

XML

Yannick Prié
UFR Informatique – Université Lyon 1

UE2.2 – Master SIB M1 – 2004-2005

Objectifs des trois cours

- Etre capable de comprendre des documents XML et des DTD
- Etre capable de construire des documents XML et des DTD
- Découverte de quelques DTD « importantes »

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

2

Un document XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE livre SYSTEM "E:\prié\Enseignement\2004-2005\Master
SIB\SIB2.2-bloc2-XML\Fichier CM XML\exemple-intro.dtd">
<livre id="561" nbpages="190" titre="La compagnie des spectres">
  <auteur>
    <nom>Salvayre</nom>
    <prenom>Lydie</prenom>
  </auteur>
  <format type="poche">
    <mesure type="largeur" unite="cm">11</mesure>
    <mesure type="longueur" unite="cm">19</mesure>
    <mesure type="hauteur" unite="mm">10</mesure>
  </format>
</livre>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

3

La DTD correspondante

```
<!ELEMENT livre (auteur, format)>
<!ATTLIST livre
  id CDATA #REQUIRED
  nbpages CDATA #REQUIRED
  titre CDATA #REQUIRED >
<!ELEMENT auteur (nom, prenom)>
<!ELEMENT format (mesure+)>
<!ATTLIST format
  type CDATA #REQUIRED >
<!ELEMENT mesure (#PCDATA)>
<!ATTLIST mesure
  type (hauteur | largeur | longueur) #REQUIRED
  unite (cm | mm | in) #REQUIRED >
<!ELEMENT nom (#PCDATA)>
<!ELEMENT prenom (#PCDATA)>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

4

Plan

- Documents XML
 - Syntaxe XML et documents bien formés
- Types de documents XML
 - DTD et documents valides
 - Introduction à XML-Schema
- Le monde XML
 - Quelques normes liés à XML
 - Quelques DTD importantes

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

5

Plan

- **Documents XML**
 - Syntaxe XML et documents bien formés
- Types de documents XML
 - DTD et documents valides
 - Introduction à XML-Schema
- Le monde XML
 - Quelques normes liés à XML
 - Quelques DTD importantes

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

6

Qu'y a-t-il dans un document XML ?

- o Prologue
 - En-tête XML
 - Déclarations de DTD
 - o Instructions pour les processeurs XML
 - Instructions de traitement
 - o Instructions pour applications externes
- o Arbre des éléments
 - Eléments
 - o Balises XML pour le marquage
 - o Contenu
 - texte
 - autres éléments
 - Attributs des éléments
 - o Information associées aux éléments
- o Commentaires

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

7

Déclaration XML

- o Syntaxe générale

```
<?xml version="1.0" [encoding = "encodage"] [standalone="yes|no" ?>
```
- o Est une des informations de traitement
- o Indique
 - Conformité du document à une version de la norme XML
 - o version="1.0"
 - Jeu de caractères utilisé dans le document
 - o encoding = "UTF-8"
 - Présence ou non de références externes
 - o standalone="yes"

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

8

Éléments : règles de base

- o Un nom d'élément
 - commence par une lettre ou souligné
 - contient des lettres, chiffres, et ".", "-", "_", ":", "*"
 - peut posséder un nom de domaine
 - o domaine:nom_element
 - o Ex.: xsl:template
- o Les noms d'éléments dépendent de la casse
 - <nom_element> ≠ <nom_Element>
- o Balises
 - de début : <nom_element>
 - de fin : </nom_element>
- o Les éléments peuvent être vides
 - pas de contenu
 - <element_vide />
 - Ex :
 - o

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

9

Arbre des éléments

- o Un seul élément racine qui contient tous les autres
- o Pas d'intersections entre éléments
 - Mauvais : <nom1><nom2>...</nom1></nom2>
 - Bon : <nom1><nom2>...</nom2></nom1>
- o Blancs ou retours chariot en général non significatifs
 - <section><p> ... </p></section>
 - <section>
<p> ... </p>
</section>
- o Les éléments sont ordonnés

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

10

Caractères spéciaux

- o Ces caractères ont une signification spéciale pour les parsers XML
- o Il faut les écrire différemment
 - < <
 - > >
 - & &
 - ' '
 - " "

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

11

Attributs : règles de base

- o Dans les balises ouvrantes
 - <el att1="valeur1" att2="valeur2">
- o Les noms d'attributs dépendent de la casse
 - <el att1="valeur1" Att1="valeur2">
- o Valeurs d'attributs entourées
 - par des guillemets (") ou des apostrophes (')
- o Les attributs sont non-ordonnés

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

12

Attributs

- Les valeurs peuvent être
 - des données textuelles
 - value="N'importe quoi"
 - des *tokens* (noms XML) simples
 - value = "blue"
 - des ensemble *de tokens*
 - value = "red green blue"
- Possibilité d'énumérer les valeurs possibles et de mettre des valeurs par défaut (voir DTD)

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

13

Attributs de type ID et IDREF(S)

- Permettent des relations non hiérarchiques entre éléments
 - ID : identificateur unique dans le document XML
 - IDREF : référence à un élément ayant un attribut de type ID
 - IDREFS : références à des éléments ayant un attribut de type ID
- Exemple

```
<société codes_services="A001 A003">
<service code="A001">
<employé code="E205" code_service="A001"> Jean Dupont </employé>
<employé code="E206" code_service="A001"> Frédéric Marc </employé>
<employé code="E207" code_service="A001"> Fabrice Detterne
</employé>
<employé code="H107" code_service="A003"> Angélique Millet
</employé>
</service>
<service code="A003">
<employé code="A115" code_service="A003"> Isabelle Mascot
</employé>
</service>
</société>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

14

Commentaires

- Les commentaires ne sont pas considérés comme faisant partie du document XML.
 - `<!-- Un commentaire -->`
- Pas de `--` dans un commentaire !
- Un commentaire ne peut pas se trouver dans une autre déclaration

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

15

Instructions de traitement

- Informations nécessaire à une application externe
- Format :
 - `<?NomApplication paramètres ?>`
- Exemples
 - Déclaration XML obligatoire en début de fichier
 - `<?xml version='1.0' ?>`
 - Déclaration de feuille de style à utiliser
 - `<?xmlstylesheet href="fichier.xsl" type="text/xsl"?>`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

