

# XML – Introduction

EBD – 21 et 30 octobre 2013

Florent Palluault – Service du livre et de la lecture (MCC/DGMIC)

# PLAN

- Qu'est-ce que le XML ?
- Balises et structure ; éditeur XML
- Schémas et DTD
- Visualiser, publier, interroger, transformer : XPath et XSLT

Connaissez-vous le XML ?

Savez-vous le repérer ?

# QU'EST-CE QUE LE XML ?

Boîte à outils

- Imprimer
- Afficher l'adresse URL
- Contacteur l'établissement



Fin d'Ézéchiel ; Daniel ; Petits Prophètes ; Job ; Psaumes, y compris le psaume 151 (version dite Gallicana) ; Livres Sapientiaux ; Tobie ; Judith ; Esther ; Esdras ; Néhémias ; Livres III et IV d'Esdras ; Macchabées ; Actes des Apôtres ; Canoniques ; Apocalypse et les Épîtres de saint Paul, y compris l'épître aux Laodicéens.

Cote : Ms 4

Description physique : 327 feuillets ; manquent les feuillets 323, 324, 326 ; 2 colonnes.

Description :

Au XIII<sup>e</sup> siècle, on a ajouté en tête sept feuillets contenant le livre de Baruch et une seconde copie d'Ézéchiel.

Exemplaire important pour la constitution du texte du livre IV d'Esdras. On y trouve en effet (f. 207-208) un passage qui manque dans tous les manuscrits connus, sauf un de la bibliothèque d'Arras. Le manuscrit de la Mazarine est même plus complet que ce dernier, et donne quelques lignes que M. Bensly n'a pu rétablir qu'à l'aide des autres traductions de l'Écriture. (Voir Bensly, *The missing... fragment of the 4<sup>th</sup> book of Ezra*. Cambridge, 1875, in-4<sup>o</sup>.)

Autre support : A fait l'objet d'une reproduction numérique.

Bibliographie :

Sixdenier, Guy-Dominique, « Le III<sup>e</sup> livre d'Esdras et la 'Vulgate' de Stuttgart », *Revue des études anciennes*, 71, 1969, p. 390-401.

- Oeuvre :
- Bible. A. T. Ezéchiel
  - Bible. A. T. Daniel
  - Petits Prophètes
  - Job



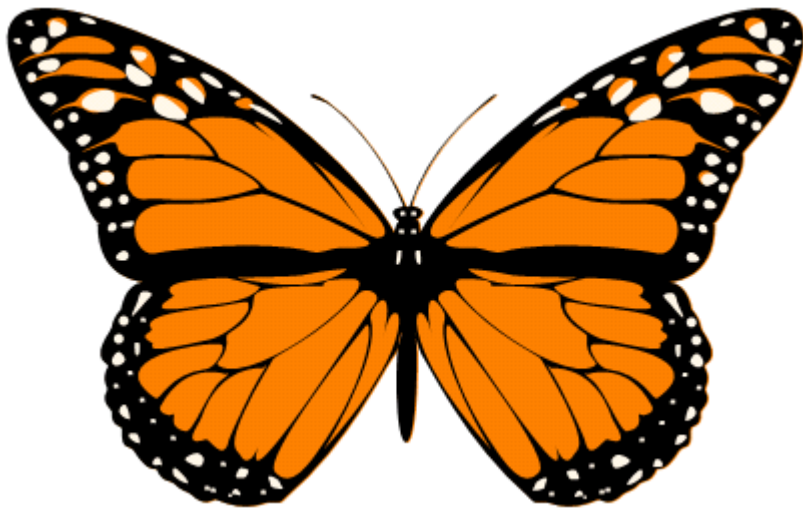
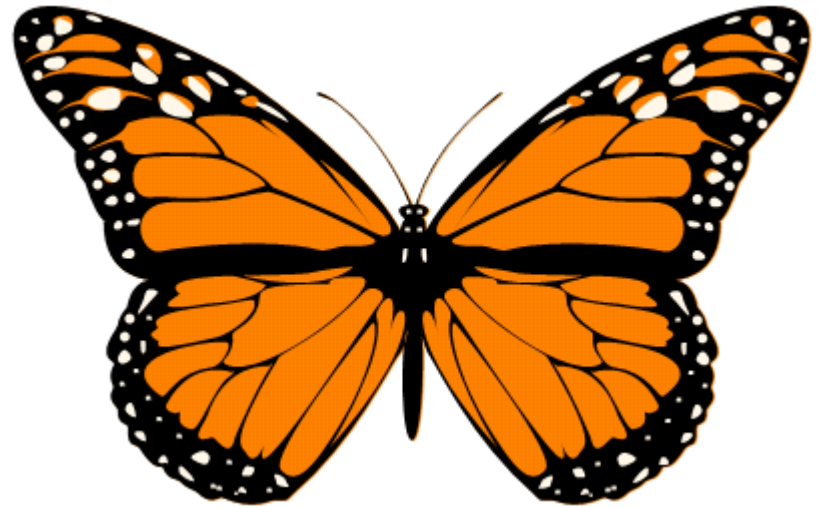
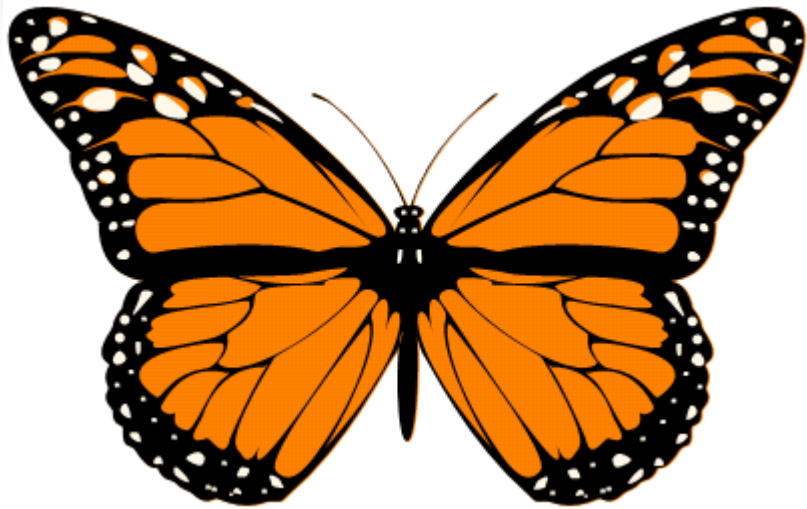
Voir en plein écran



