



L'EAD et l'EAC, deux DTD pour la description archivistique

Claire Sibille
Archives de France
Présentation Groupe PIN, 4 mai 2004

L'EAD et l'EAC, deux DTD pour la description archivistique

1. Présentation sommaire de l'EAD =
Description archivistique encodée
(version 2002)
2. Exemples d'utilisation de l'EAD
3. Les outils
4. La DTD EAC = Contexte archivistique
encodé
5. Pour en savoir plus...

Quelques définitions

Tout document ou ensemble documentaire :

- se situe dans un **contexte** de production :
une pièce fait partie d'un dossier, un dossier d'un **fonds** ou d'un sous-fonds (« unités documentaires »)

ensemble des documents d'archives de toute nature réunis par une personne physique ou morale dans l'exercice de ses activités ou de ses fonctions

La description archivistique :

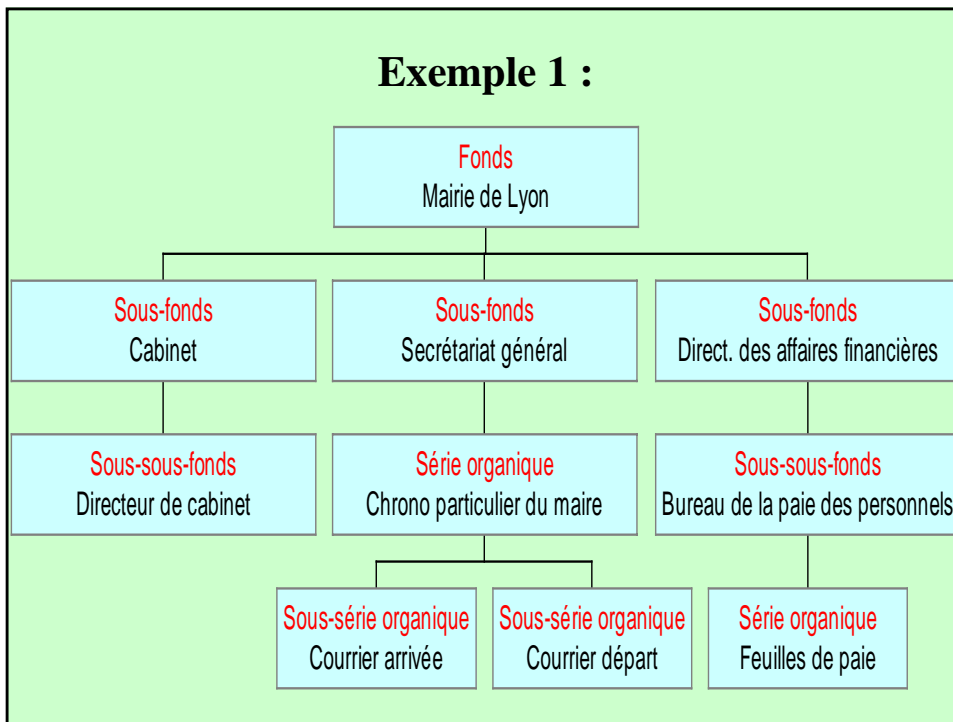
- ensemble des tâches accomplies par l'archiviste en vue d'élaborer des **instruments de recherche** qui faciliteront l'accès au fonds d'archives en général et aux documents en particulier
- s'appuie sur la norme de description archivistique générale et internationale ISAD (G)

outil papier ou informatisé énumérant ou décrivant un ensemble de documents d'archives de manière à les faire connaître aux utilisateurs

