

Schémas XML

LIONEL MÉDINI
UFR INFORMATIQUE
UNIVERSITÉ CLAUDE BERNARD LYON 1

D'après le cours de Yannick Prié
2010-2011 – Master SIB
M1 – UE 3 / Bloc 2 – Cours 1

Objectifs de ce cours

2

- Etre capable de
 - Comprendre la problématique liée aux espaces de noms XML
 - Comprendre et créer des schémas XML
 - Découvrir quelques langages basés sur XML

Plan

3

- **Espaces de noms XML**
 - URI, URL et URN
 - Position du problème
 - Principes de fonctionnement
- **Schémas XML**
 - Différences par rapport aux DTD
 - Introduction à XML-Schema
- **Le monde XML**
 - Quelques normes liées à XML
 - Quelques DTD / schémas importants

URI, URL et URN

4

- **URI : Uniform Resource Identifier**
 - But : identifier de façon unique une ressource sur le web
 - » En disant où elle se trouve
 - Donner son URL (Uniform Resource Locator)
 - Format : protocole ":" chemin "/" nom de fichier "/" requête
 - `http://www.w3.org/2001/XMLSchema`
 - Permet d'accéder réellement à la ressource (tant qu'elle existe)
 - Enregistrement des DNS auprès de l'entité concernée
 - » En disant comment elle s'appelle
 - Donner son URN (Uniform Resource Name)
 - Format : "URN:" NID (namespace identifier) ":" NSS (namespace specific string)
 - URN : ISBN : 0-395-36341-1
 - Choix plus « libre », et correspondant mieux à la définition d'un espace de noms
 - Enregistrement des NID à l'IANA (Internet Assigned Numbers Authority)
 - Syntaxe générique

URI, URL et URN

5

- **URI : Uniform Resource Identifier**
 - But : identifier de façon unique une ressource sur le web
 - Syntaxe générique
 - » « scheme » ":" autorité ":" chemin ":" requête ":" fragment
 - » Avec le temps, on s'est mis à penser que « urn » peut aussi être un URI scheme
- D'un point de vue pratique, les URL sont plus sûres afin d'éviter les conflits entre les espaces de noms
 - Un URI est uniquement un identificateur, qui n'a pas de sens en soi
 - Il ne signifie rien pour le processeur XML, qui le transmet tel quel à l'application

Espaces de noms XML

6

- **Position du problème**
 - Liberté de choix des noms de balises et des attributs XML
 - ⇒ Conflits et polysémie entre ces noms/attributs
 - Besoin d'associer plusieurs applications dans un même document
 - ⇒ « Préfixage » des noms de balises par l'URI de l'application concernée

Espaces de noms XML

- Noms qualifiés (qualified names)
 - Noms de balises appartenant à des espaces de noms
 - Syntaxe : `PrefixeDEspaceDeNoms:PartieLocale`
 - Exemple : `<xsl:stylesheet>`
 - Le préfixe fait référence à un URI
 - Les noms d'attributs peuvent également être préfixés
- Association d'un préfixe à un URI
 - Attribut `xmlns`
 - Exemple : `<xhtml:html xmlns:xhtml="http://www.w3.org/1999/xhtml">`
- Remarques
 - Portée : l'élément porteur de l'attribut `xmlns`
 - Bien entendu, un document XML peut contenir des éléments se référant à plusieurs espaces de noms
 - Le préfixe en lui-même n'a aucune signification
 - En interne, le parser passe à l'application des « noms pleinement qualifiés », où le préfixe est remplacé par la valeur de l'URI

Espaces de noms XML

- Espace de noms par défaut
 - Pas de préfixe d'espace de noms
 - Exemple : `<html xmlns="http://www.w3.org/1999/xhtml">`
- Annulation d'espaces de noms
 - Par valeur de l'attribut `xmlns` vide : `xmlns=""`
- Exemple de code

```
<?xml version="1.0"?>
<CV xmlns="http://www.univ-lyon1.fr/etds/CV/english"
  xmlns:xhtml="http://www.w3.org/1999/xhtml">
  <personne>
    <civil_status>
      <title>Mr.</title>
    </civil_status>
    ...
  </personne>
  <xhtml:html>
    <xhtml:head>
      <xhtml:title>CV of a student</xhtml:title>
    </xhtml:head>
    <xhtml:body>
      ...
    </xhtml:body>
  </xhtml:html>
</CV>
```

