

Module 6 - section 8 :

Publication électronique des instruments de recherche

CLAIRE SIBILLE – DE GRIMOÜARD
MARCEL CAYA



version 1

14 novembre é011

Table des matières

1. Introduction	7
1.1. Des principes identiques.....	7
1.2. Des modalités différentes.....	8
2. Quelle(s) solutions technologique(s) pour quel(s) usage(s) ?	9
2.1. Les instruments de recherche électroniques non structurés.....	9
2.2. Les instruments de recherche électroniques structurés.....	9
3. Outils électroniques traditionnels	11
3.1. Instruments de recherche électroniques non structurés.....	11
3.2. Bases de données.....	12
4. Langages de balisage	15
4.1. Substrat technique.....	15
4.2. La DTD EAD.....	16
4.2.1. Structure de la DTD EAD.....	16
4.2.2. Caractéristiques principales.....	16
4.2.3. Éléments de l'EAD.....	18
4.2.4. ISAD(G) et EAD.....	18
4.2.5. Exercice.....	21
4.2.6. L'EAD et l'indexation à niveaux.....	21
4.3. La DTD EAC.....	21
4.4. La mise en œuvre de l'EAD et de l'EAC.....	23
4.4.1. Stratégies de production : comment faire de l'EAD ?.....	24
4.4.2. Les outils de production.....	24
4.4.3. Stratégies de publication et de recherche : comment utiliser de l'EAD ?.....	25
4.4.4. Les outils de gestion et de publication.....	25
4.4.5. Un investissement non négligeable mais qui en vaut la peine.....	25
5. Évaluation des connaissances	27
5.1. Exercice 1.....	27
5.2. Exercice 2.....	27

Objectifs

Description du module :

Nous avons tous vu des archives définitives avant même de savoir ce qu'elles sont. L'image de l'album de photographies anciennes ou de vieux contrats d'hypothèque aux pages vieilles et fragiles vient immédiatement à l'esprit. Parce que nous avons tous visité un jour des bureaux de l'administration, nous sommes aussi familiers avec les grandes quantités de documents que les administrations modernes produisent et accumulent. Pourtant, au sein du grand public, peu connaissent la distinction entre les archives courantes et les archives définitives.

Les archivistes d'aujourd'hui travaillent à mieux faire connaître du grand public l'importance de bien gérer les archives courantes, ces archives d'aujourd'hui, pour mieux protéger éventuellement celles dont on aura besoin demain. Une fois désignées comme définitives, ces archives constituent souvent la seule trace mémorielle des faits, qu'ils soient simples reconnaissances de droits (ou de devoirs) ou le souvenir d'événements et de contextes particuliers.

Parce qu'elles ne représentent qu'une fraction de l'ensemble des documents qui ont existé, ces archives définitives deviennent le lien privilégié entre ce qui était et la représentation que nous voulons conserver des personnes, des organismes et des administrations. C'est pourquoi, afin d'en préserver l'authenticité, la signification et l'intégrité, elles doivent être traitées et conservées selon un ensemble de règles et de normes par des spécialistes qui savent comment assurer leur préservation à long terme tout en permettant leur repérage et leur utilisation dans les meilleures conditions possibles.

Le but du module est de :

- vous apprendre à reconnaître les archives définitives et l'ensemble des problèmes et des solutions liées aux traitements nécessaires à leur utilisation
- vous initier aux procédures d'évaluation, de collecte et de tri
- vous faire connaître les principaux instruments de recherche dans les archives
- vous familiariser, dans le respect des normes internationales, avec les pratiques de classement et de description des archives définitives
- vous initier plus particulièrement aux normes internationales de description (ISAD-G) et aux notices d'autorité (ISAAR) en archivistique pour les noms propres de personnes, collectivités, lieux et famille
- vous initier à l'indexation des descriptions des archives définitives
- vous faire connaître les normes de préparation d'instruments de recherche électronique en vue d'une diffusion sur Internet

L'apprenant doit être en mesure de :

- comprendre les particularités du traitement des archives définitives
- identifier les problèmes particuliers posés par la collecte et le tri des archives définitives
- classer et décrire un fonds d'archives (et ses parties) selon les normes internationales
- préparer des instruments de recherche sur papier
- comprendre, en attendant de pouvoir le réaliser, les exigences de leur diffusion sur formats électroniques

Positionnement :

L'étudiant devrait avoir déjà vu les cinq modules précédents, particulièrement les modules 4 et 5 sur la gestion des archives courantes et intermédiaires.

Conseils d'apprentissage :

Ce module est assez long. Nous conseillons de lire d'abord les pages principales, puis de revenir sur les encarts si l'on veut davantage d'informations.

La section 8 sur la publication électronique des instruments de recherche s'adresse plus particulièrement à des archivistes qui ont déjà acquis de l'expérience ; elle est cependant maintenue dans cette formation qui se veut de base afin d'avoir une information d'ensemble sur les archives définitives.

1. Introduction

L'informatique est désormais utilisée de plus en plus pour le traitement des archives et, plus particulièrement, pour toutes les étapes de leur description et de la préparation des instruments de recherche. Dans ce contexte, des instruments de recherche ont d'abord été produits et diffusés sur des modes électroniques qui ressemblaient en tout point aux instruments sur papier ; à cet égard, ce sont surtout des logiciels de traitement de texte qui ont été utilisés.

Graduellement, cependant, on a vu un nombre croissant de centres recourir à des logiciels différents qui présupposent des modes d'organisation de l'information différents faisant appel plutôt à la gestion de bases de données. En même temps, les archivistes, qui recourent de plus en plus à l'Internet pour diffuser leurs instruments de recherche, découvrent des façons plus dynamiques et efficaces de présenter l'information des instruments de recherche informatiques ; ces modes de conception des instruments de recherche permettent non seulement d'ajouter des fonctionnalités de recherche plus avancées à l'avantage des chercheurs, mais aussi d'en élargir considérablement le rayonnement.