Liber Floridus

# Le XML est largement invisible...

Source : Catalogue des manuscrits et archives de l'Enseignement supérieur  
<http://www.calames.abes.fr/pub/#details?id=MAZA10006>



## The Quadratic Formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

## Cauchy's Integral Formula

$$f(a) = \frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{z - a} dz$$

## Double angle formula for Cosines

$$\cos(\theta + \varphi) = \cos(\theta) \cos(\varphi) - \sin(\theta) \sin(\varphi)$$

## Gauss' Divergence Theorem

$$\int_D (\nabla \cdot F) dV = \int_{\partial D} F \cdot n dS$$

## Curl of a Vector Field

$$\vec{\nabla} \times \vec{F} = \left( \frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) \mathbf{i} + \left( \frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) \mathbf{j} + \left( \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \mathbf{k}$$

# Et pourtant, le XML se promène partout...



Source : Carte de migration des oies [le long du Saint-Laurent]

<https://maps.google.com/maps/ms?msa=0&msid=212893015450819139996.0004bd7fdc090bfd8861e>



MathML

$$f(a) = \frac{1}{2\pi i} \oint_{\gamma} \frac{f(z)}{z - a} dz$$



Fin d'*Ézéchiel* ; *Daniel* ; *Petits Prophètes* ; *Job* ; *Psaumes*, y compris le psaume 151 (version dite *Gallicana*) ; *Livres Sapientiaux* ; *Tobie* ; *Judith* ; *Esther* ; *Esdras* ; *Néhémias* ; *Livres III et IV d'Esdras* ; *Macchabées* ; *Actes des Apôtres* ; *Canoniques* ; *Apocalypse* et les *Épîtres de saint Paul*, y compris l'épître aux *Laodicéens*.

Cote : Ms 4

Description physique : 327 feuillets ; manquent les feuillets 323, 324, 326 ; 2 colonnes.

Description :

Au XIII<sup>e</sup> siècle, on a ajouté en tête sept feuillets contenant le livre de Baruch et une seconde copie d'Ézéchiel.

Exemplaire important pour la constitution du texte du livre IV d'Esdras. On y trouve en effet (f. 207-208) un passage qui manque dans tous les manuscrits connus, sauf un de la bibliothèque d'Arras. Le manuscrit de la Mazarine est même plus complet que ce dernier, et donne quelques lignes que M. Bensly n'a pu rétablir qu'à l'aide des autres traductions de l'Écriture. (Voir Bensly, *The missing... fragment of the 4<sup>th</sup> book of Ezra*. Cambridge, 1875, in-4<sup>o</sup>.)

Autre support : A fait l'objet d'une reproduction numérique.

Bibliographie :

Sixdenier, Guy-Dominique, « Le IIIe livre d'Esdras et la 'Vulgate' de Stuttgart », *Revue des études anciennes*, 71, 1969, p. 390-401.

Oeuvre :

- Bible. A.T. Ezéchiel
- Bible. A.T. Daniel
- Petits Prophètes
- Job
- Bible. A.T. Psaumes
- Bible. A.T. Livres sapientiaux
- Bible. A.T. Tobie
- Bible. A.T. Judith
- Bible. A.T. Esther
- Bible. A.T. Esdras-Néhémie

EAD

[Retour aux résultats](#)

[ 2 / 12 ]

Boîte à outils

- Imprimer
- Afficher l'adresse URL
- Contacter l'établissement



[Voir en plein écran](#)





# Derrière tout cela... : des balises

```
<math>
  <mrow>
    <mi>x</mi>
    <mo>=</mo>
    <mfrac>
      <mrow>
        <mrow>
          <mo>-</mo>
          <mi>b</mi>
        </mrow>
        <mo>&PlusMinus;</mo>
        <msqrt>
          <mrow>
            <msup>
              <mi>b</mi>
              <mn>2</mn>
            </msup>
            <mo>-</mo>
            <mrow>
              <mn>4</mn>
              <mo>&InvisibleTimes;</mo>
              <mi>a</mi>
              <mo>&InvisibleTimes;</mo>
              <mi>c</mi>
            </mrow>
          </mrow>
        </msqrt>
      </mrow>
    </mfrac>
  </math>
```

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <?xml-stylesheet type="text/xsl" href="http://nwda-db.wsulibs.wsu.edu/xsl/project.xsl"?>
3 <!DOCTYPE ead PUBLIC "-//ISBN 1-931666-00-8//DTD ead.dtd (Encoded Archival Description (EAD) Version 2002)//EN" "e
4 <!--Template updated 20080408 by J. Allison-Bunnell-->
5 <!--Remember to name and save documents using the OCLC code + local filename. Documents that are revised/updated a
6 <ead>
7 <!--The following section is header information that describes the finding aid-->
8 <eadheader langencoding="iso639-2b"
9   scriptencoding="iso15924" relatedencoding="dc" repositoryencoding="iso15511"
10  countryencoding="iso3166-1" dateencoding="iso8601" id="a0">
11 <eadid countrycode="us" encodinganalog="identifier"><!--Enter here the
12 unique filename for this finding aid; attributes: also enter required value from MARC code list for organizat
13 MAINAGENCYCODE and assigned NWDA ARK in IDENTIFIER. Required--></eadid>
14 <filedesc>
15 <titlestmt>
16 <titleproper encodinganalog="title">Guide to the <!--Collection title; required-->
17 <date encodinganalog="date"><!--Insert inclusive date(s), eg., 1944-1995; required--></date></titlepr
18
19 <titleproper type="filing" altrender="nodisplay"><!--Filing title of collection, eg. Adams (Glen R.) Pap
20
```

Red and blue squares [edit]