16

Déclarations

- Instructions pour le processeur XML
- Format : `<! ... >` ou `<! ... [<! ... >]`
 - Document type - `<!DOCTYPE ... >`
 - Character data - `<![CDATA[...]>`
 - Entities - `<!ENTITY ... >`
 - Notation - `<!NOTATION ... >`
 - Element - `<!ELEMENT ... >`
 - Attributes - `<!ATTLIST ... >`
 - `<![INCLUDE[...]]>` et `<![IGNORE[...]]>`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

17

Déclaration Document Type

- Identifie le nom de l'élément racine du document
 - `<!DOCTYPE My_XML_Doc>`
- Permet aussi de rajouter des définitions d'entités et des DTD
 - `<!DOCTYPE My_XML_Doc [...] >`
`<My_XML_Doc>`
...
`</My_XML_Doc>`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

18

Déclaration *Character Data*

- Dans les occasions pour lesquelles le texte doit contenir des caractères qui ne doivent pas être interprétés
- Deux textes équivalents
 - `Press <<<ENTER>>>`
 - `<![CDATA[Press <<<ENTER>>>]]>`

Au bilan : dans un document XML

- Prologue
 - en-tête
 - déclaration de DTD
 - instructions de traitement
- Eléments
 - attributs
 - contenus
- Commentaires

Plan

- Documents XML
 - Syntaxe XML et documents bien formés
- **Types de documents XML**
 - DTD et documents valides
 - Introduction à XML-Schema
- Le monde XML
 - Quelques normes liés à XML
 - Quelques DTD importantes

Traiter automatiquement un document XML

- Parser
 - Outil qui lit un document XML et construit l'arbre des éléments en mémoire
- Vérifier qu'un document répond bien à la syntaxe XML
 - Document *bien formé*
 - Possibilité de l'utiliser en tant que tel
 - ex. : le présenter à l'utilisateur
- Vérifier en plus qu'un document suit bien la grammaire définie dans une DTD
 - Document *valide*

Document Type Definition

- Définir le type de document XML voulu
 - décrire comment construire un document XML qui lui corresponde (grammaire)
- Permet de
 - valider un document XML (parser validant)
 - vérifier que tous les éléments sont présents et corrects
 - vérifier que les noms d'attributs et leurs valeurs sont corrects
 - transmettre cette connaissance à d'autres
 - ils pourront définir leurs propres documents XML dans le même cadre
 - d'où possibilité de standardisation et d'échanges

DTD

- Un fichier
 - contenant la définition formelle de la structure autorisée,
 - qui décrit donc
 - quels noms sont utilisés pour les types d'éléments
 - comment ces types d'éléments s'organisent
 - Ordre
 - Hiérarchie
 - les attributs des éléments
 - des entités analysables ou non
 - des notations pour les types de données binaires
- Liaison DTD / document XML
 - La DTD est dans le document XML (inline)
 - Le document XML réfère à la DTD avec une URI

DTD et document XML

Valide, contraint

```

<!ELEMENT auteur (nom, prenom, adresse)*
<!ELEMENT nom (#PCDATA)
<!ELEMENT prenom (#PCDATA)
<!ELEMENT adresse (#PCDATA)
<!ELEMENT mesure (type, unite, valeur)
<!ELEMENT format (type, unite, valeur)
  
```

Document XML

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPEivre SYSTEM "E:\gniel\enseignement\2004-2005\
Master SIB\SIB2-2_bloc2-XML\Fichier CM.XML\exemple-intro.dtd">
< id="561" nbpages="190" titre="La compagnie des spectres">
< auteur>
< nom>Salvayre</nom>
< prenom>Lydie</prenom>
< adresse>
< format type="poche">
< mesure type="largeur" unite="cm" >11</mesure>
< mesure type="longueur" unite="cm" >19</mesure>
< mesure type="hauteur" unite="mm" >10</mesure>
</format>
</auteur>
  
```

Est validé par, Est contraint par

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

25

Déclarations d'éléments

- Définir un élément et son contenu
 - `<!ELEMENT name (#PCDATA)>`
⇒ `<name> ... </name>`
- Un élément vide n'a pas de contenu
 - `<!ELEMENT name EMPTY>`
⇒ `<name/>`
- Si on autorise les fils
 - Quelconques :
`<!ELEMENT name ANY>`
 - Spécifiés :
`<!ELEMENT person (name, e-mail*)>`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances


26

Spécification des fils (grammaire)

- Définir le contenu des éléments
`<!ELEMENT person (name, e-mail*)>`
- ...et définir une hiérarchie d'éléments

```

<!ELEMENT name (fname, surname)>
<!ELEMENT fname (#PCDATA)>
<!ELEMENT surname (#PCDATA)>
<!ELEMENT e-mail (#PCDATA)>
  
```


- Organisation des sous-éléments
 - Connecteur de séquence ',': `(A, B, C)` [puis]
 - Connecteur de choix '|': `(A | B | C)` [ou]

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

27

Indicateurs de quantité

- Contraintes sur les éléments des DTD

A?	Possible	[0..1]
A+	1 fois et plus	[1..*]
A*	0 ou plus	[0..*]
- Exemples
`((A,B?) | C+)*`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

28

Déclaration d'attributs

- Les attributs sont associés aux types d'éléments
- Déclarés dans une déclaration ATTLIST
 - `<!ELEMENT element ... >`
 - `<!ATTLIST element ... >`
 - Il faut ensuite définir
 - le nom de l'attribut
 - le type de l'attribut
 - sa valeur par défaut

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

29

Noms et types d'attributs

- Noms d'attributs
 - `<!ATTLIST elem name type default>`
 - `<!ATTLIST elem first_attr ...`
`second_attr ...`
`third_attr ... >`
- Types d'attributs

CDATA	ID
NMTOKEN	IDREF
NMTOKENS	IDREFS
ENTITY	NOTATION
ENTITIES	name group

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

30

Types d'attributs (1)

- o CDATA
 - Chaîne de caractères
 - `<!ATTLIST person name CDATA ... >`
 - `name = "Tom Jones"`
 - o NMTOKEN
 - Token unique
 - `<!ATTLIST mug color NMTOKEN ... >`
 - `color="red"`
 - o NMTOKENS
 - Multiples tokens
 - `<!ATTLIST temp values NMTOKENS ... >`
 - `values="12 15 34"`
- Joue sur la manière dont le parser interprète l'attribut

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

31

Types d'attributs (2)

- o ENTITY
 - L'attribut est une référence d'entité
 - `<!ATTLIST person photo ENTITY ... >`
 - `photos="MyPic"`
- o ENTITIES
 - Plusieurs références d'entités
 - `<!ATTLIST album photos ENTITIES ...>`
 - `photos="pic1 pic2"`
- o ID
 - Identificateur unique
 - `<!ATTLIST person id ID ... >`
 - `ID = "P09567"`
- o IDREF
 - Référence à un ID
 - `<!ATTLIST person father IDREF ... >`
 - `IDREF="P09567"`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

32

Types d'attributs (3)

- o IDREFS
 - Référence à plusieurs ID
 - `<!ATTLIST person children IDREFS ... >`
 - `IDREFS="A01 A02"`
- o NOTATION
 - Décrit des données non XML
 - `<!ATTLIST image format NOTATION (TeX|TIFF) ...>`
 - `FORMAT="TeX"`
- o Name group
 - Liste restreinte
 - `<!ATTLIST point coord (X|Y|Z) ... >`
 - `coord="X"`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

33

Valeurs d'attributs par défaut

- o Quatre types
 - **#REQUIRED** Doit être spécifié
 - **#IMPLIED** Peut être spécifié
 - **"default"** Valeur par défaut si non spécifié
 - **#FIXED** Une seule valeur autorisée

