programmes XSL-T) et des feuilles de style CSS
- Les programmes de conversion de Word vers XML écrits par AJLSM pourront être appliqués aux 315 premières notices de l'état des fonds privés
- Les notices 316 à 630 AP ont été livrées à la section des archives privées et ont été mises à jour directement en XML par elle, ce qui a facilité la gestion de ces fichiers (une seule opération de conversion, automatique, pour la diffusion)
- Les inventaires encodés sont diffusés sur le site web du CHAN et accessibles dans la base Archival Resources de RLG
- Les programmes XSL-T et feuilles de style CSS, modulaires et génériques, éventuellement modifiés, peuvent servir à traiter et présenter sur Internet tout instrument de recherche encodé en XML/EAD  
Exemple : inventaire du fonds de l'Amicale des anciens déportés de Mauthausen, répertoire

numérique du fonds Picasso

### 3.2.2. *L'EAD au Centre historique des Archives nationales : deuxième projet*

- Concerne l'*Etat des fonds des Archives nationales*, tomes II et V (1789-1958) et l'*Etat des inventaires* pour la même période : ces livres imprimés ont été numérisés en 2000 et sont en cours d'encodage par AJLSM, sur crédits de la MRT
- L'Etat des fonds n'est pas conforme à ISAD (G) ce qui oblige à en modifier ou préciser le contenu (gros travail de marquage des niveaux, identification des parties de fonds et sous-fonds dispersées au sein du cadre de classement, problème des introductions composites)
- Les notices bibliographiques de l'Etat des inventaires seront réintégrées dans l'Etat des fonds (élément <otherfindaid>) ; AJLSM livrera ces notices sous forme de fichier ASCII sommairement structurées, susceptibles d'être converties en Biblio-ML
- Si l'essentiel de la conversion est réalisé automatiquement par programme, alors ce programme pourra être utilisé pour convertir en XML les autres volumes de l'Etat des fonds

### 3.2.3. *L'EAD au Centre historique des Archives nationales : la rétroconversion*

- Le CHAN a produit plus de 2700 inventaires dont :
  - seulement 20 % existent sous forme imprimée
  - beaucoup sont anciens et peu fiables, périmés, visuellement peu structurés
  - la présentation est très variée, ce qui nuit à leur compréhension
- Des chevauchements et imbrications rendent ce corpus difficile à manipuler

#### *Les principes*

- Ne convertir en XML que les instruments de recherche
  - qui en valent la peine (suffisamment valides et utiles pour le public)
  - qui peuvent l'être (suffisamment structurés ou compatibles avec l'esprit de la norme, ce qui n'est pas le cas des inventaires-index)
- Pour les autres inventaires, on fera une simple OCR (ou une ressaisie intégrale en interne) et éventuellement une conversion en PDF

#### *Les bases de données*

- Les bases peuvent être classifiées en différentes familles :
  - les bases bibliographiques, dont le contenu peut être converti en XML conformément à Biblio-ML
  - les bases archivistiques, dont le contenu peut être présenté au public sous forme d'arbre, donc converti en XML selon l'EAD
  - les bases documentaires dont le contenu se prête mal à une représentation hiérarchique
  - les bases fichiers d'autorités

#### *Les bases bibliographiques*

- Le catalogue de la Bibliothèque historique du CHAN servira de point de départ (structuré conformément à UNIMARC). Le Centre des archives d'outre-mer participe à un projet d'application bibliographique basé sur Biblio-ML et accessible sous SDX
- Une autre base bibliographique pourra être mise en place pour décrire les inventaires du CHAN. Le SNTIC réfléchit à sa structure avec la Bibliothèque historique

#### *Les bases documentaires*

- AJLSM vient de livrer un outil de conversion des notices au format ASCII en documents XML bien formés
- C'est ce programme, entièrement adapté, qui sert pour produire actuellement les notices d'autorité des notaires parisiens en XML, à partir de la base ETANOT du Minutier central sous CINDOC
- Ces bases devront avant être restructurées pour en rapprocher les grilles et le vocabulaire

#### *Les bases archivistiques*

- Il est important de rapprocher grilles, syntaxe et sémantique des diverses bases et de les rendre conformes à l'ISAD (G) et à l'EAD, afin de faciliter leur consultation.
- Le SNTIC a élaboré en 2001-2002 un modèle de base de données archivistiques qui permet de décrire n'importe quel fonds, série ou collection à l'aide d'une grille linéaire et du logiciel CINDOC
- Ce modèle est en cours d'application pour une partie des données de la base ARCHIM
- AJLSM vient de livrer au CHAN la première version d'un programme de conversion de données ASCII conformes à ce modèle, en documents XML conformes à l'EAD.