```
<?xml version="1.0"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 1.1//EN"
  "http://www.w3.org/Graphics/SVG/1.1/DTD/svg11.dtd">
<svg xmlns="http://www.w3.org/2000/svg"
  width="467" height="462">
  <rect x="80" y="60" width="250" height="250" rx="20"
    style="fill:#ff0000; stroke:#000000;stroke-width:2px;" />
  <rect x="140" y="120" width="250" height="250" rx="40"
    style="fill:#0000ff; stroke:#000000; stroke-width:2px;
    fill-opacity:0.7;" />
</svg>
```

Source : </oXygen>  
Création d'un fichier EAD

Source : Wikipedia  
<http://fr.wikipedia.org/wiki/MathML>

Source : Wikimedia Commons  
[http://commons.wikimedia.org/wiki/SVG\\_examples](http://commons.wikimedia.org/wiki/SVG_examples)

```
1 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
2 <html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr">
3 <head>
4 <title>Ecole Bibliothécaires Documentalistes - EBD</title>
5 <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
6 <meta http-equiv="imagetoolbar" content="no" />
7 <meta name="author" content="" />
8
9 <meta name="Robots" content="index, follow" />
10
11
12
13 <meta name="revisit-after" content="7 days" />
14 <meta name="description" content="EBD, Ecole de Bibliothécaire Documentaliste. Formation supérieure aux métiers de la documentation et des bibliothèques. Étab
    technique supérieur, Paris, France" />
15 <meta name="keywords" content="Documentaliste, Bibliothécaire, Documentation, Bibliothèque, formation professionnelle, traitement de l'information, veille, mé
    iconographie, iconographe, photothèque, photothécaire, archives, documentation audiovisuelle, gestion " />
16 <meta name="copyright" content="" />
17 <meta name="vs_defaultClientScript" content="JavaScript" />
18 <meta http-equiv="Content-Style-Type" content="text/css" />
19 <link rel="stylesheet" href="e_habillage/css/styles.php" type="text/css" />
20 <link href="e_habillage/css/print.css" media="print" rel="stylesheet" type="text/css" />
21 <script type="text/javascript" src="e_config/swfObject/swfobject.js"></script>
22
23 <!--moteur de recherche google. -->
24 <script>
25 (function() {
26     var cx = '015394576722351416745:q3-y13p17ui';
27     var gcse = document.createElement('script'); gcse.type = 'text/javascript'; gcse.async = true;
28     gcse.src = (document.location.protocol == 'https:' ? 'https:' : 'http:') +
29         '//www.google.fr/cse/cse.js?cx=' + cx;
30     var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(gcse, s);
31 })();
32 </script>
33
34 </head>
35
36 <body class="home" >
37 <div id="container">
38 <div id="skip-top"><ul><li><a href="#skipnav">Accéder directement au contenu.</a></li><li><a href="#nav">Accéder directement au menu principal.</a></li></ul></div>
39 <div id="header">
40 <h1>Page d'accueil</h1><div id="logo"><a href="index.php" accesskey="a">
42 <div id="accessibility-menu">
43 <h2>Menu d'accès rapide :</h2><ul><li><div id="sitemap"><a href="ewb_pages/p/plan-site.php" title="Plan du site." accesskey="p">Plan du site</a></div></li><li>
    only></gcse:searchbox-only></li></ul>
44 </div>
45 <!-- FIN DE LA DIV ACCESSIBILITY --><div id="slogan"></div>
46 <!-- MENU PRINCIPAL --><div id="menu-coin-g"><div id="menu-coin-dr"><div id="menu-principal"><a name="nav"></a>
```

# Qu'est-ce que le XML ?

- XML = eXtensible Markup Language
  - langage de balisage extensible
- Etape très importante dans le processus de normalisation / uniformisation des langages informatiques et dans le développement du web : création d'un éco-système de technologies permettant la création, la structuration et l'utilisation de langages de balisage
  - Structuration des données
  - Requête / manipulation / transformation
  - Feuilles de style pour multi- diffusion
  - Etc.

# Bref historique

- 1983 : Norme américaine SGML (Standard Generalized Markup Language)
  - Langage de balisage générique, issu entre autres du GML (IBM)
  - Devient norme ISO en 1986
- 1991-1992 : apparition du HTML
  - Créé à partir du SGML, mais beaucoup plus simple, afin de pouvoir être lu par des navigateurs web
  - Conçu spécifiquement pour produire des contenus hypertexte : naissance des liens
  - Evolution progressive :
    - 1995-1996 : HTML 2.0
    - 1997 : HTML 4.0
    - 2008-2012 : HTML5 (avec support natif de XML)

# Naissance du XML

- XML 1.0 introduit en 1998 par le [W3C](#)
- Objectifs : allier le caractère générique et la souplesse du SGML avec la simplicité du HTML
  - Souplesse et créativité : le XML ne définit pas d'éléments de balisage mais permet à l'utilisateur d'en créer
  - Sûreté et fiabilité : la syntaxe doit se conformer à des règles strictes, ce qui empêche tout degré "d'interprétation" par les machines et logiciels
  - Caractère intuitif et simple : facile à maîtriser
  - Versatilité : la mise en forme est traitée par programmes externes
  - Ouverture : langage indépendant des plateformes, systèmes d'exploitation ou logiciels particuliers, propice à l'échange de données, et supportant une très large palette d'écritures, de caractères et de symboles (Unicode UTF-8 par défaut)

# Philosophie du XML

- XML est vu avant tout comme
  - Un protocole de stockage et de gestion de l'information
  - Une famille de technologies associées permettant de multiples exploitations de ces informations
- Mais c'est aussi une

*"Philosophie de la gestion de l'information qui recherche un maximum d'utilité et de souplesse en organisant les données sous la forme la plus pure et la mieux structurée"*

(Erik T. Ray, *Introduction à XML*. Paris : O'Reilly, 2001)

# XML vs. bases de données

- Bases de données relationnelles : système de stockage de données sous forme de lignes dans des tables multiples
- Avantages des bases de données relationnelles
  - Meilleure gestion des gros volumes de données
  - Meilleure gestion des gros volumes de transactions entre systèmes
  - Système largement éprouvé
- Avantages de XML
  - Contient à la fois les données et la structure
  - Données arborescentes : il est possible d'emboîter des notions sémantiques
  - Auto-documenté



Il existe également des bases de données en XML

# XML vs. HTML

## HTML

- Éléments prédéfinis
- Conçu pour de la présentation à l'écran
- Aucune validation des données
- Une seule sortie / présentation possible
- Syntaxe permissive

## XML

- Jeu d'éléments extensible
- Centré sur le contenu plutôt que sur la présentation
- Structuration standardisée et validation possible
- Multiples sorties / présentations possibles
- Syntaxe stricte



# BALISES ET STRUCTURE

# Exemple

Echange sur un logiciel de messagerie instantanée

- 21:03 - **Roméo** : Salut ! Veux-tu aller voir l'exposition Astérix à la BnF ce week-end ?
- 21:04 – **Juliette** : Impossible, il faut que je finisse les *Fleurs du Mal* pour le milieu de la semaine prochaine. Je suis déjà en retard sur mon programme de travail, et j'ai promis à Claire de travailler avec elle samedi et dimanche.

# Exemple (suite)

Cet échange peut être stocké de la manière suivante