```
<ATTLIST tag name type default>  
<!ATTLIST seqlist sepchar NMTOKEN #REQUIRED  
type (alpha|num) "num"
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

34

Mise en place de DTD

- o Utiliser les composants de XML...
 - Entités, éléments, déclarations, instructions de traitements, listes d'attributs, etc.
- o ... dans des DTD pour spécifier les règles permettant de valider des documents XML
 - Définir un modèle (type) de document de façon formelle
- o Une DTD décrit
 - Quels noms peuvent être utilisés pour les types d'éléments
 - L'ordre dans lesquels ceux-ci peuvent apparaître
 - La hiérarchie documentaire
 - Les noms et les types des attributs d'éléments

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

35

Déclaration de DTD

- o La DTD est stockée
 - soit dans le fichier XML
 - soit dans un fichier extérieur
 - soit dans les deux
- o Une DTD interne peut écraser ou ajouter des ENTITY ou des ATTLIST à des définitions de DTD externes
- o Une DTD est composée de déclarations
 - ELEMENT – Définitions d'éléments
 - ATTLIST – Définitions d'attributs
 - ENTITY – Définitions d'entités
 - NOTATION – Définitions de notations

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

36

Définition interne de DTD

- o Dans la déclaration DOCTYPE

```
<?xml version="1.0" standalone="yes" ?>
<!DOCTYPE racine [
  <!-- Ici la DTD -->
  <! ... >
  <! ... >
]>
<!-- Rest of XML file -->
<racine>
...
</racine>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

37

Définition externe privée de DTD

- o Référence à la DTD externe par un chemin dans la déclaration DOCTYPE

```
<?xml version="1.0" standalone="no" ?>
<!DOCTYPE racine
  SYSTEM "../MyDoc.dtd" [
  <!-- Extra declarations -->
  <! ... >
  <! ... >
]>
<!-- Rest of XML file -->
```

DTD externe
privée

- Les déclarations spécifiques au document restent définies de façon interne

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

38

Définition externe privée de DTD

- o Référence à la DTD externe par une URL dans la déclaration DOCTYPE

```
<?xml version="1.0" standalone="no" ?>
<!DOCTYPE racine
  SYSTEM "http://.../MyDoc.dtd" [
  <!-- Extra declarations -->
  <! ... >
  <! ... >
]>
<!-- Rest of XML file -->
```

DTD externe
privée

- Les déclarations spécifiques au document restent définies de façon interne

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

39

Définition externe publique de DTD

- o Utilisation du mot-clé PUBLIC

```
<!DOCTYPE racine
  PUBLIC "identifiant public" "url" ?>
```

- o Exemple

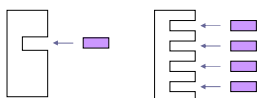
```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

40

Les entités

- o Sont des alias associant un nom à des "unités d'information"
- o Les entités spécifiques au document sont décrites dans sa DTD interne
- o Les entités plus générales sont décrites dans des DTD externes
- o Chaque entité
 - o est identifiée par un nom
 - o est définie par une déclaration d'entité
 - o est utilisée en appelant une référence d'entité



CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

41

Utilisation des entités

- o Quand l'information

- Est utilisé dans plusieurs endroits
 - o Ex. : déclaration légale, caractère spécial
- Est une partie d'un document qui doit être tronçonné pour rester gérable
 - o Ex. : livre : 1 fichier + n chapitres : n fichiers
- Est conforme à un format de donnée différent de XML
 - o Ex. : image JPEG

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

42

Types d'entités

- o Entités internes
 - générales
 - o Utilisées dans les documents XML
 - paramètre
 - o Utilisées dans les déclarations dans les DTD
- o Entités externes
 - générales
 - paramètres
- o Entités analysables
- o Entités non analysables
- o Entités caractères
 - déjà vues

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

43

Entités générales internes

- ```
<!ENTITY nom "chaîne de remplacement" >
```
- o Entités analysables utilisées uniquement dans le document
  - o Référence : `&nom_entité;`
  - o Exemple
    - Déclaration dans la DTD

```
<!ENTITY PCI "Permis de conduire informatique" >
```
    - Utilisation

```
<p>Le cours du PCI (&PCI;) se compose de...</p>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié  
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

44

## Entités générales externes

- ```
<!ENTITY nom SYSTEM "URI" >
```
- o Permet de construire un document XML à partir de plusieurs autres documents
 - o Référence : `&nom_entité;`
 - o Exemple
 - Déclaration dans la DTD

```
<!ENTITY doc SYSTEM "http://toto.org/doc.xml" >
```
 - Utilisation

```
<aide> &doc; </doc>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

45

Entités paramètres internes

- ```
<!ENTITY % nom "caractères de remplacement" >
```
- o Entités analysables uniquement utilisées dans les DTD
  - o Référence dans la DTD : `(%nom_entité;)` (parenthèses conseillées)
  - o Exemples
    - Déclarations DTD

```
<!ENTITY % tout "ANY" >
<!ENTITY % common "(para|list|table)">
```
    - Utilisations dans la DTD

```
<!ELEMENT paragraphe %tout; >
<!ELEMENT chapter ((%common;)*, section*)>
<!ELEMENT section (%common;)*>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié  
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

46

## Entités paramètres externes

