

# Introduction à XML

Georges-André Silber  
CRI/ENSMP

<http://www.cri.ensmp.fr/people/silber/xml>

# Petit tour du Web

# 1990: le temps était venu...

— [ Le World Wide Web est né dans les années 90

— [ Internet était en pleine croissance

— [ La technologie "hypertexte" était en vogue

— [ Le "Domain Name System" (DNS) existait depuis peu (1985)

— [ Il existait SGML, norme ISO 8879 depuis 1986

# Hypertexte: Bush

— [ Article "As We May Think" de Vannevar Bush dans l'Atlantic Monthly de juillet 1945

— [ Memex : machine permettant de lier des informations entre elles



# Hypertexte : Nelson

Ted Nelson (Oxford)

Inventeur du terme "hypertexte"

Systeme Xanadu (1960)

Le Web doit beaucoup à Xanadu

Ted Nelson n'aime pas le Web



HTML is precisely what we were trying to PREVENT– ever-breaking links, links going outward only, quotes you can't follow to their origins, no version management, no rights management.

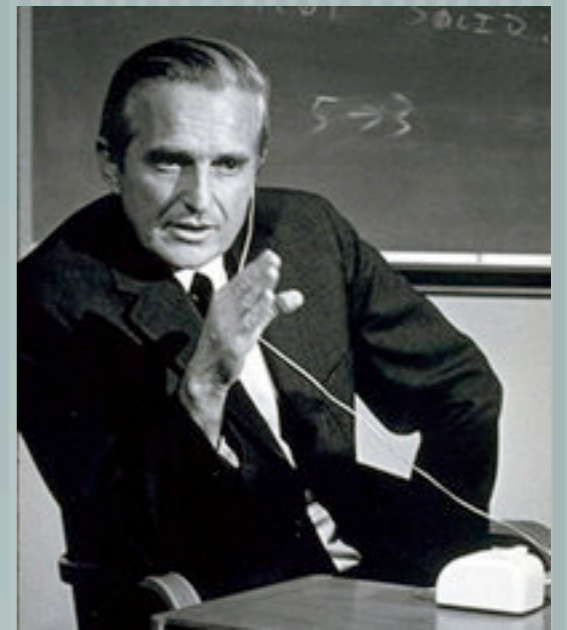
– Ted Nelson

# Hypertexte: Engelbart

— [ Douglas Engelbart (Stanford Research Institute)

— [ Développement du "On-Line System" (NLS)

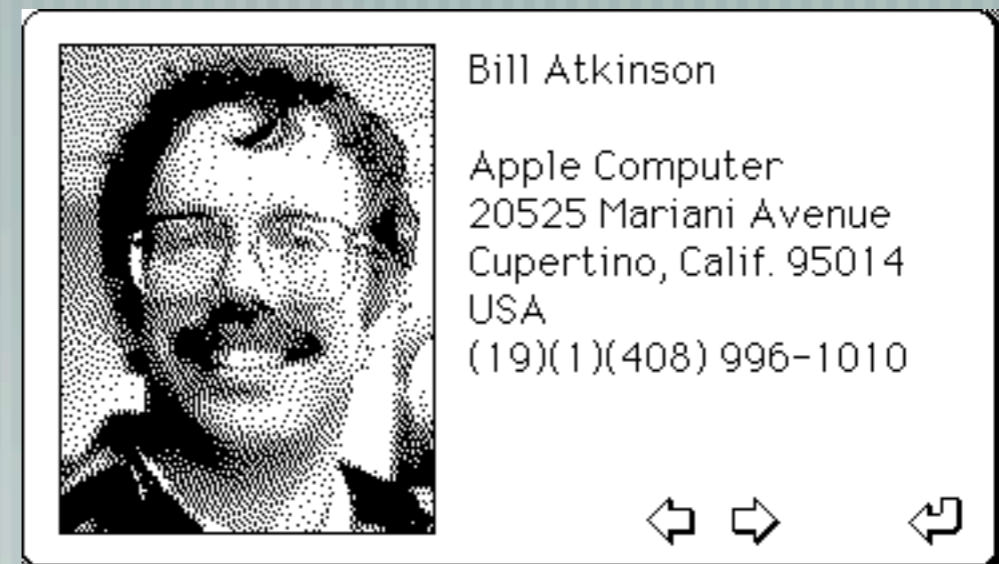
— [ Précurseur des écrans "bitmap", GUI, fenêtres, souris, hypertexte (années 70)