```
<conversation>
```

```
  <message source="456795" nomsource="Roméo" date="2013-10-20" heure="21:03">Salut ! Veux-tu aller voir l'exposition Astérix à la BnF ce week-end ?</message>
```

```
  <message source="117845" nomsource="Juliette" date="2013-10-20" heure="21:04">Impossible, il faut que je finisse les <affichage type="italiques">Fleurs du Mal</affichage> pour le milieu de la semaine prochaine. Je suis déjà en retard sur mon programme de travail, et j'ai promis à Claire de travailler avec elle samedi et dimanche.</message>
```

```
</conversation>
```

# Exemple (suite)

Ou bien de la manière suivante :

```
<sequence acte="2" scene="1">
  <intervention>
    <heure>21:03</heure>
    <personnage nom="Roméo">
      <texte>Salut ! Veux-tu aller voir l'exposition Astérix à la BnF ce week-end
?</texte>
    </personnage>
  </intervention>
  <intervention>
    <heure>21:04</heure>
    <personnage nom="Juliette">
      <texte>Impossible, il faut que je finisse les <i>Fleurs du Mal</i> pour le
milieu de la semaine prochaine. Je suis déjà en retard sur mon programme de
travail, et j'ai promis à Claire de travailler avec elle samedi et dimanche.</texte>
    </personnage>
  </intervention>
</sequence>
```

# Éléments XML

Un élément XML est constitué d'une balise ouvrante et d'une balise fermante qui encadrent son contenu...

```
<annee>1959</annee>
```

... ou d'autres éléments (emboitement)

```
<auteurs>
```

```
  <scenariste>René Goscinny</scenariste>
```

```
  <dessinateur>Albert Uderzo</dessinateur>
```

```
</auteurs>
```

Règles d'emboitement :

- La balise de fin doit arriver après la balise de début
- Les balises de début et de fin doivent être présentes au sein du même élément parent
- Entre une balise ouvrante et une balise fermante, il ne peut y avoir une unique balise (soit ouvrante, soit fermante)
  - **Incorrect** : <publication>Paris, <annee>1959</publication>.</annee>

# Éléments XML

Un élément peut également accepter un **contenu mixte** : texte et autres éléments

```
<auteurs>Les auteurs d'<titre>Astérix le Gaulois</titre> sont  
<scenariste>René Goscinny</scenariste> et <dessinateur>Albert  
Uderzo</dessinateur>.</auteurs>
```



Les éléments XML sont sensibles à la casse

- `<annee>1959</annee>`
  - `<Annee>1959</Annee>`
  - `<ANNEE>1959</ANNEE>`
- } 3 éléments différents

Le contenu textuel est appelé PCDATA (Parsed character data)

# Éléments XML

## Obligations de nommage pour les éléments

- Doit commencer par une lettre ou un \_
- Peut contenir des caractères accentués de l'alphabet latin, ainsi que des caractères d'autres alphabets

## Recommandations de nommage :

- Nom de l'élément en minuscules : `<date>1959</date>`
- Si le nom de l'élément est composé de plusieurs mots, tous les mots sont attachés et leur première lettre est en majuscule sauf pour le premier : `<dateDePublication>1959</dateDePublication>`
- Pas d'utilisation des accents



Un élément peut être vide. Exemple : `<lb></lb>`  
également écrit : `<lb/>`

# Attributs en XML

Les attributs éventuels d'un élément qualifient le contenu d'une balise

```
<oeuvre type="poésie">Les Fleurs du mal</oeuvre>
```

```
<oeuvre type="roman">Les Misérables</oeuvre>
```

Règles syntaxiques :