Plan

9

- Espaces de noms XML
 - URI, URL et URN
 - Position du problème
 - Principes de fonctionnement
- Schémas XML
 - Différences par rapport aux DTD
 - Introduction à XML-Schema
- Le monde XML
 - Quelques normes liées à XML
 - Quelques DTD / schémas importants

Limites des DTD

10

- XML est seulement une syntaxe
- XML ne porte pas de sémantique
 - ➔ DTD / schémas

Problèmes des DTD

11

- 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
 - ex. entiers, dates, etc.
 - ✱ spécifier des cardinalités simples
 - ex. « un ARTICLE aura entre 1 et 4 MOTS-CLE »
 - ✱ spécifier des contraintes simples
 - ex. entier positif

Comparaison DTD/Schémas

Caractéristique	DTD	Schémas
Syntaxe	Notation EBNF + pseudo-XML	XML 1.0
Outils	Outils SGML existants (chers et complexes)	Tous les outils XML existants et à venir
Supports DOM/SAX	Non	Oui (comme pour les fichiers XML)
Modèles de contenu	<ul style="list-style-type: none"> - Listes : ordonnées ou de choix - Cardinalité : 0, 1 ou plusieurs occurrences - Pas d'éléments nommés ou de groupes d'attributs. 	<ul style="list-style-type: none"> - Listes : ordonnées et de choix (détails de contenus mixtes) - cardinalité : spécification d'un nombre exact d'occurrences possible - groupes de modèles nommés
Typage des données	Faible (chaînes, jetons nominaux, ID...)	Fort (nombres, chaînes, date/heure, booléen, structures...)
Héritage	Non	Oui
Extensibilité	Non (pas sans modification de la recommandation XML)	Oui (puisque fondés sur l'extensibilité de XML)
Contraintes légales	Compatibilité avec SGML	Aucune (simplement des « emprunts » aux DTD, comme pour les types de données)
Nombre de vocabulaires supportés	Une seule DTD par document	Autant que nécessaire (grâce aux espaces de noms)
Dynamisme	Aucune : les DTD sont en lecture seule	Peuvent être modifiés dynamiquement

Schémas XML

13

- Principes généraux
 - 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

Schémas XML

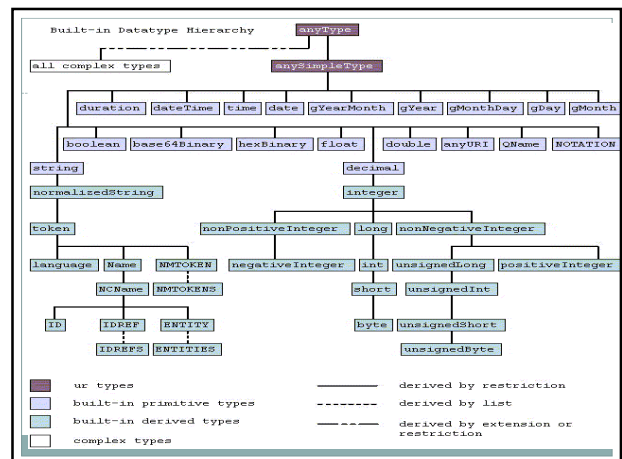
14

- Aspects techniques
 - Utilisation de la syntaxe et des outils XML
 - Extensibilité
 - Dynamicité
 - Possibilité de définir ses propres types de données et modèles de contenus
 - Un schéma définit une **classe** de documents dont chaque document est une **instance**
 - S'appuient sur les notions de
 - Types de données
 - Structures