Tous les centres d'archives qui ont adopté des systèmes informatiques pour la diffusion de leurs instruments de recherche doivent évidemment se poser le problème de la conversion rétrospective de tous les instruments de recherche produits avant le virage informatique.

C'est pourquoi, d'emblée, on doit se demander si les principes de conception des instruments de recherche doivent être les mêmes en mode papier et en mode électronique.

A. 1.1. Des principes identiques

Que la diffusion se fasse sur papier ou par l'informatique, les principes généraux de l'élaboration des instruments de recherche peuvent être considérés comme identiques.

- la description archivistique repose sur le respect des fonds et la structuration en fonction de la logique de production des documents ;
- l'organisation des instruments de recherche peut se fonder sur un grand nombre de niveaux (avec non redondance des informations d'un niveau à un autre) ;
- la présentation de la description se fait du général au particulier en situant l'unité décrite dans sa hiérarchie ;
- la nécessité d'une structuration des instruments de recherche est indépendante de la forme et du support de ceux-ci.

B. 1.2. Des modalités différentes

Toutefois, bien des modalités sont propres aux nouvelles formes électroniques. L'archiviste est confronté aux impératifs :

- de pérennisation des instruments de recherche qui maintiennent l'accès aux données (migration des données, traçabilité des mises à jour les plus importantes) ;
- de création partagée des descriptions ;
- d'aide plus efficace au public.

2. Quelle(s) solutions technologique(s) pour quel(s) usage(s) ?

Aujourd'hui, les instruments de recherche produits avec des outils informatiques peuvent recourir à :

- un simple traitement de texte,
- une base de données documentaire,
- des langages de structuration électronique.

La consultation des instruments de recherche par des lecteurs peut elle aussi recourir aux nouvelles technologies (cédérom, sites Web...). Ces différentes solutions aboutissent à des résultats différents.

A. 2.1. Les instruments de recherche électroniques non structurés

Malgré leurs limites, les instruments de recherche électroniques non structurés sont les plus couramment employés par les services d'archives :

- ils peuvent être produits avec un traitement de texte pour être imprimés, ou en HTML pour une diffusion sur Internet ;
- ils nécessitent au départ peu d'investissement et de compétences nouvelles ;
- ils permettent une mise à disposition rapide d'importantes ressources.

Par contre,

- ils ne peuvent être réutilisés sans interventions manuelles importantes et coûteuses à moyen terme ;
- ils n'offrent pas de possibilité d'échange, de pérennisation et de création partagée.

B. 2.2. Les instruments de recherche électroniques structurés

Les instruments de recherche électroniques structurés (bases de données documentaires, documents XML) :

- favorisent la mise à jour des informations et leur réutilisation ;

2. Quelle(s) solutions technologique(s) pour quel(s) usage(s) ?

- permettent de réaliser des instruments de recherche imprimés, des cédéroms ou des bornes interactives ;
- permettent de réaliser des pages HTML statiques ou des pages dynamiques (avec un moteur de recherche) sur des sites Web.

Attention

Le fait de structurer l'information et de la stocker dans un format ouvert indépendant des plates-formes logicielles, comme le xml, garantit la conservation à long terme des instruments de recherche eux-mêmes.

Ils présentent néanmoins deux inconvénients :

- lourdeur des enrichissements,
- nécessité de compétences informatiques et archivistiques de haut niveau.

3. Outils électroniques traditionnels

Il s'agit des instruments électroniques non structurés et des bases de données.

A. 3.1. Instruments de recherche électroniques non structurés

Plusieurs services ont mis en ligne des instruments de recherche au format PDF :

- soit à partir de documents en traitement de texte,
- soit à partir de documents scannés et convertis en mode texte.

Inconvénients du PDF : non retravaillés dans une perspective Web, les documents PDF ne peuvent pas être complétés et modifiés. En effet :

- leur forme est figée ;
- aucun lien ne peut être établi entre le contenu de divers instruments de recherche ;
- il n'y a pas de recherche possible sur des descripteurs ;
- le temps de chargement est très long pour l'internaute, d'où parfois la nécessité de découper les instruments de recherche en chapitres.

Le HTML (langage de balisage hypertextuel) :

- a permis à des services qui n'avaient jamais informatisé de mettre en ligne des instruments de recherche créés pour leur site Internet ;
- est très simple à écrire ;
- permet de coder une page à l'aide de commandes de mise en forme, qui seront ensuite interprétées par un navigateur.

Standard de fait pour les publications sur l'Internet, le HTML a cependant certaines limites :

- descriptions physique et structurelle sont mêlées, d'où une espérance de vie courte et une accessibilité limitée des documents ;
- HTML a surtout une fonction de navigation : il sert à créer des liens dans la même page, vers une autre page, vers des objets multimédias, etc. ;
- HTML oblige à écrire des liens entre différentes pages très lourds à maintenir ;
- HTML peine à gérer les documents de gros volume ou fortement structurés où la recherche plein texte est insuffisante ;
- HTML ne permet pas de vérifier le contenu des données et ne donne donc pas de

possibilité d'aider l'internaute dans ses requêtes.

B. 3.2. Bases de données

Une base de données est d'abord un outil de production et de gestion d'informations, adapté à une organisation, c'est-à-dire le service ou le centre d'archives dans toutes ses dimensions. Bien que la base de données soit d'abord créée pour mieux contrôler les informations des instruments de recherche, elle peut être utilisée à toutes les fins de gestion d'un service ou centre d'archives dont les impératifs plus larges interagissent en permanence avec la production d'inventaires.

Les bases de données documentaires comprennent différents champs structurant l'information (intitulé, dates extrêmes, présentation du contenu, description physique...).