- ```
<!ENTITY % nom SYSTEM "URI" >
```
- o Pour construire une DTD complexe à partir d'autres DTD complémentaires
 - o Référence dans la DTD : `%nom_entité;`
 - o Exemple
 - Déclaration dans la DTD

```
<!ENTITY % règles SYSTEM "http://toto.org/regles.dtd" >
```
 - Utilisation dans la DTD

```
%règles;
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

47

Entités analysables

- o Le texte de remplacement fait partie intégrale du document
 - Les données sont analysées correctement par le parser XML
- o Déclaration dans la DTD comme ENTITY
- o Utilisation avec `&nom;` ou `%nom;`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

48

Entités non analysables

- ```
<!ENTITY % nom SYSTEM "URI" NDATA notation >
```
- Pour déclarer un contenu non XML dans un document XML
    - Fichier image, audio, etc.
  - Référence : `&nom_entité;` uniquement comme attribut de type ENTITY
  - Exemple
    - Déclaration DTD

```
<!ENTITY photo SYSTEM "photo.tif" NDATA TIFF>
<!ELEMENT pic EMPTY>
<!ATTLIST pic name ENTITY #REQUIRED>
```
    - Utilisation dans le document XML

```
<pic name="photo" />
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié  
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

49

## Déclarations de notations

- ```
<!NOTATION nom SYSTEM "URI" >
```
- Pour
 - Identifier par un nom le format des entités non XML externes
 - Définir les formats des données et les applications qui permettent de les traiter
 - Exemple

```
<!NOTATION GIF SYSTEM "GIF" >
<!NOTATION GIF89a PUBLIC "-//Compuserve//NOTATION
Graphics Interchange format 89a//EN" >
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

50

Identificateurs publics

- ```
<!DOCTYPE mybook PUBLIC "-//EBI//DTD My book//EN" "url">
```
- PUBLIC keyword
  - Identifier type - Registered + / - / ISO
  - Owner identifier
  - Public text class
    - DTD, NOTATIONS, ENTITIES, TEXT
  - Public text description
  - Public text language
  - url non obligatoire mais conseillée
- ```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

51

Construire une DTD

- Non trivial : il faut éviter de se tromper
 - Changer une DTD XML a des conséquences sur les documents qui la suivent
- Ressemble à la création d'un schéma de base de données
- Il faut considérer
 - Le problème de la granularité
 - La questions des attributs et des éléments
 - Les limitations inhérentes aux DTD

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

52

Identifier les données qui nécessitent d'être balisées

- Pour chaque unité d'information, déterminer
 - Peut-on lui donner un nom ?
 - Apparaît-elle tout le temps ?
 - Peut-il y en avoir plusieurs ?
 - Peut-on la décomposer en des unités plus petites ?
 - Y-a-t'il du contenu textuel qui ne change pas ?
 - Comment est-elle associée aux autres unités ?

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

53

Granularité

- **<PERSON>**
<NAME>Jon Smith</NAME>
</PERSON>
- **<PERSON>**
<FORENAME>Jon</FORENAME>
<SURNAME>Smith</SURNAME>
</PERSON>

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

54

Éléments ou attributs ?

- Comment les données doivent-elles être encapsulées ?
 - `<book>`
`<title>The Forty-nine Steps</title>`
...
 - `</book>`
 - `<book title="The Forty-nine Steps">`
...
 - `</book>`
- Tout dépend de ce que l'on veut faire...
- Il existe des avis tranchés...

Éléments ou attributs ? (2)

- Séparer le contenu des métadonnées
 - Données qui doivent être imprimées comme du texte
 - Métadonnées comme attributs
- Règles générales
 - Si on enlève toutes les balises, le document doit encore être lisible et utilisable
 - S'il y a doute, utiliser un attribut

Limites des DTD / XML

- XML est seulement une syntaxe
- XML ne porte pas de sémantique
- Uniquement description de structure
- Pas de types
- Un des moyens de pallier certains problèmes
 - XML-schema

DTD

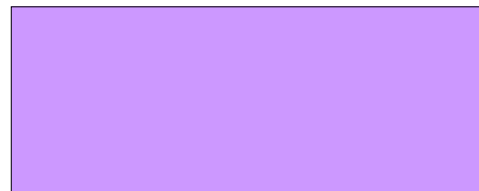
- Une syntaxe de description non-XML, héritée de SGML
 - Oblige à apprendre un langage supplémentaire
 - Ne permet pas de manipuler les DTD avec des outils XML
- Pas assez de contraintes sur les données manipulées
 - Toute donnée est une chaîne de caractères
 - Impossible de
 - spécifier des types simples
 - Entiers, dates, etc.
 - spécifier des cardinalités simples
 - Un ARTICLE aura entre 1 et 4 MOTS-CLE
 - spécifier des contraintes simples
 - Entier positif

XML-Schema

- Autre manière de spécifier des types de documents XML
- Le schéma est exprimé en XML
- Possibilité de spécifier plus de contraintes sur les données
- Possibilités avancées d'extension des schémas
- On élargit l'approche de gestion documentaire à celle plus générale de gestion de données

Exercice pratique

- Soit le document XML suivant



- En proposer une DTD

Plan

- Documents XML
 - Syntaxe XML et documents bien formés
- Types de documents XML
 - DTD et documents valides
 - Introduction à XML-Schema
- Le monde XML
 - Quelques normes liés à XML
 - Quelques DTD importantes

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

61

Standardisation

- XML permet de définir des DTD
 - modèles de documents
 - modèles de représentation de données
- Dès qu'on a un groupe, partage de données/documents
 - nécessité de partager les manières de décrire
 - accord
 - local
 - global → standardisation
- Des standards sous la forme de DTD (ou de schémas),
 - Stricts
 - Qui peuvent être raffinés
 - Les spécialiser avec des DTD internes
 - N'en utiliser que des parties

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

62

Avantages et applications XML

- Avantages
 - Réutilisabilité, partage
 - Pérennité
 - Intégrité
 - Portabilité
- Applications
 - Documents
 - Echange de données
 - Bureautique
 - Web
 - BDD semi-structurées
 - Commerce électronique
 - ...

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

63

Quelques standard XML

- The XML Bookmark Exchange Language (XBEL)
- Open eBook Publication Structure
- SportsML
- NewsML
- XML Book Industry Transaction Standards (XBITS)
- ...
- DocBook
- ebXML (electronic Business)
- Universal Description, Discovery & Integration (UDDI)
- Text Encoding and Interchange (TEI)
- XTM (XML Topic Maps)
- ...

<http://publishing.xml.org/standards/>
<http://www.oasis-open.org/specs/index.php>

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

64

Quelques spécifications XML (W3C)

- XML Schema
- XLink et XPointer
- XPath
- XSL et XSLT
- XML Query
- Namespaces
- SAX
- DOM
- MathML
- OWL
- RDF
- SMIL
- SOAP
- SVG
- XHTML

Voir <http://www.w3c.org/>

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

65

XPath

- Standard permettant d'identifier et de spécifier toutes données dans un document XML
- Exemples
 - `//toto[@name]`
→ Tous les élément toto qui ont un attribut name
 - `//tata/descendant::*`
→ Tous les descendants des éléments tata
- Voir cours Xpath

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

66

XLink

- Objectif
 - Donner la possibilité de liens riches
- XLink
 - XML Linking Specification
 - Liens
 - simples (1:1) et étendus (n:n),
 - types
 - internes ou externes
- Exemple