```
<nomElement attribut1="valeur1" attribut2="valeur2">contenu  
de l'élément</nomElement>
```

- Un élément peut avoir un nombre quelconque d'attributs
- Au sein d'un élément donné, chaque attribut a un nom unique ; autrement dit, un attribut n'y est pas répétable

 démonstration



# Eléments ou attributs : que choisir ?

Création d'un document XML: l'utilisation d'éléments ou d'attributs pour porter l'information dépend du modèle de données

```
<oeuvre type="poésie">Les Fleurs du mal</oeuvre>
```

```
<oeuvre type="roman">Les Misérables</oeuvre>
```

ou

```
<poesie publication="1857">Les Fleurs du mal</poesie>
```

```
<roman publication="1862">Les Misérables</roman>
```

Eléments :

- Contenu potentiellement supérieur à plusieurs mots
- A priori affiché à l'écran

Attributs :

- Information non destinée à être affichée telle quelle
- Information servant à faire des tris et filtres informatiques
- Possible d'avoir une liste fermée de valeurs

# Entités

**Entités caractères** prédéfinies : permettent de rendre des caractères qui sont réservés dans la syntaxe XML

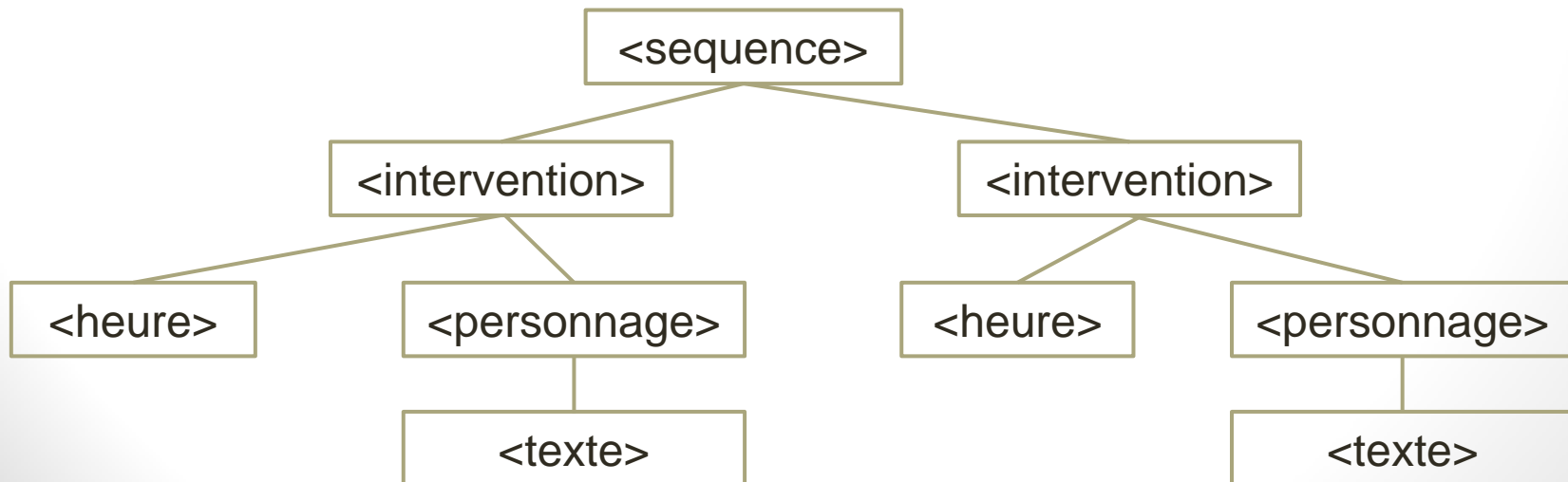
- & : &amp; | ' : &apos; | > : &gt; | < : &lt; | " : &quot;

**Entités de contenu mixte** : permettent de déclarer de l'information réutilisable dans le document

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE lemien SYSTEM "madtd.dtd"
[<!ENTITY URLsite "http://www.ebd.fr">
<!ENTITY nomEcole "Ecole de bibliothécaires documentalistes">]>
<texte>
  <message>Retrouvez toute l'information sur les formations de l'&nomEcole; sur notre
site à l'adresse &URLsite;</message>
  <autreMessage>Vous trouverez d'ailleurs plein d'autres choses à propos de
l'&nomEcole; à &URLsite;</autreMessage>
</texte>
```

# Arborescence

- Une séquence XML (et a fortiori, un document XML) forme une **structure arborescente**
  - Un élément racine
  - Les éléments de premier niveau sous cette racine, puis de 2<sup>e</sup> niveau, etc.
  - Les éléments de dernier niveau ("feuilles") : contiennent du PCDATA ou sont vides



# Document XML "bien formé"

- Caractéristiques nécessaires
  - Existence d'un prologue qui indique qu'il s'agit d'XML
    - `<?xml version="1.0" encoding="utf-8"?>`
  - Un élément racine englobe tous les autres
  - Arborescence correcte et imbrication logique des balises
- Le caractère bien formé peut être vérifié dans un éditeur XML
- Nommage : un fichier XML comporte une extension .xml

# Ex. de document XML bien formé

```
<?xml version="1.0" encoding="utf-8"?>
<notice>
  <cote type="actuelle">Mss. M. 1001-2800</cote>
  <cote type="ancienne">Mont. 1-1864</cote>
  <intitule>Fonds Montesquieu</intitule>
  <date>1201-1900</date>
  <descriptionPhysique>
    <importanceMaterielle>1800 manuscrits</importanceMaterielle>
  </descriptionPhysique>
  <producteur>
    <nomPersonne normal="Montesquieu, Charles-Louis de Secondat (1689-1755 ;
baron de La Brède et de)">Montesquieu </nomPersonne>
  </producteur>
  <!-- Ceci est un commentaire, qui n'est pas parsé par le validateur -->
  <acquisition>Achats, dons, legs et dation</acquisition>
  <acces>Consultation dans la salle de la Réserve uniquement</acces>
  <note> Fonds consacré à Montesquieu : livres et manuscrits du philosophe,
archives de la seigneurie de La Brède dont les plus anciennes pièces remontent au
XIIIe siècle.</note>
</notice>
```