Schémas XML

15

- Les types de données : 3 dichotomies
 - Hiérarchie arborescente à partir d'un *ur-type*
 - Types primitifs : premier niveau de décomposition
 - Types dérivés : tous les niveaux suivants
 - La recommandation définit un ensemble de types
 - Types intégrés
 - Types dérivés par l'utilisateur
 - Atomicité
 - Types atomiques : dont les valeurs ne peuvent pas être décomposées
 - Types listes : ensembles de valeurs atomiques
- Remarques
 - Tous les types primitifs sont intégrés. La réciproque est fausse
 - string est un type atomique



Schémas XML

16

- Types de données
 - Les types de données comportent 3 caractéristiques
 - Espace lexical : définit tous les caractères représentant les valeurs possibles
 - Espace de valeurs : ensemble des valeurs exprimé dans l'espace lexical
 - Facettes : propriétés définitionnelles de l'ensemble des valeurs
 - Facettes fondamentales : propriétés abstraites (égalité, bornes, ordre, cardinalité, numérique ou non)
 - Facettes de contraintes : limitent certaines propriétés (12 facettes : length, enumeration, minExclusive...)

Schémas XML

17

- Les structures
 - Permettent de définir des types de données (contenus et attributs) selon deux méthodes
 - SimpleType : dérivation de types atomiques
 - Par restriction (par intension)
 - Par liste (par extension)
 - Par union (sur-ensemble de types existants)

```
<xsd:simpleType name="myInteger">
  <xsd:restriction base="xs:integer">
    <xsd:minInclusive value="-2"/>
    <xsd:maxExclusive value="5"/>
  </xsd:restriction>
</xsd:simpleType>
<xsd:simpleType name="myIntList">
  <xsd:list>
    <xsd:restriction base="xs:integer">
      <xsd:maxInclusive value="100"/>
    </xsd:restriction>
  </xsd:list>
</xsd:simpleType>
<xsd:simpleType name="IntOrUndefined">
  <xsd:union>
    <xsd:simpleType>
      <xsd:restriction base="xs:integer"/>
    </xsd:simpleType>
    <xsd:simpleType>
      <xsd:restriction base="xs:NMTOKEN">
        <xsd:enumeration value="undefined"/>
      </xsd:restriction>
    </xsd:simpleType>
  </xsd:union>
</xsd:simpleType>
```

Schémas XML

• Les structures

- Permettent de définir des types de données (contenus et attributs) selon deux méthodes
 - **ComplexType** : autres types de dérivation
 - Dérivation
 - par restriction d'un type de base complexe,
 - par extension d'un type de base (simple ou complexe),
 - par restriction de l'ur-type definition
 - En pratique : la définition d'un type complexe est une composition
 - De séquences (ET)
 - ordonnées : `xsd:sequence`
 - non-ordonnées : `xsd:all`
 - De choix (OU) : `xsd:choice`

Schémas XML

• Les structures

- Définition d'un élément
 - Avec la balise `xsd:element`
 - En utilisant le type choisi (simple ou complexe)
- Définition d'un attribut
 - Avec la balise `xsd:attribute`
 - En utilisant un type simple
- Pour aller plus loin : un cours très instructif
<http://globalcomputing.cpl.ch/unifr/seance02-xml-schema-1/xml-schema-notes.pdf>

```
<xsd:element name="recette">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="titre" type="xs:string"/>
      <xsd:element name="commentaire" type="xs:string" minOccurs="0"
        maxOccurs="unbounded"/>
      <xsd:element name="item" maxOccurs="unbounded">
        <xsd:complexType>
          <xsd:sequence>
            <xsd:element name="entete" type="xs:string" minOccurs="0"/>
            <xsd:choice maxOccurs="unbounded">
              <xsd:element name="ingredient" type="xs:string"/>
              <xsd:element name="preparation" type="xs:string"/>
            </xsd:choice>
          </xsd:sequence>
        </xsd:complexType>
      </xsd:element>
    </xsd:sequence>
  </xsd:complexType>
</xsd:element>

<xsd:element name="MusicDescription">
  <xsd:complexType>
    <xsd:all>
      <xsd:element name="country" type="xs:string"/>
      <xsd:element name="originalTitle" type="xs:string"/>
      <xsd:element name="author" type="xs:string"/>
    </xsd:all>
  </xsd:complexType>
</xsd:element>

<xsd:element name="picture" minOccurs="0">
  <xsd:complexType>
    <xsd:attribute name="source" type="xs:anyURI"/>
  </xsd:complexType>
</xsd:element>
```