# Hypertexte : Atkinson

— [ Bill Atkinson (Apple) crée Hypercard, permettant de naviguer avec des liens entre des "cartes"

— [ Logiciel très populaire



# Domain Name System

— [ Depuis la mi-1980, nouveau système permettant de nommer facilement des machines sur Internet

— [ Idée de "nom de domaine", organisation hiérarchique

— [ Correspondance entre un numéro IP (193.48.171.253) et un nom ([www.cri.ensmp.fr](http://www.cri.ensmp.fr))

— [ Logiciel "Domain Name System" (DNS), annuaire distribué

# Markup Languages

— [ “Markup”: historiquement, ce qui permet au typographe de composer un texte pour l’impression

— [ “Markup codes”: passage à l’ère électronique

— [ Problème universel: codage permettant de rendre explicite l’interprétation d’un texte

— où est-ce que les choses commencent, finissent, et comment elles se structurent

— Livre, chapitres, pages, paragraphes, phrases, mots...

# Markup: GML

## — [ IBM Generalized Markup Language (années 70)

```
:h1.Chapter 1: Introduction
:p.GML supported hierarchical containers, such as
:ol
:li.Ordered lists (like this one),
:li.Unordered lists, and
:li.Definition lists
:eol.
as well as simple structures.
:p.Markup minimization (later generalized and formalized in SGML),
allowed the end-tags to be omitted for the "h1" and "p" elements.
```

# Markup: SGML

— [ Standard Generalized Markup Language (SGML)

— [ Norme ISO 8879 (années 80)

— [ SGML est un métalangage, un moyen de décrire formellement un langage, dans ce cas un langage de marquage (ou de balisage)

— [ “Grammaire” d’un langage décrite avec le formalisme DTD (Document Type Definition)

# Exemple SGML

```
<title>SGML example</title>

<h1>Chapter 1: Introduction</h1>
<p>SGML supported hierarchical containers,
such as</p>
<ol>
<li>Ordered lists (like this one)</li>
<li>Unordered lists, and</li>
<li>Definition lists</li>
</ol>
<p>as well as simple structures.</p>
```

# L'invention du Web

— [ 1989: Tim Berners-Lee travaille au CERN et invente le Web

— [ 1994: Premier navigateur Web diffusé: NCSA Mosaic

— [ 200x: HTML 4.01, CSS 2.0, JavaScript, AJAX

— [ Etat de l'art des technologies Web: Google Maps

# Tim Berners-Lee

— [ Le père du WWW

— [ Inventeur des deux  
protocoles fondateurs du  
World Wide Web en 1989,  
HTTP et HTML

— [ Directeur du W3C depuis  
1994



# Les débuts du Web

— [ Le World Wide Web est né au CERN

— [ L'idée de base: lier des articles scientifiques entre eux grâce à des liens "hypertexte"

— Création d'une toile (web) de documents

— [ Premier prototype de Tim en 1990 sur un ordinateur NeXT

# HTML

— [ Création d'un nouveau langage pour représenter des documents: HTML (HyperText Markup Language)

— [ Inspiration: SGML

— [ Récupération de certaines balises de SGML (<h1>, <h2>, <p>, <ol>, etc...)