# L'éditeur XML

- C'est un logiciel incontournable pour tout utilisateur de données XML car il permet
  - de créer des documents XML et de les modifier
  - d'en vérifier la structure et validité : indication d'éventuelles erreurs à corriger
  - d'y effectuer des recherches et des requêtes
  - d'effectuer des manipulations sur plusieurs documents à la fois
  - éventuellement de les visualiser sous une autre forme
- Il existe de multiples éditeurs, dont certains gratuits
  - Plus ou moins sophistiqués
  - Plus ou moins ergonomiques

 démonstration du logiciel </oxygen>

# XML en résumé

- Structure
  - Une balise contient de l'information textuelle et/ou d'autres balises
  - Chaque balise caractérise l'information
  - Chaque attribut caractérise la balise
  - **Héritage implicite** : les éléments « fils » héritent des informations de leur élément « père »
  - Liens faciles à créer entre éléments d'un même document ou vers l'extérieur
- Affichage multiple possible
  - mise en forme assurée par une feuille de style externe (CSS, XSL)
  - consultation statique ou dynamique
- Possibilité de créer ses propres éléments :
  - Intitulés significatifs pour l'utilisateur
  - structuration des éléments en fonction des objectifs



**Inconvénient majeur : risques pour l'interopérabilité**



**Exemple**

181 entrée vacante

LEGRAND (Jacques)

182 Sophologium. — Lyon: Nicolaus Philippi et Marcus Reinhart, [circa 1480]. — 2°.

120 fnc. n. sig. [a<sup>10</sup> b-g<sup>8</sup> h i<sup>10</sup> k-n<sup>8</sup> o<sup>10</sup>], 2 col., 46 ll.; justif. 167 × 124 mm — Car. G 121 (explicit et colophon), G 72 (texte). BMC VIII 241 et pl. xxxv, xxxvi.

Fnc. [a<sub>2</sub>]<sup>ro</sup>, col. 1, incipit: Sequitur tabula capitulorum fophologij: Et primo capitula primi libri || Capim p̄mū quō sapiā facit felices || Secūdū... — fnc. [a<sub>3</sub>]<sup>ro</sup>, l. 44:

Decimum septimum de viduis || Finit tabula huius libri || — fnc. [a<sub>3</sub>]<sup>ro</sup>, blanc.

— fnc. [a<sub>4</sub>]<sup>ro</sup>, col. 1, incipit: Doctissimi atq̄ excellentissimi patris tris sacrae litterarum doctoris de||uotissimi fratris Jacobi magni || religionis fratru heremitaru san||cti Augustini sophologium incipit || Cuius p̄ncipalis intentio est inducere legētim animū ad sapiē amozē || SERENISSIMI atq̄ christi||anissimi p̄ncipis francoz regis cons||effo||i d̄ Michaeli diuina p̄uiden||te gracia episcopo antiodoēnsi sui || p̄conij humilis capellan<sup>9</sup> frater iaco||bus magni ordinis fratrum heremi||tarum...

l. 30: Capitulum primum qualiter sapien||tia reddit homines felices. || — col. 2: [7]

Icit. Aristo. x. ethi||coz/ q̄ hō sapiens || maxime felix est || Et sene ad lucil. || Beata (inquit) vi||tā sapiā p̄fecta ef||ficat Et sequitur || Scio neminem... — fnc. [b<sub>1</sub>]<sup>ro</sup>,

incipit: docent/ legunt vel student. Nam (tes||te ouidio .ij. de arte amandi)... — fnc. [o<sub>5</sub>]<sup>ro</sup>, l. 10: est z suscipe quod p̄ se bonum non est || scz nubere z loquitur vidue. || Jacobi magni sopholo||giū finit feliciter || — fnc. [o<sub>10</sub>]<sup>ro</sup>: Epigramma ad huius opis conspectorem. || [5 distiques:] Istuc claroz cōtendunt dogmata patrum! — l. 11:

Mille voluminibus clauditur hoc opere. || Imp̄essum lugduñ per Nicolau Philippi de || benszheym et Marcū Reinhart de Argentina. || — v° blanc.

HC 10476. Pell 7105-7107 (7051<sup>e</sup>-i). IGI 5961. Goff M-46. IBE 3665. Pr 8527A. Sheppard 6579-6580. Claudin III p. 123. Klebs 595,10.

Mouliens BM Inc 4° 28092. (Prou. [fnc. 2<sup>ro</sup>] "Hunc lib. dederunt DD. de Cigogne et Desmarests ff. Augustin. S<sup>4</sup> Petri a Mon. 1639"; [fnc. 1<sup>ro</sup>, bl.] "Cœnobii Annæi PP. Augustinianorum S<sup>4</sup> Petri a Monasterio c.B. Scup. s.g." (XVII<sup>e</sup> s.). — Ment. Foliotation et titre cour. mss d'une main du XV<sup>e</sup> s.; notes mss, XV<sup>e</sup>-XVI<sup>e</sup> s.; [contreplat sup.] anciennes cotes de la bibliothèque des Augustins de Saint-Pierre-le-Moutier "q3" [fnc. 2<sup>ro</sup>] "J9"; cote ancienne 7571. Ripoud 135. — Rel. Ais couverts de basane fauve estampée de filets à froid (par 6) déterminant quatre losanges et 8 triangles muets; restes de deux fermoirs dont subsistent sur le plat sup. les départs de lanière à 3 clous et sur le plat inf. les 2 contre-agraves à gouttière; dos à 4 nerfs, XV<sup>e</sup> s.)

LEGRAND (Jacques)

Lyon: Guillaume Balsarin, circa 1488]. — 4°.