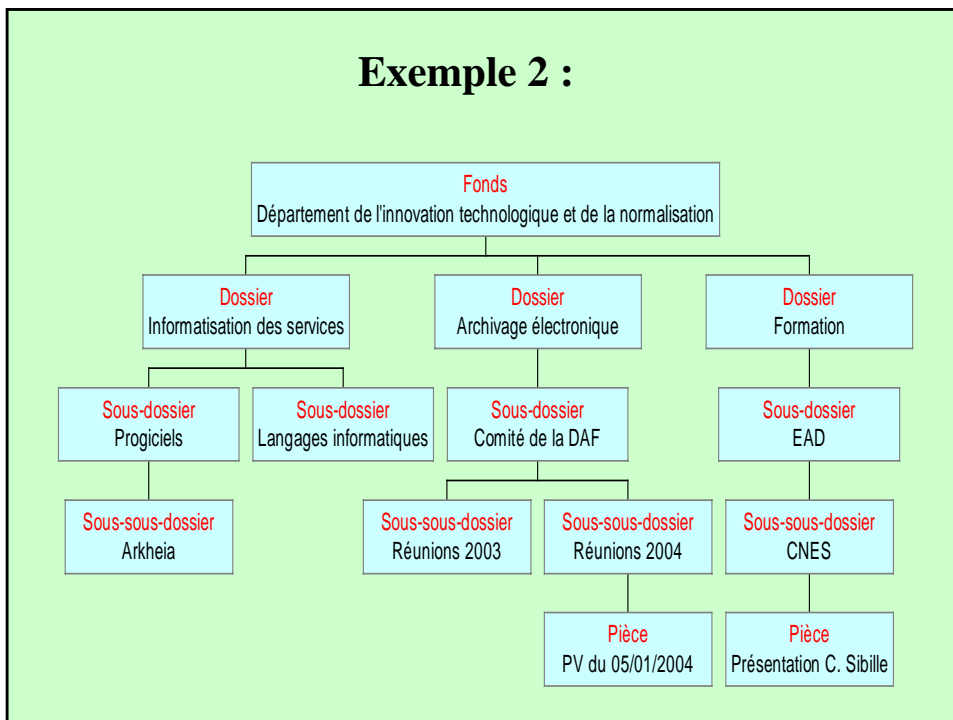
Niveaux de classement et niveaux de description

- Le classement d'un fonds d'archives est préalable à sa description, car il en détermine la structure organique
- Un fonds se décompose en sous-fonds, série organique, sous-série organique, dossier et pièce
- Ces niveaux de classement constitueront par la suite autant de niveaux de description

Exemple 1 :



Exemple 2 :



La description à plusieurs niveaux

- On peut établir une description unique pour l'ensemble du fonds, ou établir une description particulière pour chaque subdivision, en descendant au niveau de la pièce, du dossier ou de la série organique selon le cas
- La description à plusieurs niveaux consiste à fournir des éléments de description pour l'ensemble du fonds, d'autres propres à une série, à chaque dossier, voire à chaque pièce

Quatre règles fondamentales

- présenter le contexte et la structure hiérarchique du fonds et de ses composantes (aller du général au particulier)
- donner les renseignements appropriés au niveau décrit
- relier chaque description à celle du niveau de description immédiatement supérieur
- ne pas répéter à un niveau inférieur l'information déjà présente au niveau supérieur

**A cela s'ajoute pour
l'archiviste la nécessité :**

- de pérennisation des instruments de recherche qui maintiennent l'accès aux données (migration)
- de création partagée des descriptions
- d'aide plus efficace au public (besoin d'un repérage immédiat de l'information)

**L'EAD 2002 : principes,
objectifs, structure**

Les origines de l'EAD

- Née dans le monde des bibliothèques (université de Berkeley, 1993) pour remplacer le standard MARC-AMC (Archival and Manuscript Control), insuffisant pour les descriptions à plusieurs niveaux
- Propriété de la Society of American Archivists
- Maintenance par la Library of Congress
- 1996 : version expérimentale
- août 1998 : première version officielle (1.0)
- septembre 2002 : révision (EAD 2002)

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 ») a traduit en français la Tag Library (version 1.0) ; la version officielle de cette traduction est en ligne sur le site de la DAF

www.archivesdefrance.culture.gouv.fr/fr/archivistique/

- Le groupe AFNOR/EAD traduit la Tag Library (version 2002) et prépare un manuel d'application de l'EAD

Caractéristiques essentielles (1)

- D'abord écrite sur SGML, l'EAD est parfaitement compatible avec XML
- S'adapte à tous types d'archives, quels qu'en soient le supports, la forme et le volume, ainsi qu'aux différents types d'instruments de recherche (sauf aux guides et aux bordereaux de versement)
- Pas de limite de taille pour les instruments de recherche encodés

Caractéristiques essentielles (2)

- 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

Caractéristiques essentielles (3)

- L'EAD utilise des standards complémentaires (on peut faire des liens depuis une notice MARC décrivant un fonds vers l'inventaire détaillé encodé en EAD ; inversement, on peut générer l'ossature d'une notice MARC à partir des éléments de haut niveau d'un document EAD)
- L'EAD permet de créer de nouveaux instruments de recherche, mais elle est aussi très bien adaptée à des opérations de conversion rétrospective, notamment par la souplesse de l'ordre de ses éléments