Contrairement à ce qui s'écrit parfois, une base de données est compatible avec la description par niveaux et, encore plus, avec la création de liens multimédia. On peut même dire que ce sont les bases de données qui ont permis d'inventer les relations entre documents et les catégories d'informations hétérogènes.

En pratique, créer une édition respectant la description à plusieurs niveaux à partir d'une base de données décrivant des archives ne pose aucune difficulté majeure.

Exemple

Certains services d'archives français, les Archives départementales de la Savoie notamment, ont réalisé un export des données saisies dans la base pour les mettre en ligne sur Internet en utilisant des outils de bases de données libres (MySQL et développements JAVA).

En France, généralement, les bases de données ont été particulièrement utilisées :

- pour les versements administratifs contemporains, qui ne peuvent être exploités qu'avec une recherche indexée permettant de regrouper les versements de même origine et de retrouver les versements intéressant un même sujet (base Priam 3 des Archives nationales de Fontainebleau) ;
- pour les ensembles sériels de dossiers individuels (applications des Archives nationales, comme Nat, dossiers de naturalisation, ou Paris, dossiers de personnel des administrations centrales).

Une base de données est un outil indispensable pour "balayer" la totalité des fonds et trouver des informations qu'on ne penserait jamais à rechercher dans certaines sources. L'interrogation structurée et le thésaurus ont prouvé depuis longtemps leur supériorité sur les moteurs de recherche textuels.

Mais la base de données demeure un outil complexe à interroger pour l'utilisateur non averti, qui préférera toujours lire un texte et être guidé. L'interface d'interrogation ne remplace pas la lecture de l'inventaire, elle la complète.

L'orientation actuelle de certains services est donc de développer de nouvelles interfaces diversifiées pour leurs bases de données :

- une interface cartographique
- et des interfaces "guides".

Dans les deux cas, en cliquant par étapes successives dans des menus, l'utilisateur reproduira l'équivalent d'une interrogation de la base de données.

Soulignons enfin que l'existence de bases de données spécifiques à un service n'interdit pas l'interrogation croisée de toutes les bases existantes (par exemple tous les services d'Archives départementales) : à très faible coût, il est possible de mettre en place des outils d'interrogation qui s'adressent aux différentes interfaces des bases de données et sont en mesure de fournir les réponses disponibles dans ces bases sans capter de trafic (exemple

des bases bibliographiques de Savoie).

Attention

L'idée que les bases de données ne seraient pas en format pérenne appartient elle aussi au passé : la plupart des logiciels sont aujourd'hui capables d'exporter leurs données dans des formats textes structurés qui sont au niveau inférieur au XML, en termes informatiques. En langage ordinaire, cela signifie qu'ils peuvent générer des structures plus simples que le XML (type données en tableau par exemple) et donc que leur intégration dans une structure plus complexe ne pose pas de problème technique.

Complément : Pour plus de détails...

Pour en savoir plus : étude de la direction des Archives de France¹ sur les fonctionnalités documentaires des logiciels utilisés dans les archives publiques françaises.

La direction des Archives de France mène depuis 2006 une étude sur les fonctionnalités de description documentaire offertes par les logiciels de gestion des archives au regard des normes ISAD(G) et ISAAR(CPF) et de la DTD EAD.

L'étude prévue vise, pour chaque logiciel pris en compte, à répondre aux questions suivantes :

- est-il possible, avec le logiciel, de décrire les documents d'archives conformément à la norme ISAD(G) et les producteurs d'archives conformément à la norme ISAAR(CPF) ?
- si oui, de quelle manière faut-il utiliser le logiciel pour aboutir à des descriptions conformes à ces normes ?
- le logiciel permet-il de produire automatiquement des instruments de recherche au format XML suivant la DTD EAD ? Est-il possible d'intégrer dans le logiciel des documents structurés selon la DTD EAD ?

Le but de cette étude, toujours en cours, est triple :

- fournir une information précise aux services d'archives engagés dans le choix d'un outil ;
- donner des conseils pour l'utilisation de chaque logiciel ;
- apporter une aide aux éditeurs qui souhaiteraient améliorer la conformité de leur produit avec les normes citées.

Les résultats de cette étude sont disponibles sur demande.

1 - <http://www.archivesdefrance.culture.gouv.fr/gerer/informatisation/etude/>

4. Langages de balisage

Afin de faciliter le recours à l'informatique pour la création, la gestion et la diffusion de données à toutes les fins requises, les archivistes ont développé des outils permettant de définir et créer les outils indispensables à ces types d'informatisation du travail archivistique.

En adaptant des définitions de types de documents au contexte de la gestion des bases de données archivistiques, ces outils permettent à tous les centres d'archives d'adopter des modes de travail facilitant l'utilisation de l'Internet et de plusieurs autres méthodologies.

Le recours à ces méthodes, cependant, requiert un minimum de connaissances techniques pour comprendre l'environnement global dans lequel ces nouveaux modes de fonctionnement s'insèrent.

A. 4.1. Substrat technique

SGML, XML, HTML sont des langages créés pour des usages différents. Ils sont tous utilisés pour caractériser les éléments d'information diffusés sur Internet. Usages :

- Version allégée du SGML (Langage normalisé de balisage généralisé), le XML (Langage de balisage extensible) est un langage ouvert qui permet de créer ses propres balises.
- Alors que le HTML s'attache à la présentation d'un contenu, XML porte sur la description de ce contenu.
- Tout document XML est un arbre d'éléments, permettant de décomposer le texte en unités d'informations. L'arbre d'éléments est formé d'un et d'un seul élément racine qui contient tous les autres éléments.

Les éléments :

- sont définis par des balises de début de contenu `< >` et de fin de contenu `</ >` ;
- ont un code donné entre les chevrons identifié avec un nom en clair; ex. : `<nom>` pour nom ;
- peuvent contenir :
 - directement du texte,
 - d'autres éléments (éléments englobants),
 - des liens vers des ressources électroniques externes ou internes au document (éléments de lien),
 - peuvent être qualifiés par des attributs qui en en modifient le sens.