(7043). Goff M-47. IDL 3041 (Johann Neumeister). IBE 3668. CIBN

(Incompl. des f. sig. a<sub>1</sub>, [a<sub>8</sub>], [B<sub>8</sub>] et [E<sub>8</sub>]. — Ment. [f. sig. a<sub>2</sub>]<sup>ro</sup> Notice bibliographique avec référence au cat. La Vallière, XVIII<sup>e</sup>; notes mss; ancienne cote 141. — Rubr. Seuls les cahiers a-1 sont rubriqués; titre courant ms. donnant le numéro des livres, XV<sup>e</sup> s. — Rel. Ais couverts de basane; un fermoir central dont le départ de lanière et la contre-agrave trapézoïdale à gouttière sur le plat inf. sont encore en place; les gardes sont constituées

Défi : comment exprimer le mieux possible en XML la notice descriptive d'un incunable ?

ANTTIUS (Lucius Coelius Firmianus)

Opera : De Diuinis institutionibus. De Ira Dei. De Opificio Dei vel de formatione hominis. De Phœnice carmen. Epitome diuinarum institutionum. cap. lvi-lxxxiii. Éd. Johannes Andreas de Buxis. — FORTUNATUS (Venantius Honorius Clementianus). De Die resurrectionis. — Venezia: Vincenzo Benali, III 1493. — 2°.

9816\*. Polain 2423. IGI 5628. Goff L-11. BS 3342. Sajó-Soltész 2021. IDL 2871. IBE 9. BMC V 525. CIBN L-11. BSB-Ink L-10.

BM Inc 10 (1) (Rec. R3/1. — Incompl. du fnc. sig. [a<sub>1</sub>], titre. — Table reliée à la fin. Notes mss. — Rel. Dérelé, sans couverture, dos à 4 nerfs.)

ANTTIUS (Placidus). Comment. STATIUS (Publius Papinus): 257.

ANTTIUS (Theodorus). Éd. HIERONYMUS, S.: 155.

ANTTIUS (Christophorus). Comment. HODARTIUS

ANTTIUS astronomicus, in JOHANNES de SACRO



Mieux contrôler et échanger

# SCHÉMAS ET DTD

# Solution : DTD et schémas

**DTD** : définition de type de document

- C'est à la fois un **vocabulaire** :
  - définition des éléments et attributs autorisés
  - mention du caractère obligatoire ou facultatif, répétable ou unique
- ...et une **grammaire** : endroit et ordre d'utilisation des éléments et attributs

Evolution des DTD vers des **Schémas XML**  avantages

- Création en XML, avec utilisation possible d'attributs XML génériques
- Davantage de contrôle des données saisies
- Auto-documentation du schéma
- Peut être associé à d'autres schémas.

# Document XML valide

**Validation** : le fichier est **valide** s'il est conforme à la DTD ou au schéma déclaré

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<!DOCTYPE ead PUBLIC "-//ISBN 1-931666-00-8//DTD ead.dtd (Encoded Archival Description (EAD) Version 2002)//EN" "ead.dtd">
```

- La validité est vérifiée par un parseur intégré à l'éditeur ou au navigateur
- Le fichier DTD ou le schéma de référence peuvent être en ligne ou en local.
- Si le fichier est déclaré en local, il faut qu'il soit présent dans le répertoire indiqué pour que la validation soit effectuée

# Espaces de noms

Permettent d'utiliser plusieurs vocabulaires dans un même doc. XML, grâce à une déclaration spécifique

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<bibliographie>
```

```
  <noticeArch xmlns:noticeArch="http://www.example.com">
```

```
    <noticeArch:cote>Mss. M. 1001-2800</noticeArch:cote>
```

```
    <noticeArch:intitule>Fonds Montesquieu</noticeArch:intitule>
```

```
    <noticeArch:descPhysique>1800 manuscrits</noticeArch:descPhysique>
```

```
    <noticeArch:note>Manuscrits et correspondance du  
philosophe</noticeArch:note>
```

```
  </noticeArch>
```

```
  <noticeLivre xmlns:noticeLivre="http://www.example2.com">
```

```
    <noticeLivre:auteur>Montesquieu, Charles-Louis de Secondat (1689-1755  
; baron de La Brède et de)</noticeLivre:auteur>
```

```
    <noticeLivre:titre>Lettres persanes</noticeLivre:titre>
```

```
  <noticeLivre:dateDePublication>1721</noticeLivre:dateDePublication>
```

```
  </noticeLivre>
```

```
</bibliographie>
```

# Quelques DTD / schémas utilisés en documentation et bibliothèques

Avantage de XML dans le domaine de la documentation :  
**alliance de la structuration et de l'intégralité du texte**

- EAD : Description archivistique encodée
  - Cf. <http://www.bonnespratiques-ead.net/>
- MarcXchange : schéma de données pour l'Unimarc
  - MarcXML : idem pour le Marc21
- MODS : schéma simplifié de description bibliographique
- MARC-TEI : schéma de description des incunables
- ONIX : schéma de données du domaine de l'édition
- TEI (Text Encoding Initiative) : ensemble de recommandations pour l'encodage de textes (discours, pièces de théâtre, etc.)
  - TEI-MS : sous-ensemble TEI pour la description des manuscrits (en particulier les manuscrits médiévaux)



# Liens

## Exemples de construction de liens en XML

- XHTML : `<a href="files/cons/Charte_de_la_conservation.pdf" class="lien">Charte de la conservation dans les bibliothèques</a>`
- EAD : `<separatedmaterial><p>Une partie des papiers de Georges Cuvier (<archref href="http://www.calames.abes.fr/pub/#details?id=PA2011342" actuate="onrequest" show="new">oeuvres scientifiques</archref>) a été donnée à la Bibliothèque du Muséum. Cf. <bibref href="http://gallica.bnf.fr/ark:/12148/bpt6k4081g568h" actuate="onrequest" show="new" >G. Dupont, Vie et œuvre de Georges Cuvier</bibref></p></separatedmaterial>`
- Liens internes
  - Utilisation d'un attribut de type ID, dont la valeur doit être unique dans tout le document XML
  - `<employe id="#0123456">Marina</employe>`  
....  
`<employe id="0789654" binome="#0123456">Arthur</employe>`