Caractéristiques essentielles (4)

- 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

Structure

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

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

<eadheader>En-tête EAD

- Il fournit des informations sur l'instrument de recherche lui-même : titre, rédacteur, date de rédaction, informations sur le fichier...
- Il peut être sommaire ou contenir de nombreux sous-éléments spécifiques
- Il contient les métadonnées
- Il est directement inspiré par la TEI (Text Encoding Initiative)

<eadheader>En-tête EAD

<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**

<langusage>**Langue utilisée**

<descrules>**Règles de description**

<revisiondesc>**Descriptions des révisions**

</eadheader>

The screenshot shows a software interface for editing XML documents. On the left, a tree view displays the document's structure, including elements like <eadheader>, <eadid>, <filedesc>, <titlestmt>, <editionstmt>, <publicationstmt>, <seriestmt>, <notestmt>, <profiledesc>, <creation>, <langusage>, <descrules>, and <revisiondesc>. The main window shows the content of the selected element, which is <titlestmt>. It contains <titleproper> 'Répertoire numérique de la sous-série 10 M' and <author> 'par Germaine GALL, sous la direction de Dominique Dupont'. Below this, there is a <profiledesc> element with <creation> 'Cet instrument de recherche a été encodé en XML à l'aide du logiciel XMetaL, conformément à l'EAD (version 2002), par Marie-Thérèse Thompson, lors de son stage à la DAF dans le cadre de sa formation ENSIB' and <langusage> 'Instrument de recherche rédigé en français'. The interface also shows a toolbar at the top and a status bar at the bottom.

<archdesc> Description archivistique

- Englobe toutes les informations décrivant un ensemble documentaire 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>), etc. ; 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

<archdesc> Description archivistique

- L'élément <archdesc> a plusieurs attributs spécifiques
- Par exemple, l'attribut LEVEL, obligatoire, permet de définir le niveau de description de l'ensemble documentaire : "collection" (collection), "fonds" (fonds), "subfonds" (sous-fonds), "series" (série organique), "subseries" (sous-série organique) ou "otherlevel" (autre niveau)

<archdesc> Description archivistique

<archdesc>

<did> Identification et description

<bioghist> Biographie/histoire institutionnelle

<custodhist> Historique de la conservation

<acqinfo> Informations sur les modalités d'entrée

<accessrestrict> Conditions d'accès

<userrestrict> Conditions d'utilisation

<arrangement> Classement

<bibliography> Bibliographie

<scopecontent> Présentation du contenu

<controlaccess> Point d'accès contrôlé

<dsc> Description des composants

<c> niveaux de description définis grâce à un attribut

</archdesc>

<did> Identification et description de l'unité documentaire

<did>

<origination> Origine

<unittitle> Intitulé de l'unité documentaire

<unitdate> Date de l'unité documentaire

<physdesc> Description physique

<repository> Organisme responsable de l'accès intellectuel

<unitid> Identifiant de l'unité documentaire

<langmaterial> Langue de l'unité documentaire

<physloc> Localisation physique

</did>

Exemple d'utilisation de <did>

```
<archdesc level= "series"
  <did>
    <origination>Préfecture de la Somme</origination>
    <unittitle>Personnel de la préfecture, sous-préfectures
    et services annexes</unittitle>
    <unidade>an VIII-1950</unidade>
    <physdesc>82 cartons et registres, 9 mètres
    linéaires</physdesc>
    <repository>Archives départementales de la
    Somme</repository>
    <unitid>2 M 1-82</unitid>
  </did>
  <!-- Etc. : les autres éléments de archdesc-->
```

Informations complémentaires

- <accessrestrict><p> *Restrictions d'accès*
- <acqinfo><p> *Informations sur les modalités d'entrée*
- <altformavail><p> *Documents de substitution*
- <appraisal><p> *Informations sur l'évaluation*
- <custodhist><p> *Historique de la conservation*
- <userrestrict><p> *Restrictions d'utilisation*
- <accruals><p> *Accroissements*
- <bibliography><bibref> *Bibliographie*
- <relatedmaterial><p> *Documents en relation*
- <separatedmaterial><p> *Documents séparés*
- <otherfindaid><p> *Autre instrument de recherche*
- ou <descgrp> *Groupe d'éléments de description*

<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*

Exemples d'utilisation de <controlaccess>

```
<controlaccess>  
  <corpname>Préfecture des Vosges</corpname>  
  <subject>guerre de 1939-1945</subject>  
  <geogname>Bourges (Cher)</geogname>  
  <geogname>Indre (département)</geogname>  
  <famname>Dampierre (famille de)</famname>  
</controlaccess>
```