```
<site xlink:type="local"
      xlink:href=http://www.xml.fr/FAQ.xml
      xlink:label="fr"
      xlink:title="Version française"/>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

67

XPointer

- Objectif
 - Pointer précisément dans un document XML
- XPointer
 - XML Extended Pointer Specification
 - Une référence absolue (le document XML)
 - et une référence relative (à l'intérieur du document)
 - expression XPATH
- Exemples

```
http://www.toto.org/xml/doc.xml#xptr(/intro/title)
élément titre de doc.xml
http://site.fr/page.xml|id('ref12').child(1,session)
premier élément session enfant de l'élément
identifié par ref12, dans le document page.xml
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

68

XSL

- Ensemble d'outils permettant de
 - Visualiser les documents XML sous forme lisible, pour de multiples supports
 - Transformer les documents XML en d'autres documents XML (changement de format)
- eXtensible Style Language
 - XSL-FO (« XSL Formatting Objects »)
 - présenter des informations
 - XSLT (XSL transformation)
 - transformer un arbre XML en un autre arbre XML
 - Voir cours XSL-XSLT

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

69

Xquery

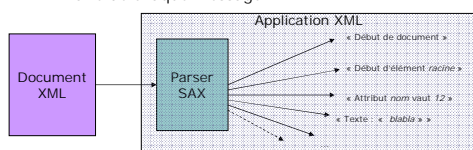
- Standard XML permettant d'exprimer des requêtes dans les documents XML
- Syntaxe XML ou non
- Utilisation de Xpath

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

70

XML et les applications : SAX

- SAX : Simple API for XML
 - Principe : un parser SAX lit un document XML, et envoie un message à une application dès qu'il
 - rencontre un début ou une fin de document
 - rencontre un début ou une fin d'élément
 - trouve des caractères dans un événement, etc.
 - Le programmeur qui écrit l'application XML décide quoi faire de chaque message

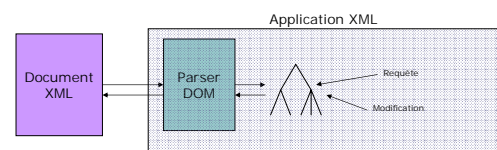


CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

71

XML et les applications : DOM

- DOM : Document Object Model
 - Principe : un parser lit un document XML et fabrique un arbre des éléments en mémoire.
 - Le programmeur qui écrit l'application XML peut alors accéder aux informations de cet arbre, les modifier, enregistrer celui-ci, etc.



CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

72

SVG

- o Objectif
 - o Description de schémas
 - o Eléments
 - o Formes (lignes, courbes, triangles, rectangles, etc.), images, textes, groupes d'éléments, ...
 - o Affichages
 - o opacités, redimensionnements, masques, ...
 - o Hypermédia
 - o liens, animations (changements de propriétés, déplacements), ...

Un exemple qui bouge !

```
<?xml version="1.0" standalone="no"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG December 1999//EN"
"http://www.w3.org/Graphics/SVG/SVG-19991203.dtd">
<svg width="300" height="300" viewBox="0 0 300 300">
<desc>Exemple anim01 - demonstrate animation elements</desc>
<rect id="RectElement" x="300" y="100" width="300" height="100" style="fill:rgb(255,255,0)"
/>
<animate attributeName="x" attributeType="XML" begin="0s" dur="9s" fill="freeze"
from="300" to="0"/>
<animate attributeName="y" attributeType="XML" begin="0s" dur="9s" fill="freeze"
from="100" to="0"/>
<animate attributeName="width" attributeType="XML" begin="0s" dur="9s" fill="freeze"
from="300" to="800"/>
<animate attributeName="height" attributeType="XML" begin="0s" dur="9s" fill="freeze"
from="100" to="300"/>
</rect>
<g transform="translate(100,100)" >
<text id="TextElement" x="0" y="0" style="font-family:Verdana; font-size:35.27;
visibility:hidden" > It's alive!
<set attributeName="visibility" attributeType="CSS" to="visible" begin="3s" dur="6s"
fill="freeze"/>
<animateMotion path="M 0 0 L 100 100" begin="3s" dur="6s" fill="freeze" />
<animateColor attributeName="fill" attributeType="CSS" from="rgb(0,0,255)"
to="rgb(128,0,0)" begin="3s" dur="6s" fill="freeze" />
<animateTransform attributeName="transform" attributeType="XML" type="rotate" from="-
30" to="0" begin="3s" dur="6s" fill="freeze" />
<animateTransform attributeName="transform" attributeType="XML" type="scale" from="1
to="2" begin="3s" dur="6s" fill="freeze" />
</text>
</g>
</svg>
```

MathML

- o Objectifs
 - Intégrer proprement des expressions mathématiques dans les pages Web
 - Permettre l'échange de formules entre logiciels mathématiques
 - Représenter la structure de présentation et la structure mathématique des formules
- o Exemples
 - $(a + b)^2$

Mathématique	Présentation
<pre><apply> <power/> <apply> <plus/> <ci>a</ci> <ci>b</ci> </apply> <cn>2</cn> </apply></pre>	<pre><msup> <mrow> <mo>(</mo> <mi>a</mi> <mo>+</mo> <mi>b</mi> <mo>)</mo> </mrow> <mn>2</mn> </msup></pre>

Espaces de noms

- o Problème
 - Deux schémas ou DTD peuvent définir des éléments qui ont le même nom
 - Exemple :
 - o DTD biblio : `<!ELEMENT name (nom,prénom) >`
 - o DTD vcard : `<!ELEMENT name (titre,prénom,nom) >`
 - Question
 - o Comment utiliser plusieurs DTD dans un unique document en évitant les collisions de noms ?
- o Solution
 - Utiliser des "espaces de nom", "espaces de nommage", "vocabulaires" (*namespaces*)

Namespaces

- o Spécification W3C
- o Principes
 - On considère qu'un schéma (DTD) définit son propre espace de nom, dans lequel tous les noms d'éléments et d'attributs sont uniques
 - On dispose d'un mécanisme pour
 - o identifier les espaces de nom utilisés dans le document,
 - o identifier pour chaque élément ou attribut à quel espace de nom il appartient.
 - Ainsi,
 - o Toute référence à un nom d'élément est non ambiguë
 - o Un document unique peut contenir des informations définies dans plusieurs espaces de nom.

Identification des namespace

- o Beaucoup de standards ont une URI officielle
 - une URI est unique
- o On peut utiliser l'URI pour identifier l'espace de nom
 - Pas forcément besoin d'un accès à Internet
 - o L'URI devient une simple chaîne de caractères identifiant un schéma
- o On "marque" les noms d'éléments et d'attributs en les préfixant avec l'URI ou un raccourci
 - `préfixe:nom`
 - Aussi appelé QName (nom qualifié)

Exemple d'utilisation

- On définit les espaces de nom avec des attributs
- Le nom de l'attribut est `xmlns`
 - On peut le spécifier n'importe où (auquel cas il est valable pour tous les sous-éléments)