VISUALISER, PUBLIER

# Visualisation de documents XML

- Via l'éditeur XML
  - Vision du code
  - Avec ou sans balises
- Via un navigateur
  - Données brutes : contenu des éléments sans mise en forme  
     Exemple
  - Données XML mises en forme par utilisation d'une feuille CSS  
     Exemple **(ajout d'une instruction)**
- Pour rappel : CSS = feuilles de style en cascade
  - Langage très répandu, servant à décrire la présentation de document HTML et XML



# Publication de documents XML

- Statique : aisée, rapide et peu chère
  - Transformation de XML en HTML ou PDF via des feuilles de style CSS ou de transformation (XSLT, XSL-FO)
  - Instruction `<xsl:stylesheet>` en début de fichier

➡ Exemple

Complément éventuel avec CSS et JavaScript

➡ Exemple

- Dynamique : permet indexation, recherches, rebonds
  - Solution logicielle dédiée à la mise en ligne de DTD ou de schémas particuliers
  - Développements propres à des projets spécifiques

# Créer des documents XML

- Par création directe dans un éditeur XML
- Par création automatique à partir de données hébergées dans des logiciels de bureautique
  - Word / OpenOffice Writer : export possible en XML
  - Excel / OpenOffice Calc : idem
- Par export de bases de données (MySQL, etc.) dans un format XML

# LOCALISER ET MANIPULER LE XML

# Localiser et sélectionner des données XML

**XPath** : langage incontournable pour rechercher efficacement dans un document XML

- Langage non XML d'interrogation et de localisation de portions de documents XML à la syntaxe assez simple et robuste
- Utilise des pas (étapes) de localisation, absolus ou relatifs, afin de créer un chemin de localisation
- Peut utiliser des tests et des prédicats afin de sélectionner des nœuds XML répondant à certaines conditions
- Les éditeurs XML possèdent désormais tous un module XPath

**XQuery** : langage de requête informatique permettant des extractions d'informations d'un document XML

- Equivalent du SQL pour les bases de données relationnelles

# XPath : notions de syntaxe

/	Nœud document
/*	Élément racine
*/elem	Tout élément <elem> sous-élément immédiat de l'élément racine
..	Remonte au nœud parent
.	Nœud courant
*	Tout nœud élément
@	Nœud attribut
//elem	Tous les éléments <elem> du document
//oeuvre[@type="roman"]	Tous les éléments <oeuvre> dont la valeur de l'attribut type est "roman"

## Quelques fonctions utilisées dans les prédicats :

contains(chaine,argument) : validation si chaîne présente dans le contenu

starts-with(chaine,argument) : validation si le début de la chaîne est présent

substring(chaine,position de départ,longueur) : renvoie la chaîne commençant à la position de départ et de la longueur spécifiée

//chapitre/paragraphe[2][contains(.,'longtemps')] Tous les éléments <chapitre> dont le 2<sup>e</sup> élément <paragraphe> contient la chaîne de caractères 'longtemps'.



**Exemples d'utilisation**

# Manipulation des données XML

- Manipulation des données :
  - extractions ciblées
  - ajouts ciblés
  - Tous types de manipulations
- En particulier : création d'index
  - Eviter le magma des recherches en texte intégral
  - Exploiter la sémantique de l'XML
  - Exploiter la valeur des attributs par usage de référentiels, thésaurus, listes contrôlées, etc.
- Création de métadonnées (par exemple, Dublin Core)
  - Par correspondance automatique de balises

# Exemple : de l'EAD au Dublin Core

- Obtention de métadonnées Dublin Core par mapping à partir de notices descriptives en EAD : ex [Gallica](#)

EAD	Dublin Core
<unittitle>	dc.title
<unitdate>	dc.date
<unitid>	dc.identifiant
<langmaterial>	dc.language
<scopecontent>	dc.description
<controlaccess> <...>	dc.subject

- Export OAI-PMH : métadonnées en Dublin Core éventuellement accompagnées d'EAD

# XSLT

## Extensible Stylesheet Language for Transformation

- Langage (lui-même XML) de transformation XML, permettant notamment de transformer un document dans un autre format (PDF, XHTML) pour affichage sur le web
- Parcourt l'arborescence du document et exécute des modèles d'actions prévus par le programmeur (assez différent en cela d'un langage de programmation)
- La sélection des nœuds se fait grâce au langage XPath
- Quelques éléments XSLT
  - `<xsl:template match="elementcible">`
  - `<xsl:apply-templates>`
  - `<xsl:value-of>` : donne la valeur d'un noeud cible
- Exemples : [XSLT](#)



# Conclusion : les raisons du succès

- Standard ouvert, indépendant de plateformes informatiques, systèmes d'exploitation ou logiciels particuliers
- Stockage et organisation de n'importe quel type d'informations
- Gestion par défaut de l'Unicode : jeu de caractères le plus développé jusqu'à présent
- Légèreté et clarté de la syntaxe
- Système de validation, permettant des vérifications multiples sur la qualité et le contenu du document
- Facilité de la diffusion par divers moyens
- Intégration dans le web

# Merci de votre attention

- Bibliographie : Erik T. Ray, *Introduction à XML*. Paris : O'Reilly, 2001.

Florent Palluault

Ministère de la Culture et de la Communication

Service du livre et de la lecture, Département des Bibliothèques

[florent.palluault@culture.gouv.fr](mailto:florent.palluault@culture.gouv.fr)