<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 les niveaux si on utilise ces composants numérotés, et on est contraint par le nombre de niveaux (12)

<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> niveau fonds
 <dsc>
 <c> niveau dossier
 <c> niveau dossier
 <c> niveau dossier

<dsc>

Ou des structures plus complexes du type :

<archdesc> fonds

<dsc>

<c> sous-fonds

<c> série organique

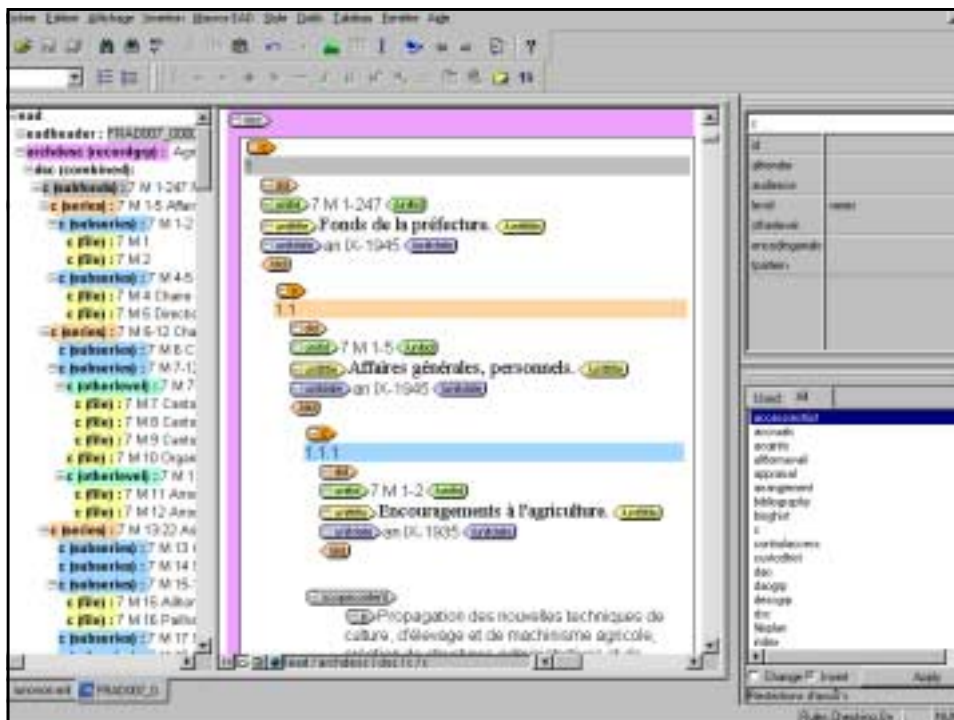
<c> dossier

<c> sous-fonds

<c> série organique

<c> sous-série organique

<c> dossier



Exemples d'utilisation de l'EAD

Exemples étrangers d'utilisation de l'EAD

- Tous sites EAD du monde sur
<http://jefferson.village.virginia.edu/ead/sitesann.html>
- PRO sur http://catalogue.pro.gov.uk/ListInt/browse_lettercode_frameset.asp
- Duke university sur
<http://scriptorium.lib.duke.edu/findaids/>
avec outil Dynaweb
- Research Library Group sur
<http://eureka.rlg.org>

L'EAD en France

- Direction du livre et de la lecture : rétroconversion du catalogue général des manuscrits
- Bibliothèque nationale de France : conversion des données en EAD et structuration en XML du dernier volume paru des *Nouvelles acquisitions françaises* (NAF)
- Institut national d'histoire de l'art : base de données sur les archives d'artistes et de galeries
- Projet européen MALVINE : conversion en EAD de catalogues réalisés sous 12 formats locaux divers

<http://helmer.hit.uib.no/malvine/EAD/page.html>

Dans le réseau des Archives de France

- Test de conversion en EAD réalisé en 1999 par Jouve à partir du répertoire numérique de la série M des AD des Pyrénées-Orientales
- Encodage, en 2001-2002, de quatre inventaires par la société AJLSM : état des fonds d'archives privées et inventaire du fonds Napoléon du CHAN, état des fonds d'archives d'entreprises du CAMT, inventaire du fonds de l'abbaye de Clairvaux aux AD de l'Aube
- Groupe de travail 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

Dans le réseau des Archives de France

- Les sociétés auteurs des principaux logiciels utilisés dans les services territoriaux commencent à mettre en place des fonctions d'export de données en XML (ex. : ANAPHORE avec ARKHEIA, SICEM avec THOT)
- Divers projets pilotes de conversion sont en cours aux Archives nationales et dans des services territoriaux (AD des Pyrénées-Atlantiques, du Puy-de-Dôme, des Yvelines, AM de Toulouse,...) ; mais la saisie directe reste peu pratiquée
- La formation continue se développe (DAF, CHAN, CAOM, INP....)