Exemple

Dans le cas présenté ci-dessous, nous avons un élément balise qui contient deux attributs :

- un premier nommé attribut1 dont la valeur est valeur1
- un deuxième nommé attribut2 dont la valeur est valeur2.

Le caractère "=" sépare le nom et la valeur, les guillemets délimitent la valeur et un espace sépare les attributs entre eux.

```
<balise attribut1="valeur1" attribut2="valeur2">...</balise>
```

La structure en arborescence d'un document XML peut être déclarée dans le corps du document XML ou dans un fichier à part :

- cette déclaration ou Définition de Type de Document (DTD) est facultative ;
- son avantage est de faire partager à une communauté d'utilisateurs une même structuration de l'information.

B. 4.2. La DTD EAD

Développée par la Société des archivistes américains, l'EAD (Description archivistique encodée) est une DTD pour créer, publier/diffuser sur Internet, conserver des instruments de recherche (inventaires, catalogues...) en version électronique.

1. 4.2.1. Structure de la DTD EAD

La documentation officielle de l'EAD se compose de trois parties (l'ensemble est disponible sur le site officiel de EAD, à la Bibliothèque du Congrès (<http://loc.gov/ead/>). Elle comprend :

- la DTD proprement dite, écrite en anglais (non modifiable sauf par ses auteurs)
- le dictionnaire des balises (Tag Library)
- le guide d'application (Application Guidelines).

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.

Complément : Pour plus de détails...

Pour en savoir plus : traduction française de la documentation de l'EAD

La documentation de l'EAD a été traduite en français par un groupe d'experts de l'AFNOR. Cette traduction est disponible sur le site de la direction des Archives de France (<http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/ead/>).

2. 4.2.2. Caractéristiques principales

- L'EAD est conçue pour être utilisée avec des normes de description, en particulier ISAD(G) , qui normalisent les contenus.
- L'EAD facilite l'échange des instruments de recherche : il est possible d'utiliser le

même document encodé pour des restitutions multiples.

- 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.
- L'EAD permet d'intégrer aux instruments de recherche des liens vers des représentations numériques des documents décrits.
- Il est possible de choisir la finesse de l'encodage en fonction des besoins et des moyens.

Attention

La première version officielle de l'EAD, publiée en août 1998, a été suivie d'une nouvelle édition en septembre 2002.

L'EAD est une norme de la Société des archivistes américains (SAA). Elle demeure la propriété de la SAA, c'est-à-dire qu'elle ne peut être modifiée sans l'accord de cette association des archivistes des États-Unis.

Complément : Pour aller plus loin...

Pour aller plus loin : la transformation de la DTD EAD en schéma XML

Un groupe de travail a été constitué en juillet 2004, à la demande du groupe de travail international chargé de la maintenance intellectuelle de l'EAD, afin de transformer la DTD EAD en schéma.

La version 1.0 du schéma XML de l'EAD est disponible depuis mars 2007 à : <http://www.loc.gov/ead/eadschema-test.html>

Pourquoi cette évolution ?

- Pour permettre une saisie plus cohérente des données :
 - la DTD est plus permissive et convient bien à la conversion rétrospective ;
 - le schéma impose plus de contraintes à la saisie courante (respect de listes de valeurs pour les attributs par exemple) ;
 - le développement de l'encodage direct en ligne par des équipes d'archivistes ou de bibliothécaires exige plus de contraintes à la saisie pour obtenir plus de cohérence dans les résultats.
- Pour être en mesure d'inclure le schéma EAD dans d'autres schémas (par exemple METS ou *Metadata Encoding and Transmission Standard* qui permet l'encodage des métadonnées de gestion des images numériques) ou d'inclure d'autres DTD ou d'autres schémas dans l'EAD (par exemple la TEI ou *Text Encoding Initiative* qui permet l'encodage de documents textuels).
- Pour gérer la *Tag Library* et les règles de gestion dans un document unique :
 - la *Tag Library* peut être intégrée dans le schéma EAD
 - chaque pays peut utiliser l'EAD dans sa propre langue (définir pour chaque langue une table de correspondance avec les noms des éléments, des attributs et des valeurs des attributs en anglais ; la définition de profils nationaux pour le schéma sera possible en fonction de la langue utilisée pour les noms des éléments, des attributs et de leurs valeurs et en fonction des valeurs retenues pour chaque attribut).

La DTD EAD version 2002 reste la version de référence jusqu'à nouvel ordre : le groupe chargé de la définition du schéma n'a ajouté aucun élément par rapport à la DTD.

Des outils de conversion ont été livrés avec le schéma pour rendre conformes par rapport au schéma les fichiers produits avec la DTD.

Néanmoins, le schéma reste peu utilisé pour le moment. La transformation en schéma XML du format EAC (Contexte archivistique encodé) pour encoder en XML les notices d'autorité

de producteurs devrait avoir des incidences sur les évolutions de l'EAD.

3. 4.2.3. Éléments de l'EAD

L'EAD comprend 146 éléments, dont un tout petit nombre sont obligatoires (en-tête EAD, description du fonds, niveaux hiérarchiques).

Trois éléments principaux s'inscrivent sous un élément racine <ead> :

- <eadheader> En-tête EAD, qui contient des informations bibliographiques et descriptives sur l'instrument de recherche, et non sur les documents d'archives que celui-ci décrit (obligatoire) ;
- <frontmatter> Préliminaires, qui sert à créer une première de couverture et à présenter les parties liminaires de l'instrument de recherche (facultatif) ;
- <archdesc> Description archivistique qui correspond à l'instrument de recherche lui-même (obligatoire).