```
<x:html xmlns:x="http://www.w3c.org/TR/REC-html40"
        xmlns:alan="file:/DTD/myDTD.dtd">
  <x:p>An HTML paragraph</x:p>
  <alan:p>My own special p-value markup</alan:p>
</x:html>
```
 - L'espace de nom par défaut peut être spécifié sans identificateur


```
<book xmlns="file:/DTD/myDTD.dtd"
        xmlns:="http://www.w3c.org/TR/REC-html40" >
  <x:p>An HTML paragraph</x:p>
  <p>My own special p-value markup</p>
</book>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

79

Autre exemple avec MathML

```
<?xml version="1.0"?>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>...</head>
<body>
<h1>Exemple</h1> ....
<math
  xmlns="http://www.w3.org/1998/Math/MathML">
  <mi>x</mi><mo>+</mo><mi>y</mi></math>
</body>
</html>
```

Petite démo

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

80

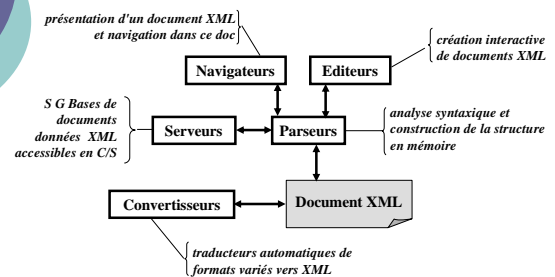
Espaces de noms et DTD

- On peut utiliser des préfixes dans les DTD
 - `<!ELEMENT document (feature, gene, sequence, collection:sequence, collection:list)*>`
- On peut inclure les définitions d'espace de nom dans les DTD
 - `<!ATTLIST document xmlns:collection #FIXED "file:/DTDs/collection.dtd">`
 - `// Implique un attribut fixé à l'élément "document",`
 - `<document xmlns:collection = "file:/DTDs/collection.dtd">`

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

81

Différents types d'outils XML



CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

82

Exemples d'outils

- Parseurs
 - SAX et DOM souvent intégrés directement dans les langages (Java, .NET, etc.)
- Éditeurs
 - XML-Spy
 - Cooktop
 - XMetal...
- Navigateurs
 - Firefox, IE, etc.
- Convertisseurs
 - Nombreux outils avec format de sortie textuel
- SGB données/documents XML
 - Évolutions des SGBD classiques
 - SGBD dédiés

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

83

Autres outils

- XHTML / CSS
 - Dreamweaver...
- XSL
 - Style-vision...
- RDF
 - Outils du web sémantique...
- SMIL
 - Player : REAL ...
- SVG
 - Adobe...

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

84

EAD / EAC

- o Objectif
 - Normalisation des instruments de recherche et des descriptions de contextes pour les archives
- o EAD
 - Encoded Archival Description
- o EAC
 - Encoded Archival Context
- o Un point d'entrée
 - <http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/DAFlangage.html>
- o Il y a des spécialistes à l'ENSSIB

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

85

Les trois éléments principaux de l'instrument de recherche en XML/EAD

- o Sous l'élément racine <ead> :

<eadheader>	en-tête EAD (description bibliographique de l'IR) (obligatoire)
<frontmatter>	préliminaires (page de titre, introduction, préface...)
<archdesc>	description archivistique (obligatoire)

Ce transparent et les suivants sur l'EAD sont extraits de « L'EAD, une DTD pour la rédaction, l'archivage et la diffusion des instruments de recherche archivistiques » F. Clavaud / 5 avril 2004

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

86

En-tête EAD <eadheader>

```
<eadheader>
  <eadid> Identifiant EAD (du fichier électronique)
  <filedesc> Description du fichier
    <titlestmt> Mention de titre et de responsabilité
    <titleproper> Titre propre de l'instrument de recherche
    <subitle> Sous-titre de l'instrument de recherche
    <author> Auteur de l'instrument de recherche
    <editionstmt> Mention d'édition
  <publicationstmt> Mention de publication
  <seriesstmt> Mention de collection
  <notestmt> Mention de note
  <profiledesc> Description du profil
  <creation> Informations sur la création de l'inventaire
  <dates> Date de l'inventaire
  <languages> Langue utilisée
  <descrules> Règles de description (archivistique utilisée)
  <revisiondesc> Descriptions des révisions ; permet de gérer les versions successives de l'instrument de recherche encodé
</eadheader>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

87

En-tête EAD <eadheader>