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

L'EAD à l'épreuve des fonds ouverts

- idée (née en 2000) de faire élaborer une chaîne de traitement complète, de la rédaction du bordereau de versement par le producteur à la réalisation du répertoire pour un fonds; crédits Recherche pour 2002
- tests faits avec le ministère de la Justice et les AD de Lozère (TGI de Mende)
- prise en compte des documents électroniques, dont certains peuvent contenir des métadonnées alimentant automatiquement la description au sein du bordereau

Le cahier des charges portait sur le fait :

- qu'il fallait mettre la DTD EAD à l'épreuve des fonds ouverts
 - en créant une plate-forme de production des bordereaux
 - en faisant évoluer l'instance EAD répertoire au fur et à mesure des versements
- qu'il fallait transformer la DTD EAD en schéma XML
- qu'il fallait intégrer des documents numériques à l'instance EAD

La société à qui a été sous-traité ce travail a réalisé :

- une plate-forme de production de bordereaux
- un modèle de bordereau de versement qui a fait l'objet d'un traitement en XML
- un formulaire de saisie en HTML d'une partie de ce bordereau pour le service versant
- Reste à faire le lien vers la création automatique d'un répertoire en EAD à partir de plusieurs bordereaux

Les outils de mise en œuvre et de diffusion

Les outils de mise en œuvre

- saisie directe avec un éditeur de documents XML (XMetaL, XMLSpy, XMLOperator...)
- saisie à partir d'un simple éditeur de texte (Notepad, Wordpad) ou d'un traitement de texte (Word, Wordperfect 9...) puis conversion vers XML
- export automatique de données à partir d'un outil de production d'instruments de recherche archivistiques (Arkhéïa, Clara...)

Quelques mots sur XMetaL

- Permet la production de documents textuels en XML
- Interfaces graphiques d'un traitement de texte : ouverture, fermeture et sauvegarde, travail avec plusieurs documents, impression copier/coller, recherche de caractères
- Tâches d'édition de documents XML : insertion d'un élément, ajout d'un attribut à un élément, création et insertion d'une entité, validation du document
- Permet de visualiser un document sous 3 formes, ainsi que sa structure arborescente, les listes d'éléments et d'attributs utilisables
- Contrôle la validité par rapport à une DTD déclarée (ici l'EAD)
- Les modes d'affichage ont été configurés par AJLSM

Les outils de diffusion

- Les programmes de transformation des documents XML en HTML, fondés sur le langage XSL-T, sont indispensables à la chaîne de production et de diffusion des instruments de recherche en XML
- Ces programmes peuvent soit servir à déposer sur un serveur de pages Web des versions HTML statiques des instruments de recherche, soit être intégrés à des systèmes de publication dynamique

Les outils de diffusion

- Des applications Web permettant d'interroger de façon ciblée le contenu d'un ensemble de documents XML ont été développées récemment :
 - ActionArchives de la société Anticipation (AD Deux-Sèvres, Côte-d'Or)
 - le ministère de la Culture a fait développer par AJLSM, pour sa documentation scientifique, une plate-forme d'outils libres SDX : <http://sdx.culture.fr>
Exemples d'applications SDX : base BORA, guide de l'Internet culturel, calendrier du site Victor Hugo 2002
 - les sociétés AJLSM et Anaphore ont développé deux applications génériques sous SDX : Pleade et Navimages

Les outils de diffusion

- Pleade est une application générique pour la recherche et la consultation en ligne d'instruments de recherche en XML/EAD. Le site Web de ce projet est consultable à : <http://www.pleade.org>
- Navimages est une application générique pour la gestion, la recherche et la consultation en ligne de lots d'images numériques, développée par AJLSM dans le cadre d'un marché public passé par la direction des Archives de France. Le site Web de ce projet est consultable à : <http://sdx.archivesdefrance.culture.gouv.fr/gpl/navimages>
- Ces deux outils, conçus pour être utilisables conjointement, servent à construire des applications documentaires à partir de ressources archivistiques. Ils fonctionnent sur la plateforme de logiciels libres SDX