4. 4.2.4. ISAD(G) et EAD

Idéalement, l'EAD utilise les éléments dont le contenu est déjà défini par la norme de contenu ISAD(G) ; néanmoins elle en modifie la présentation pour mieux utiliser toutes les possibilités de l'informatique.

- Développée ultérieurement, au départ sans lien avec le groupe d'experts de l'ICA 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.
- À chacun des éléments définis dans ISAD(G) correspond un élément 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.

Comparons :

ISAD(G) permet de produire un instrument de recherche dont le résultat ne diffère pas des répertoires numériques classiques :

- les informations de haut niveau (contexte de production, historique de la conservation, description physique, etc.) sont regroupées dans l'introduction de l'inventaire ;
- les parties propres au répertoire figurent dans le corps de l'instrument de recherche.

En EAD, la **description du producteur** fait déjà partie de la description archivistique.

Ainsi, l'élément Description archivistique <archdesc> comprend d'abord des informations de haut niveau :

- l'identification du fonds (élément <did>), avec la mention de l'institution où est conservé le fonds (<repository>),
- la cote du fonds (<unitid>),
- l'intitulé (<unittitle>),
- les dates extrêmes (<unitdate>),
- le nom du producteur (<origination>),
- la description physique globale (<physdesc>),
- des précisions sur les modalités d'accès (<accessrestrict>),
- éventuellement un aperçu du contenu du fonds (<scopecontent>).

On décrit **ensuite les parties du fonds**, dans l'élément englobant Description des composants <dsc> :

- on commence par le niveau hiérarchique le plus élevé et on descend si on le souhaite jusqu'à la pièce;

- douze niveaux hiérarchiques sont disponibles, mais on peut aller encore plus loin si nécessaire pour décrire les fonds complexes;
- à chaque niveau, tous les éléments de description peuvent s'appliquer dans un ordre très souple.

L'élément <archdesc> a plusieurs attributs spécifiques.

L'attribut LEVEL, obligatoire, permet de définir le niveau de description auquel on se situe :

- "fonds" (fonds),
- "collection" (collection),
- "subfonds" (sous-fonds),
- "series" (série organique),
- "subseries" (sous-série organique),
- "file" (dossier),
- "item" (pièce),
- ou "otherlevel" (autre niveau).

	ISAD(G)	EAD
3.1.1	Référence	<eadid> avec les attributs COUNTRYCODE et MAINAGENCYCODE <unitid> avec les attributs COUNTRYCODE et REPOSITORYCODE
3.1.2	Intitulé/analyse	<unittitle>
3.1.3	Dates	<unitdate>
3.1.4	Niveau de description	attribut LEVEL de <archdesc> et <c>
3.1.5	Importance matérielle et support de l'unité documentaire	<physdesc> et ses sous-éléments <extent>, <dimensions>, <genreform>, <physfacet>
3.2.1	Nom du producteur	<origination>
3.2.2	Histoire administrative/notice biographique	<bioghist>
3.2.3	Historique de la conservation	<custodhist>
3.2.4	Modalités d'entrée	<acqinfo>
3.3.1	Présentation du contenu	<scopecontent>
3.3.2	Évaluation, tris et éliminations, sort final	<appraisal>
3.3.3	Accroissements	<accruals>
3.3.4	Mode de classement	<arrangement>
3.4.1	Conditions d'accès	<accessrestrict>
3.4.2	Conditions de reproduction	<userrestrict>
3.4.3	Langue et écriture des documents	<langmaterial>
3.4.4	Caractéristiques matérielles et contraintes techniques	<phystech>
3.4.5	Instruments de recherche	<otherfindaid>
3.5.1	Existence et lieu de conservation des originaux	<originalsloc>
3.5.2	Existence et lieu de conservation de copies	<altformavail>
3.5.3	Sources complémentaires	<relatedmaterial> <separatedmaterial>
3.5.4	Bibliographie	<bibliography>

3.6.1	Notes	<odd> <note>
3.7.1	Notes de l'archiviste	<processinfo>
3.7.2	Règles ou conventions	<descrules>
3.7.3	Date(s) de la description	<processinfo><p><date>

Tableau 1 **Tableau de correspondance entre la Norme générale et internationale de description archivistique ISAD(G) et la DTD EAD 2002**

5. 4.2.5. Exercice

Cliquez sur le lien pour accéder à l'exercice.

6. 4.2.6. L'EAD et l'indexation à niveaux

L'EAD permet l'**indexation à niveaux**, corollaire de la description à niveaux :

- L'indexation à niveaux consiste à saisir au niveau de description pertinent les descripteurs utiles pour la recherche, seulement à ce niveau, sans les répéter plus bas dans l'arborescence du document EAD/XML, puisqu'ils qualifient de fait les niveaux de description inférieurs.
- Aux niveaux inférieurs, on pourra entrer d'autres descripteurs (points d'accès), qui pourront préciser ceux entrés plus haut (par exemple : en Nom géographique <geogname>, on saisira "Bruxelles", objet d'un sous-dossier contenu dans le dossier concernant la Belgique, "Belgique" étant enregistré comme point d'accès au niveau de la description du dossier).
- Deux types d'indexation sont possibles (les deux méthodes peuvent être utilisées conjointement dans un document EAD).
- L'indexation dans le texte permet de créer des liens entre un descripteur là où il se trouve et une recherche sur ce descripteur.

L'indexation hors du texte avec l'élément Vedettes et accès contrôlés <controlaccess> 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.

Complément : Pour plus de détails...

Pages consacrées à l'EAD sur le site de la direction des Archives de France

<http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/ead/> : lien vers le site officiel de l'EAD, traduction française de la Bibliothèque des balises version 2002, manuel d'encodage, référentiel des applications françaises de l'EAD, Bulletin d'information francophone sur l'EAD (37 numéros parus depuis 2000), informations sur les journées européennes de 2004, etc.