```
<eadheader audience="external" finalstatus="edited-partial-draft" encodinganalog="DC">
  <eadid encodinganalog="Identifier">FRDAFANCH00AP_000000001</eadid>
  <filedesc>
    <titlestmt>
      <titleproper encodinganalog="title">Etat sommaire des fonds d'archives privées du Centre historique des Archives nationales</titleproper>
      <subitle> Série 317AP à 637AP et AB XIX (AB XIX/368 à AB XIX/525)</subitle>
      <author encodinganalog="creator">Instrument de recherche rédigé par Claire Sibille, [...] encodé en XML conformément à la DTD EAD puis converti en HTML par Martin Sévigny (société AJLSM) et Florence Clavaud (service des nouvelles technologies du CHAN)</author>
    </titlestmt>
    <editionstmt>
      <edition>Quatrième édition</edition>
    </editionstmt>
    <publicationstmt>
      <publisher encodinganalog="publisher">Centre historique des Archives nationales de France (CHAN)</publisher>
      <address>
        <addressline>60 rue des Francs-Bourgeois</addressline>
        <addressline>F-75141 PARIS CEDEX 03</addressline>
      </address>
      <date encodinganalog="date">août 2003</date>
    </publicationstmt>
  </filedesc>
  <creation>
    <creation>Instrument de recherche produit sous Word et converti en XML [...] en <date>novembre 2001</date>. Encodage relu et complété par [...]</creation>
    <language>Instrument de recherche rédigé en <language langcode="fr">français</language>.</language>
  </creation>
  <revisiondesc>
    <change>
      <date>8 septembre 2003</date>
      <item>fichier converti en EAD 2002 par programme XSL-T fourni par D. Pitt (SAA)
    </change>
  </revisiondesc>
</eadheader>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

88

Préliminaires <frontmatter> : exemple

```
<frontmatter>
  <titlepage>
    <publisher>Centre historique des Archives nationales</publisher>
    <titleproper>Etat sommaire</titleproper>
    <subitle>Série AP (317AP à 637AP) et AB XIX (AB XIX/368 à AB XIX/525)</subitle>
    <author>par Claire SIBILLE</author>
    <author>avec la collaboration de Fanny AUDOUS et de Violaine LE NENAON</author>
    <publisher>Direction des Archives de France</publisher>
    <date>août 2003</date>
  </titlepage>
  <div>
    <head>Introduction</head>
    <p>Paru en 1973, l'état sommaire des fonds de la série AP (archives de personnes et de familles) rédigé par Suzanne d'Huart et Chantal Bonazzi traitait des fonds privés 1AP à 315AP. Depuis, plus de 300 nouveaux fonds d'archives privées sont entrés aux Archives nationales. [...] Ces notices sont également consultables dans la base <extref>
    href="http://sdx.archivesdefrance.culture.gouv.fr/ep/BORA</extref>
  </div>
  <div>
    <head>Mode d'emploi</head>
  </div>
  [...]
</frontmatter>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

89

Description archivistique <archdesc>

```
<archdesc level="niveau de description"> (nota : L'attribut level est obligatoire)
  <did> Identification et Description →
  <accessrestrict> et <userestrict> Conditions d'accès et Conditions d'utilisation
  <acqinfo> Acquisitions
  <acqinfo> Informations sur les modalités d'entrée
  <altformavail> Documents de substitution
  <appraisal> Informations sur l'évaluation
  <arrangement> Classement
  <biography> Bibliographie
  <bioghist> Biographie ou Histoire (du producteur)
  <controlaccess> Vocabulaires et accès contrôlés →
  <custodhist> Historique de la conservation
  <date> et <dategrp> Objet archivistique numéroté
  <note> Note
  <odd> Autres données descriptives
  <originals> Existence et lieu de conservation des documents originaux
  <otherfindaid> Autre instrument de recherche
  <physdesc> Caractéristiques matérielles et contraintes techniques
  <preference> Mention conseillée
  <processinfo> Informations sur le traitement
  <relatedmaterial> et <separatedmaterial> Documents en relation et documents séparés
  <scopecontent> Présentation du contenu
  <disc> Description des sous-composants →
</archdesc>
```

CM2-3-4 : eXtensible Markup Language – Yannick Prié
UE2.2 – Master SIB M1 – 2004-2005 : Représentation des données et des connaissances

90

Les sous-éléments de <did> Identification et description

- o <head> Titre
- o <repository> Organisme responsable de l'accès intellectuel
- o <originat> Origine
- o <unititle> Intitulé de l'unité documentaire
- o <unitdate> Date de l'unité documentaire
- o <unitid> Identifiant de l'unité documentaire
- o <physdesc> Description physique
- o <physloc> Localisation physique
- o <container> Unité de conditionnement
- o <note> Note
- o <dao> Objet archivistique numérique
- o <daogrp> Groupe d'objets archivistiques numériques
- o <abstract> Résumé
- o <langmaterial> Langue des unités documentaires
- o <materialspec> Particularités matérielles

Exemple d'utilisation de <did>

```
<archdesc level="fonds">
  <did>
    <unitid>320 AP</unitid>
    <unititle>Fonds Bazaine</unititle>
    <unitdate normal="1801/1900">XIX<emph
      render="super">e</emph> siècle</unitdate>
    <originat>Bazaine, Achille</originat>
    <physdesc>7 cartons (320AP/1 à 320AP/7),
    0,75 mètres linéaires</physdesc>
  </did>
  <!-- Etc. : les autres sous-éléments de archdesc -->
```

<archdesc> : exemple de description globale d'un fonds