Les outils de diffusion

- Le cœur de Pleade est constitué par la publication des documents EAD, document par document ou groupe de documents par groupe de documents
- Le processus de publication peut inclure des regroupements par types d'instruments de recherche ou de documents décrits, ainsi qu'une fragmentation des différents documents EAD en unités documentaires interrogeables
- Les documents EAD sont stockés sur un serveur Web au format natif XML et présentés après requête sous la forme de documents HTML générés à la volée

La DTD EAC (Encoded Archival Context)

La DTD EAC

- Destinée à encoder des notices d'autorité
- Premiers travaux en mars 2001
- Octobre 2003 : version Bêta de la Tag Library
- EAC compatible avec ISAAR (CPF)
- Comptes rendus du Working Group de l'EAC disponibles sur :
<http://jefferson.village.virginia.edu/eac/>

La DTD EAC

- Contient deux éléments obligatoires, l'élément En-tête EAC <eachheader> et l'élément Description du contexte <condesc>
- <eachheader> contient des données utilisées lors du contrôle de la description du producteur et pour fournir le contexte de la description
- <condesc> englobe la description du producteur

Sous-éléments de <eachheader>

- <eacid> Identifiant EAC (obligatoire)
- <mainhist> Historique des mises à jour (obligatoire)
- <language decl> Déclaration de langue (facultatif)
- <rule decl> Déclaration de règles (facultatif)
- <source decl> Déclaration de source (facultatif)
- <auth decl> Déclaration d'autorité (facultatif)

Attributs de <eachheader>

- L'attribut TYPE est utilisé pour signaler le type de producteur décrit dans l'instance EAC : collectivité, famille ou particulier
- L'attribut STATUS est utilisé pour signifier le statut de la description : brouillon, publié ou détruit
- L'attribut ENCODINGANALOGSYS est utilisé pour signaler le système dans lequel se trouvent des zones analogues à celles utilisées dans l'instance EAC

Attributs de <eachheader>

- LANGENCODING pour le codage de la langue de la description. Par défaut à "iso639-2b".
- SCRIPTENCODING pour le codage d'écriture utilisé. Par défaut à "iso15924".
- DATEENCODING pour la formulation de dates normalisées. Par défaut à "iso8601".
- COUNTRYCODE pour le codage des pays. Par défaut à "iso3166-1".
- OWNERENCODING pour le codage des services responsables de l'accès intellectuel ou des propriétaires. Par défaut à "iso11551".

<condesc> Description du contexte

- <identity> Identité (obligatoire)
- <desc> Description (facultatif)
- <eacrels> Relations avec une autre description (facultatif)
- <resourcerels> Relations avec une source d'information (facultatif)
- <funactrels> Relations de la fonction ou de l'activité (facultatif)

<identity> Identité

- Élément le plus complexe de l'EAC
- En plus d'associer un ou plusieurs noms utilisés par ou pour l'entité, l'élément <identity> doit aussi permettre d'associer deux ou plus de deux formes parallèles du nom dans différentes langues ou écritures
- Dans les pays où il y a plus d'une langue officielle, les noms de personnes morales sont souvent fournies dans plus d'une langue

Pour en savoir plus...

Ressources utiles et bibliographie

- Les sites officiels de l'EAD et de l'EAC :
<http://www.loc.gov/ead/>
<http://jefferson.village.virginia.edu/eac>
- La table ronde de la SAA (et son « cookbook » sur l'EAD) : <http://jefferson.village.virginia.edu/ead/>
- Sur le site de la DAF, deux rubriques du chapitre Archivistique sont consacrées à l'EAD et à l'EAC:
[http://www.archivesdefrance.culture.gouv.fr/fr/archivistique /index.html](http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/index.html)
- Pour tout renseignement complémentaire :
claire.sibille@culture.gouv.fr

Ressources utiles et bibliographie

- *Encoded archival description : context theory and case studies*, SAA, août 1998
- Revues *The American archivist*, *Journal of Internet Cataloging*, *Journal of Archival Organization*
- *Encoded Archival Description on the Internet*, par Daniel V. Pitti et Wendy Duff, 2002
- *Abrégé d'archivistique* (Association des archivistes français), chapitre sur les nouvelles technologies appliquées aux instruments de recherche (Florence Clavaud et Martin Sévigny), 2004