C. 4.3. La DTD EAC

L'EAD contient des éléments pour les noms de personnes, de familles et de collectivités, avec des attributs permettant des liens vers des notices d'autorité, mais ce format ne supporte pas des fichiers d'autorité et d'information contextuelle séparés.

Parfaitement compatible avec la deuxième édition de la norme ISAAR(CPF)²⁶, la DTD EAC (Contexte archivistique encodé) permet de structurer et d'échanger de l'information sur les

producteurs et sur le contexte de production des fonds d'archives.

Les fichiers XML créés à partir de l'EAC, sont similaires aux fichiers d'autorité des bibliothèques, mais avec en plus des renseignements biographiques ou administratifs sur la personne physique ou morale concernée.

Cette DTD a été appelée EAC pour mettre en valeur ses relations avec l'EAD (Description archivistique encodée) :

- ce nouveau standard permet en effet de structurer l'information habituellement placée dans l'élément Biographie ou histoire <bioghist> de l'EAD, qui correspond à l'élément biographie/histoire administrative d'ISAD(G) ;
- l'avantage d'extraire cette information de la structure d'un instrument de recherche et de la maintenir dans un fichier EAC séparé est de lier plus facilement la même information sur une personne physique ou morale à plusieurs instruments de recherche.

Complément : Pour plus de détails : l'EAC

Le groupe d'experts « Données d'autorité » de l'AFNOR (CG46/CN357/GE4), piloté par la Bibliothèque nationale de France, a commencé la traduction en français du dictionnaire des balises de la version beta, consultable sur Internet à : <http://www3.iath.virginia.edu/eac/> (dernière mise à jour : 29 novembre 2004).

Par ailleurs, une page est consacrée à cette nouvelle DTD sur le site de la direction des Archives de France (<http://www.archivesdefrance.culture.gouv.fr> rubrique archivistique/description archivistique/ Informatisation des noms de personnes, familles et collectivités : la DTD EAC).

Toutefois, l'EAC n'est pas stabilisée puisqu'elle en est toujours, depuis août 2004, à une version Beta. Un groupe international a été recréé, piloté par la Société des archivistes américains pour faire évoluer le format en une version stable et également pour le rendre plus conforme aux règles de développement des schémas XML. Cette nouvelle version devrait être rendue disponible d'ici la fin 2009. Voir le site officiel de développement de cette nouvelle version ou EAC-CPF (Contexte archivistique encodé - Collectivités, personnes, familles) : <http://eac.staatsbibliothek-berlin.de/>

Quelques sites donnent déjà accès à des notices d'autorité encodées en XML conformément à la DTD EAC.

En France, les Archives nationales ont mis en ligne l'application **ETANOT** (ETat des NOTaires parisiens) (<http://chan.archivesnationales.culture.gouv.fr/sdx/etanot/index.xsp>), qui est une base d'autorités, décrivant non pas des documents d'archives, mais des producteurs d'archives (notaires parisiens du XVe siècle à nos jours). Un travail est en cours pour publier sur Internet les instruments de recherche du Minutier central décrivant les archives des notaires recensés dans la base ETANOT, en commençant par les répertoires (listes d'actes) produits par ces notaires. ETANOT s'inscrit en effet dans un projet plus général, baptisé NOEMI (NOTaires et MInutes), qui doit être, à terme, un ensemble de bases de données coordonnées et accessibles sur Internet, sous la forme d'un portail.

La direction des Archives de France a développé une application BORA Archives photographiques (<http://daf.archivesdefrance.culture.gouv.fr/sdx23/pl/>), qui donne accès aux descriptions des fonds et collections photographiques conservés dans les services d'archives publics français. BORA Archives photographiques contient aussi quelques notices d'autorité structurées en XML/EAC relatives aux photographes auteurs des documents et liées aux fichiers XML/EAD décrivant les sources photographiques.

En 2003, les **Archives de la ville de Genève** ont lancé sur le web un module consacré à la description des fonds recourant à ISAD(G) et XML/EAD. En décembre 2007, elles ont mis en ligne un second module de description, consacré, cette fois, aux notices d'autorité. Ce module suit la norme internationale ISAAR(CPF) et fonctionne grâce à une solution informatique en XML/EAC. Les notices réalisées peuvent être consultées sur le site des AVG

en se rendant sur la page <http://w3public.ville-ge.ch/seg/xmlarchives.nsf/RechercheFonds?OpenForm> et en cliquant ensuite sur les différentes catégories de fonds proposées sous la rubrique « Accès direct aux fonds ».

Le portail **Censo Guia de los Archivos**² est un projet collaboratif impliquant plus de 40 000 institutions d'Espagne et d'Amérique latine (voir présentation générale du projet avec liste des conventions passées). Le portail comprend trois types de descriptions, toutes structurées en XML : descriptions des fonds (EAD), de leurs producteurs (EAC) et de leurs lieux de conservation (DTD EAG ou Guide archivistique encodé, DTD développée par le ministère de la Culture espagnol pour ce projet).

D. 4.4. La mise en œuvre de l'EAD et de l'EAC

L'EAD a rencontré d'emblée un vif succès et est utilisée désormais dans de nombreux pays, notamment francophones (France, Suisse, Belgique, etc.) alors que la mise en œuvre de l'EAC relève encore dans la plupart des cas d'une phase expérimentale.