### 3.3. Produire (et mettre à jour) les inventaires directement en XML

- Le CHAN dispose en tout de 9 licences du logiciel XMetaL v2.1 et v. 3, éditeur XML adapté à la production d'instruments de recherche en long
- Ce logiciel a été doté d'une configuration spécifique pour les archivistes, conçue par AJLSM.
- Installé dans différents services (Archives privées, sections ancienne, du XIXe siècle, du XXe siècle,...), ce logiciel a commencé à servir pour les premières expériences de production directe en XML

#### *Quelques mots sur XMetaL*

- Permet la production de documents textuels en XML
- Propose trois représentations de chaque document ainsi que la visualisation de leur arbre, des listes d'éléments et attributs utilisables
- Contrôle la validité d'un document par rapport à la DTD déclarée (ici l'EAD)
- L'intitulé en clair et en français de chaque élément est affiché en bas à droite de l'écran
- Les modes d'affichage en arbre, avec et sans balises ont été configurés de façon spécifique et modifiable (couleurs, indentations, comptage des niveaux)
- Un modèle a été développé pour saisir une notice de l'état des fonds d'archives privées

#### *PROXIMUS*

- XMetaL n'est ni un outil réseau ni une plate-forme capable de gérer des quantités importantes de documents XML
- Le ministère de la Culture a lancé un appel d'offres pour la réalisation d'une plate-forme d'outils logiciels libres, pour produire, gérer et archiver en réseau la documentation scientifique de ses services. La plate-forme s'appelle PROXIMUS. Elle devra être générique, paramétrable et évolutive.
- PROXIMUS sera déployée fin 2003 au CHAN et remplacera alors XMetaL (et progressivement Word)

### 3.4. Diffuser les documents EAD sur Internet

- Les programmes de transformation des documents XML en HTML sont indispensables à la chaîne de production et de diffusion des instruments de recherche en XML.
- Ces programmes fondés sur le langage XSL-T peuvent soit servir à fabriquer des versions HTML statiques des instruments de recherche, soit être intégrés à des systèmes de publication dynamique.
- Dans l'immédiat, le CHAN s'est procuré grâce à AJLSM des feuilles de style XSL-T pour fabriquer des versions HTML statiques (état sommaire du fonds Napoléon, état des fonds AP, répertoire du fonds Picasso)
- On pourra bientôt charger sur le serveur web des documents en XML natif, en donnant à l'internaute la possibilité d'interroger un document XML en texte intégral mais aussi de consulter des index de certains éléments ou des listes prédéfinies de termes.
- Le ministère de la Culture a fait développer par AJLSM, pour sa documentation scientifique, une plate-forme d'outils libres SDX
- Exemples d'application SDX : base BORA, guide de l'Internet culturel, calendrier du site Victor Hugo 2002

## 4. Pour en savoir plus...

### *Ressources utiles et bibliographie*

- Le site officiel de l'EAD : <http://www.loc.gov/ead/>
- La table ronde de la SAA sur l'EAD (et son « cookbook ») : <http://jefferson.village.virginia.edu/ead/>
- Sur le site de la DAF, une rubrique du chapitre Archivistique est consacrée à l'EAD : <http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/index.html>
- Pour tout renseignement complémentaire : [catherine.dherent@culture.gouv.fr](mailto:catherine.dherent@culture.gouv.fr)
- Nombreux articles et revues en anglais dont *Encoded archival description : context theory and case studies*, SAA, août 1998, et revue *The American archivist*
- **Sur le XML :**
  - deux sites en français :
    - <http://www.educnet.education.fr/dossier/xml/>
    - <http://xmlfr.org>
  - deux ouvrages :
    - MICHARD, Alain, *XML : langage et applications*, Editions Eyrolles, Paris, 2001 ;
    - RAY, Erik T., *Introduction à XML*, O'Reilly, Paris, 2001
- Le support de cours rédigé par Martin Sévigny sur XML  
<http://www.ajlsm.com/formation/xml>