Schémas XML

• Préambule d'un schéma

- Avec gestion des espaces de noms


```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xsd:targetNamespace="http://www.monsite.com/monnamespace"
  xmlns:tns="http://www.monsite.com/monnamespace">
```
- Sans gestion des espaces de noms


```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xsd:noTargetNamespace="noTargetNamespace">
```

ou simplement

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
```

Schémas XML

• Association d'un document à un schéma

- Avec gestion des espaces de noms


```
<ici:element
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://monsite.com/monnamespace
    http://monsite.com/monnamespace/schema/MonSchema.xsd"
  xmlns:ici="http://monsite.com/monnamespace">
```
- Sans gestion des espaces de noms


```
<element
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:noNamespaceSchemaLocation="http://monsite.com/monnames
    pace/schema/MonSchema.xsd">
```
- Dans tous les cas, il faut fournir une URI vers le schéma

Plan

24

- Espaces de noms XML
 - URI, URL et URN
 - Position du problème
 - Principes de fonctionnement
- Schémas XML
 - Différences par rapport aux DTD
 - Introduction à XML-Schema
- Le monde XML
 - Quelques normes liées à XML
 - Quelques DTD / schémas importants

Avantages et applications XML

25

- **Avantages**
 - réutilisabilité, partage
 - pérennité
 - intégrité
 - portabilité
- **Applications**
 - documents
 - échange de données
 - bureautique
 - Web
 - BDD semi-structurées
 - commerce électronique
 - ...

Quelques standards XML

26

- The XML Bookmark Exchange Language (XBEL)
- Open eBook Publication Structure
- SportsML
- NewsML
- XML Book Industry Transaction Standards (XBITS)
- OpenDocument...
- 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>

Quelques spécifications du W3C liées à XML

27

- 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/>

Conclusion

28

- **XML**
 - standard sorti en 1998
 - Unicode / généricité
 - documents / données
 - mondialement adoptée
- **Standards et normes**
 - variés : dans tous les domaines nécessitant
 - pérennité
 - échange
 - plus ou moins adaptés et adoptés
 - questions récurrentes
 - évolution
 - interopérabilité
 - guerre des standards et le contrôle de l'information

Annexes

29

- **Quelques normes XML utiles aux sciences de la documentation**
 - Docbook
 - OpenDocument
 - EAD/EAC
 - TEI

DocBook

30

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

Open Document

31

- Objectif
 - codage des documents de suites bureautiques
 - textes, feuilles de calcul, présentations, dessins, formules, bases de données...
 - modèles pour ces documents
- Mis en place par OASIS
 - à partir des formats de OpenOffice (SUN)
- Mai 2005 : version 1.0
- 2006 : ISO
- Utilisé dans Open Office 2.0
- Standard ouvert
 - intérêt pour le partage, la récupération, etc.
 - enjeu politique (Massachusetts, sept 2005)
 - finalement MS Office exportera en Open Document
 - tentative Microsoft ISO / OOXML

EAD / EAC

32

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

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

33

- 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

En-tête EAD <eadheader>

34

```
<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
      <subtitle> 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
    <date> Date de l'inventaire
    <language> Langue utilisée
    <desrules> Règles de description (archivistique utilisée)
  <revisiondesc> Descriptions des révisions ; permet de gérer les versions successives de l'instrument de recherche encodé
</eadheader>
```

En-tête EAD <eadheader>

35