Quelques éléments de réflexion peuvent être dégagés avec ces différentes expériences :

- l'application des normes ISAD(G) et ISAAR(CPF) est un préalable indispensable à la mise en œuvre de l'EAD et de l'EAC ; il s'agit aujourd'hui d'élaborer des outils de recherche cohérents dépassant la simple description de ses fonds propres ; or, la normalisation des descriptions est un impératif qui n'est pas toujours intégré dans les pratiques au quotidien ; par ailleurs, chaque institution doit définir ses propres règles d'utilisation des formats ;
- les problèmes rencontrés lors des projets de conversion rétrospective demandent des adaptations techniques (numérisation des documents imprimés, reconnaissance optique des caractères ou saisie manuelle si la qualité des documents originaux est trop mauvaise...), mais aussi une reprise de l'analyse archivistique pour la rendre plus cohérente et conforme aux normes de description ;
- la mise en œuvre de l'EAD et de l'EAC se heurte souvent aux limites financières et techniques des institutions. De plus, informaticiens et professionnels des archives et des bibliothèques doivent apprendre à travailler ensemble pour mettre en œuvre de véritables projets collectifs. Ce sont surtout des choix stratégiques à long terme que les responsables de services sont conduits à prendre en adoptant l'EAD. Il s'agit essentiellement de déterminer quels fonds encoder en priorité en fonction des besoins du public et des moyens dont on dispose ;
- les différentes expériences en cours pour la publication sur Internet de corpus organiques complets de documents d'archives numérisés ont montré que la mise en relation informatique des instruments de recherche avec les documents qu'ils décrivaient n'était pas sans conséquence pour le contenu des descriptions documentaires. Il est probable qu'en préparant cette mise en relation, l'archiviste ressent la nécessité d'affiner l'instrument de recherche pour rendre plus commodes la visualisation ou le feuilletage des images numériques ;
- par ailleurs, si le substitut numérique du document primaire est visualisé à l'écran en même temps que sa description, il sera peut-être nécessaire de modifier cette description, de la vérifier et de la compléter par des informations sur le support ou le format du document original ou par des éléments d'indexation ;
- enfin, l'environnement numérique, en dématérialisant le document, induit des évolutions qui sont des facteurs de rapprochement de métiers d'essences différentes. D'autres communautés patrimoniales peuvent avoir besoin des normes et formats d'échange des archivistes. C'est ainsi que depuis peu, on voit se développer des

2 - http://aer.mcu.es/sgae/index_censo_guia.jsp

relations de plus en plus étroites avec d'autres professions apparentées, surtout les bibliothèques.

Vous trouverez dans les écrans suivants des éléments de stratégie pour la production et la publication de descriptions archivistiques en EAD.

1. 4.4.1. Stratégies de production : comment faire de l'EAD ?

Avant de produire des descriptions XML, il faut se poser certaines questions pour connaître les besoins du service (et choisir les réponses appropriées) :

- Quel est l'environnement d'informatique documentaire : GED, SGBD, bureautique, sécurité du réseau, schémas directeurs ?
- Qui va produire : combien de personnes, quelle formation ?
- Quels fonds vont être traités : fonds sériel, fermé, ouvert, manuscrits ?
- Quel type d'instruments de recherche va être produit : répertoire synthétique ou analytique, catalogue ?
- Quel va être l'usage du document EAD : publication en ligne (intranet ou internet), publication papier ?

En fonction des réponses apportées, on peut être amené à choisir :

- de confier la production à une société extérieure, si la production est trop faible pour investir (matériel, formation),
- s'il y a de grandes quantités à produire rapidement (conversion rétrospective), de s'orienter vers une production interne avec un outil de production.

Attention

Les circonstances propres à chaque service peuvent amener à prendre des décisions qui mêlent le recours à une société extérieure et une production en interne.

2. 4.4.2. Les outils de production

Il existe trois grandes familles d'outils de production :

- les éditeurs XML, propriétaires (XMetaL, XMLSpy, Oxygen, ArborTEXT Editor, etc.) ou gratuits (EditXML, XMLMind, etc.),
- les logiciels documentaires qui ont une fonction d'exportation en EAD (le logiciel ICA-AtoM en cours de développement pour le Conseil international des archives ; pour la France, Arkheïa, Gaïa et Thot),
- l'utilisation des produits bureautiques classiques, avec une saisie directe en XML, des modules de conversion de bases de données, etc.

Les éditeurs XML :

- sont les outils les plus appropriés pour produire des inventaires rédigés de fonds complexes (structure compliquée, description très détaillée, etc.),
- peuvent être paramétrés en fonction des besoins particuliers (macros, habillages, etc.),
- mais sont d'une prise en main plus ou moins ergonomique et, comme tout éditeur de texte, il faut se les approprier (pratique régulière).

Les logiciels documentaires qui ont des fonctions d'export en EAD :

- se réfèrent tous à ISAD (G) (de façon plus ou moins complète),
- sont les outils les plus appropriés pour décrire des documents dont la nature se prête à une saisie dans une base de données structurée,
- offrent des possibilités plus ou moins souples d'adaptation (dépendantes du

prestataire et de l'évolution du produit) et les exports EAD doivent souvent être retravaillés (enrichissements possibles),

- mais sont très ergonomiques et peuvent être pris en main rapidement.

Dans un service d'archives qui utilise un logiciel de gestion, il peut suffire d'utiliser les fonctions d'export du logiciel en spécialisant une ou deux personnes (équipées d'un éditeur de XML) dans le contrôle et l'enrichissement des documents EAD automatiquement produits.

Un service peut aussi envisager une solution mixte en fonction de ses ressources humaines et financières.

3. 4.4.3. Stratégies de publication et de recherche : comment utiliser de l'EAD ?

Les questions de publication et de recherche sont indissociables des choix qui sont faits lors de la production des instruments de recherche en EAD.

Avant d'offrir au public ses instruments de recherche, il faut (aussi) se poser certaines questions (et choisir les réponses appropriées) :

- Souhaite-t-on diffuser les instruments de recherche en intranet ou en internet ?
- Souhaite-t-on publier les instruments de recherche sous plusieurs formes : inventaire imprimé et inventaire en ligne ?
- Souhaite-t-on se contenter d'un affichage statique type HTML ?
- Souhaite-t-on donner la possibilité de manipuler les informations pour la recherche (Web dynamique, conversion des données vers d'autres formats) ?