```
<archdesc level="fonds">
  <did>
    <repository>Centre historique des Archives nationales (France)</repository>
    <originat>Ce fonds a été produit par <persname>Henry-Louis Duhamel du Monceau</persname>, par son frère
    <persname>Achille Duhamel de Monceau</persname>, par leur neveu <persname>Auguste-Louis Fougereux de Bondary</persname> et par le
    frère de ce dernier <persname>Fougeroux de Secval</persname></originat>
    <unitid>127AP</unitid>
    <unitdate>Fonds Duhamel du Monceau, Duhamel de Desautiers, Fougereux de Bondary et Fougeroux de Secval</unitdate>
    <unitid normal="1700/1788-1788-1788">1700-1788</unitid>
    <physdesc>4 cartons (127AP1 à 127AP4), 127AP(2), extens=1 mètre linéaire<extens></physdesc>
    <langmaterial>Langue laquée</langmaterial>
  </did>
  <acqinfo>
    <acqinfo>Achat, 7 juin 1957, 24 mars 1960, 16 juin 1977, 11 décembre 1991</acqinfo>
  </acqinfo>
  <bioghist>
    <bioghist>Henry-Louis Duhamel du Monceau (1700-1782) fit des études scientifiques au jardin du Roy, actuel jardin des plantes, en 1724.
    Associé honoraire à l'Académie des sciences en 1759, inspecteur général de la Marine en 1759, membre de l'Académie des sciences de Paris et de la
    Société royale de Londres, Duhamel du Monceau achève en 1745 un Essai sur la conservation des grains et en 1747, un Traité de la fabrication des
    mannequins pour les vêtements ou l'art de la corolle perfectionnée. On lui doit aussi [...]</bioghist>
  </bioghist>
  <scopecontent>
    <scopecontent>Commencé vers 1720 par Le Masson du Parc, commissaire ordinaire de la Marine et inspecteur général des pêches, le Traité
    général des pêches s'enrichit de la monumental projet éditorial entreprise par l'Académie des sciences sous l'impulsion de Colbert puis du Régent;
    la Description des Arts et Métiers. Après la mort de Le Masson du Parc, survient en 1741, le projet fait repris par Henry-Louis Duhamel du Monceau,
    alors inspecteur général de la Marine, qui s'attache personnellement à cette entreprise. Aidé de son neveu Fougeroux de Secval, Duhamel du Monceau en
    poursuivit la réalisation pendant une trentaine d'années, mais n'en publia qu'une petite partie (Traité général des pêches et histoire des poissons) : sa
    mort l'empêcha de la mener à son terme. Le Masson du Parc avait rassemblé une importante documentation [...]</scopecontent>
  </scopecontent>
  <accession>
    <accession>Il n'y a pas de restriction juridique à la consultation de ces documents, qui se fait selon les modalités des documents réunis en
    vigueur dans le service d'archives.</accession>
  </accession>
  <scopecontent>
    <scopecontent>Documentation scientifique plus que fonds d'archives personnelles, cette sous-série rassemble des documents réunis
    tout au long du XVIIIe siècle sur deux thèmes principaux : la pêche et l'agriculture.</scopecontent>
  </scopecontent>
  <materialspec>
    <materialspec>Description numérique établie dactylographie (127AP1 à 127AP7), par B. GIL, 1957, 19 pages, inventaire sans
    analyse dactylographie (127AP8 à 127AP19), 7 pages, Supplément dactylographie (127AP(2)), par Ariane Ducrot, 1998, 2
    pages.</materialspec>
  </materialspec>
  <materialspec>
    <materialspec>Les articles 127AP1 à 127AP19 sont conservés ensemble sous forme de microfilm.</materialspec>
  </materialspec>
  <materialspec>
    <materialspec>La documentation sur les pêches rassemblée par Le Masson puis Duhamel du Monceau est répartie dans
    différents bibliothèques : aux États-Unis (Harvard Library) et en France (Bibliothèque municipale de Lyon, Académie des sciences à Paris).</materialspec>
  </materialspec>
  <biography>
    <biography>Informations essentielles de l'introduction du répertoire numérique détaillé.</biography>
  </biography>
```

TEI

- o Text Encoding Initiative
- o Lancé en 1987
 - incompatibilité totale des formats
 - chercheurs du domaine de l'archivage structuration et analyse des textes électroniques
 - pour permettre la préparation et l'échange de textes électroniques
- o Principes
 - être aussi complet que possible,
 - être simple, clair et concret,
 - être facile à utiliser sans logiciel particulier,
 - être rigoureusement défini,
 - permettre un traitement efficace,
 - être ouvert à des extensions définies par les utilisateurs,
 - être compatible avec les standards existants ou en développement.

TEI : suite de l'histoire

- o Financé sur des programmes « Humanités »
- o Sortie de la norme (Recommandation) en mai 1994
- o Ensemble de conventions de codage utilisables dans une grande variété d'applications
 - publication électronique, analyse littéraire et historique, lexicographie, traitement automatique des langues, recherche documentaire, hypertexte, etc.
 - Textes écrits ou parlés, sans restriction de langue, de période, de genre ou de contenu
- o Répondent aux besoins fondamentaux de nombreux utilisateurs
 - lexicographes, linguistes, philologues, bibliothécaires
 - tous ceux qui sont concernés par l'archivage et l'accès à des documents électroniques.

TEI : codage

- o Construction de la DTD de façon modulaire :
 - jeu de balises «noyau» (*core tag set*) composé d'éléments communs à tous les types de textes (divisions, paragraphes, etc.)
 - des ensembles de balises de base (*base tag sets*) pour chaque type particulier de texte (prose, poésie en vers, etc.)
 - des jeux de balises additionnelles (*additional tag sets*) pour des mécanismes particuliers qui peuvent se superposer à n'importe quel type de texte (liens hypertextuels, etc.).
- o Noyau obligatoire, autre éléments facultatifs
- o Importance de l'en-tête
 - codage systématique des méta-données de n'importe quel document électronique

TEI : où en est on ?

- Au départ : SGML, puis XML
- Utilisation dans de très nombreux projets
- Reste la question du codage
 - Pour qui code-t-on ?
 - Un chercheur peut-il réutiliser le codage d'un autre ? Ajouter au codage d'un autre ?
 - Possibilité de mixer avec d'autres formats ? (particulièrement docbook)
 - Etc.

DocBook

- Objectif
 - Codage de texte de documentation
 - Sorties multi-formats
- <http://www.oasis-open.org/docbook/>
- A voir en TP
 - sdocbook : sous-ensemble de balises

Conclusion

- XML
 - Norme sortie en 1998
 - Unicode / généricité
 - Documents / données
 - Mondialement adoptée
- Standards et normes
 - Variés : dans tous les domaines nécessitant
 - Pérennité
 - Echange
 - Plus ou moins adaptés et adoptés
 - Questions récurrentes
 - Evolution
 - Interopérabilité

Remerciements

- Ce cours s'appuie largement sur celui d'Alan Robinson
<http://industry.ebi.ac.uk/~alan/XMLWorkshop/>
- Il reprend des éléments du cours CNAM de Tiphaine Accary, lui-même basé sur celui de Jean-Marie Pinon