```
<eadheader audience="external" filedescstatus="edited-partial-draft" encoding="utf-8">
  <eadid encodinganalog="Identifier">FRDAFANCH05AP_000000001</eadid>
  <filedesc>
    <titleproper encodinganalog="Title">Etat sommaire des fonds d'archives privées du Centre historique des Archives nationales</titleproper>
    <subtitle> Série 317AP à 317AP et ABXIX0268 à ABXIX0225</subtitle>
    <author encodinganalog="creator">Instrument de recherche rédigé par Claire Sibille, [...] encodé en XML, conformément à la DTD EAD pub converti en HTML par Martin Bégin (jocelyn.AULNE) et Florence Clavaud (service des nouvelles technologies du CHAN)</author>
    <editionstmt>
      <edition>Quatrième édition</edition></editionstmt>
    <publicationstmt>
      <publisher encodinganalog="publisher">Centre historique des Archives nationales de France (CHAN)</publisher>
      <address>
        <addressline>40 rue des Francs Bourgeois</addressline>
        <addressline>75141 PARIS CEDEX 05</addressline>
      </address>
      <date encodinganalog="date">août 2003</date>
    </publicationstmt>
    <profiledesc>
      <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 languagecode="fr">français</language></language>
      <revisiondesc>
        <change>
          <date>8 septembre 2003</date>
          <desc>fichier converti en EAD 2002 par programme XSL-T fourni par D. Pili (SAA) adapté</desc>
        </change>
      </revisiondesc>
    </profiledesc>
  </filedesc>
</eadheader>
```

Préliminaires <frontmatter> : exemple

36

```
<frontmatter>
  <titlepage>
    <publisher>Centre historique des Archives nationales</publisher>
    <titleproper>Etat sommaire</titleproper>
    <subtitle>Série AP (317AP à 317AP) et ABXIX (ABXIX0268 à ABXIX0225)</subtitle>
    <author>par Claire SIBILLE</author><conservateur du Patrimoine></author>
    <author>avec la collaboration de Fanny AUDOUS et de Violaine LE NENAON</author>
    <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 TAP à 315AP. Depuis, plus de 300 nouveaux fonds d'archives privées sont entrés aux Archives nationales. [...] Ces notices sont également consultables dans la base <entref href="http://sdx.archivesdefrance.culture.gouv.fr/ap">BORA</entref></p>
  </div>
  <div>
    <head>Mode d'emploi</head>
    <p>[...]</p>
  </div>
</frontmatter>
```

Description archivistique <archdesc>

37

<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
- <bibliography> Bibliographie
- <bioghist> Biographie ou Histoire (du producteur)
- <controlaccess> Vedettes et accès contrôlés →
- <custodhist> Historique de la conservation
- <dao> et <daogrp> Objet archivistique numérique
- <note> Note
- <dd> Autres données descriptives
- <originaldoc> Existence et lieu de conservation des documents originaux
- <otherfindaid> Autre instrument de recherche
- <physdesc> Caractéristiques matérielles et contraintes techniques
- <prefercite> 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>

Les sous-éléments de <did> Identification et description

38

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

Exemple d'utilisation de <did>

39

<archdesc level="fonds">

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

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

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