En fonction des réponses apportées, on peut être amené à choisir :

- L'utilisation techniquement simple d'une feuille de style CSS pour gérer l'affichage du document,
- Une solution un peu plus complexe d'affichage à travers l'utilisation de transformations XSL-T pour obtenir une version statique affichable et imprimable,
- L'installation d'un outil de gestion et de publication de documents XML.

Attention

Pour une publication papier de l'instrument de recherche, on s'orientera de préférence vers une transformation XSL-FO.

4. 4.4.4. Les outils de gestion et de publication

Les logiciels propriétaires de gestion et de consultation de descriptions XML/EAD sont encore peu nombreux.

- On peut citer pour la France : Digiteo de V-Technologies (Archives départementales du Puy-de-Dôme, du Maine-et-Loire, de la Vendée), Mnesys (Archives départementales du Doubs, de la Drôme, du Cantal).

Les logiciels libres ne sont pas plus nombreux et ne constituent pas (encore) une solution simple et miraculeuse pour publier en ligne les documents EAD.

- On peut citer pour la France : Pleade (Archives départementales de la Haute-Marne, du Val-de-Marne, de Loire-Atlantique, etc.).

5. 4.4.5 Un investissement non négligeable mais qui en vaut la peine

Dans tous les cas, il n'existe pas de solution qui ne demande pas un investissement de la

part de l'établissement :

- lors de la rédaction d'un cahier des charges complet et cohérent (pour acquérir un logiciel propriétaire ou pour faire paramétrer un logiciel libre),
- pour s'approprier les logiciels (« libres » ou non),
- pour garantir l'interopérabilité entre les systèmes.

Mais cet investissement en vaut la peine :

- des inventaires pérennes,
- un public élargi,
- un meilleur service,
- de nouveaux usages.

Complément : Pour plus de détails...

Pour en savoir plus : quelques exemples de publication en ligne de descriptions EAD et EAC (cf. Publication en ligne de descriptions EAD et EAC)

5. Évaluation des connaissances

Les exercices sont, à l'image de cette section, conçus pour des archivistes ayant déjà une bonne expérience de leur profession.

Avant de tenter de les résoudre, pour vous éviter tout découragement, il vous est conseillé de bien relire cette section, surtout le chapitre 4, et de suivre avec attention la correction de l'exercice du paragraphe 4.2.5. de façon à parfaitement comprendre.

A. 5.1. Exercice 1

Cliquez sur le lien pour accéder à l'exercice.

B. 5.2. Exercice 2

Répondre aux questions suivantes :

Question 1

1. Quelle est la différence entre le XML et le HTML ?

Question 2

2. Qu'est-ce qu'une balise ? Qu'est-ce qu'un élément ? Qu'est-ce qu'un attribut ? Qu'est-ce qu'une DTD ?

Question 3

3. Quelle est la différence entre un document valide et un document bien formé ?

Question 4

4. Qu'est-ce qu'un parseur ?

Question 5

5. Quelle est la norme de description sur laquelle est basée l'EAD 2002 ? Pouvez-vous citer une ou deux équivalences en EAD des éléments de description définis par cette norme ?

Question 6

6. Quelle est la structure de base d'un document EAD ?

Question 7

7. Quelles sont les trois grandes familles d'outils de production de documents XML/EAD ?

Question 8

8. Comment trouver de l'information sur l'EAD ?

Question 9

9. Citer deux exemples de sites utilisant l'EAD

Question 10

10. Quelle est la norme de description sur laquelle est basée l'EAC ?

Question 11

11. Comment trouver de l'information sur l'EAC ?

Question 12

12. Citer un exemple de site utilisant l'EAC

Question 13

13. Quels avantages et inconvénients trouvez-vous à l'EAD et à l'EAC ?

Glossaire

ISAAR (CPF)

Norme internationale sur les notices d'autorité archivistiques relatives aux collectivités, aux personnes et aux familles.

ISAD (G)

Norme générale et internationale de description archivistique qui fournit, dans le cadre du respect des fonds, une structure type de description pour les archives de toute forme, de tout support et de toute époque.

Bibliographie

[Bureau canadien des archivistes]

Comité de planification sur les normes de description. *Règles pour la description des documents d'archives*. Ottawa, Bureau canadien des archivistes, 1990-2005 (tous les chapitres se retrouvent sur le site du Conseil canadien des archives à l'adresse : <http://www.cdncouncilarchives.ca/f-archdesrules.html>).

[Charbonneau]

NORMAND CHARBONNEAU. «La diffusion» in *Les fonctions de l'archivistique contemporaine*, COUTURE CAROL ET COLLABORATEURS, Québec, Presses de l'Université du Québec, 1999.

[Conseil international des Archives]

La norme générale et internationale de description archivistique [ISAD (G)]. Paris, Conseil international des Archives, 2000, 2e éd. (<http://www.ica.org/en/node/30001>).

[Conseil international des Archives]

Norme internationale sur les notices d'autorité archivistiques relatives aux collectivités, aux personnes et aux familles [ISAAR (CPF)]. Paris, Conseil international des Archives, 1996 (<http://www.ica.org/en/node/30231>).

[Maurel, Champagne]

DOMINIQUE MAUREL ET MICHEL CHAMPAGNE. «La description et l'indexation» in *Les fonctions de l'archivistique contemporaine*, COUTURE CAROL ET COLLABORATEURS, Québec, Presses de l'Université du Québec, 1999.

[Nougaret]

CHRISTINE NOUGARET. « Classement et description : des principes à la pratique » in *La Pratique archivistique française*, DIR. FAVIER JEAN. Paris, Direction des Archives de France, 2008, p. 257.

[Nougaret]

CHRISTINE NOUGARET . *Les instruments de recherche en archivistique*. Paris, Direction des Archives de France et La Documentation française, 1999.