```
<archdesc level="fonds">
  <did>
    <repository>Centre historique des Archives nationales (France)</repository>
    <origination>Ce fonds a été produit par <persname>Henry-Louis Duhamel du Monceau</persname>, par son frère
    <persname>Duhamel de Denainvilliers</persname>, par leur neveu <persname>Auguste Louis Fougeroux de Bondaroy</persname> et par le
    frère de ce dernier <persname>Fougeroux de Secval</persname></origination>
    <unitid>127AP</unitid>
    <unittitle>Fonds Duhamel du Monceau, Duhamel de Denainvilliers, Fougeroux de Bondaroy et Fougeroux de Secval</unittitle>
    <unitdate normal="1700/1788">1700 - 1788</unitdate>
    <physdesc> <extent>8 cartons</extent> (127AP/1 à 127AP/7, 127AP/20), <extent>1 mètre linéaire</extent></physdesc>
    <langmaterial><language langcode="fre">français</language></langmaterial>
  </did>
  <acqinfo><p>achats, 7 juin 1957, 24 mars 1960, 16 juin 1977, 11 décembre 1991</p></acqinfo>
  <bioghist>
    <p>Henry-Louis Duhamel du Monceau (1700-1788) fit des études scientifiques au jardin du Roy, actuel jardin des plantes, en 1724. Associé-botaniste à l'Académie des sciences en 1739, inspecteur général de la Marine en 1739, membre de l'Académie des sciences de Paris et de la Société royale de Londres, Duhamel du Monceau écrivit en 1745 un Essai sur la conservation des grains et en 1747, un Traité de la fabrication des manœuvres pour les vaisseaux ou l'art de la corderie perfectionnée. On lui doit aussi [...] </p> [...]
    <p>Commencé vers 1720 par Le Masson du Parc, commissaire ordinaire de la Marine et inspecteur général des pêches, le Traité général des pêches s'inscrivait dans le monumental projet éditorial entrepris 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, survenue en 1741, le projet fut repris par Henry-Louis Duhamel du Monceau, alors inspecteur général de la Marine, qui s'attacha personnellement à cette entreprise. Aidé de ses neveux Fougeroux, 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 [...] </p> [...]
    <p>Commencé vers 1720 par Le Masson du Parc, commissaire ordinaire de la Marine et inspecteur général des pêches, le Traité général des pêches s'inscrivait dans le monumental projet éditorial entrepris 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, survenue en 1741, le projet fut repris par Henry-Louis Duhamel du Monceau, alors inspecteur général de la Marine, qui s'attacha personnellement à cette entreprise. Aidé de ses neveux Fougeroux, 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 [...] </p> [...]
  </bioghist>
  <custodhist>
    <p>Il n'y a pas de restriction juridique à la consultation de ces documents, qui se fait selon les modalités matérielles en vigueur dans le service d'archives</p></accessrestrict>
    <p>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. </p> [...] </scopecontent>
    <p>Le Traité général des pêches est une œuvre majeure de l'histoire de la pêche et de l'agriculture. Elle est constituée de deux volumes : le Traité général des pêches (1727AP/1 à 127AP/7) et le Traité de la fabrication des manœuvres (127AP/8 à 127AP/20). Le Traité général des pêches est une œuvre majeure de l'histoire de la pêche et de l'agriculture. Elle est constituée de deux volumes : le Traité général des pêches (1727AP/1 à 127AP/7) et le Traité de la fabrication des manœuvres (127AP/8 à 127AP/20). Le Traité général des pêches est une œuvre majeure de l'histoire de la pêche et de l'agriculture. Elle est constituée de deux volumes : le Traité général des pêches (1727AP/1 à 127AP/7) et le Traité de la fabrication des manœuvres (127AP/8 à 127AP/20).
  </custodhist>
  <altformavail><p>Les articles 127AP/8 à 127AP/19 sont conservés uniquement sous forme de microfilm. </p></altformavail>
  <separatedmaterial><p>La documentation sur les pêches rassemblée par Le Masson puis Duhamel du Monceau est répartie dans différentes bibliothèques, aux États-Unis (Houghton Library d'Harvard mais aussi American Philosophical Society de Philadelphie, Newberry Library de Chicago) ou en France (Bibliothèque municipale de Lyon, Académie des sciences à Paris). </p></separatedmaterial>
  <bibliography><p>Informations extraites de l'introduction du Répertoire numérique détaillé. </p></bibliography>
```

TEI

41

- Text Encoding Initiative
- 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
- 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

42

- Financé sur des programmes « Humanités »
- Sortie de la norme (Recommandation) en mai 1994
- 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
- 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

43

- 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.).
- Noyau obligatoire, autre éléments facultatifs
- Importance de l'en-tête
 - codage systématique des méta-données de n'importe quel document électronique

TEI : où en est on ?

44

- 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)
 - multiplier les éléments / aux besoins
 - ou bien limiter volontairement le nombre d'éléments
 - etc.