— [ Rajout de balises originales

```
<a href="http://www.w3.org">World Wide Web Consortium</a>
```

# HTTP

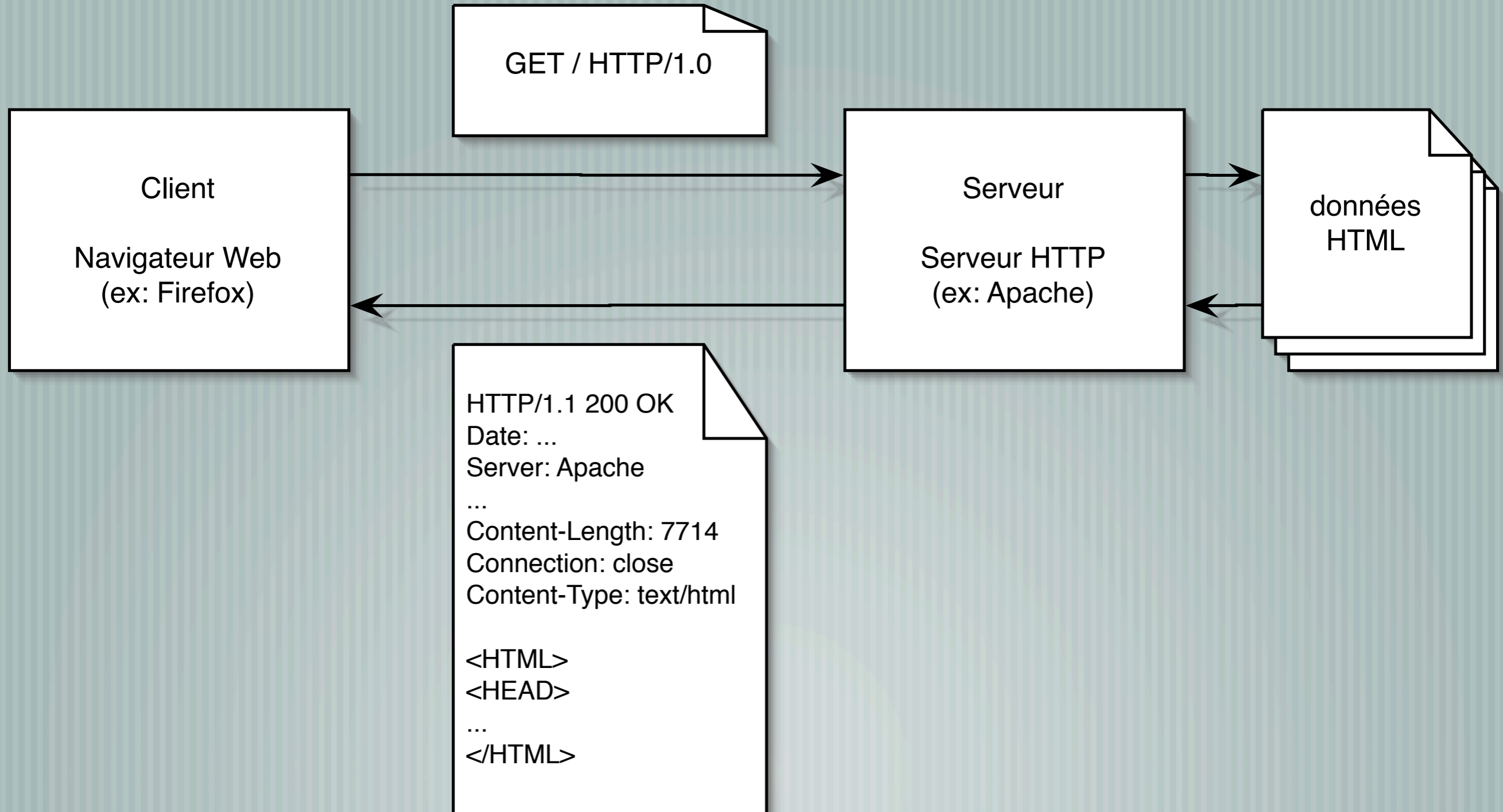
— [ HyperText Transfer Protocol (HTTP)

— [ Protocole très simple permettant de retrouver des documents sur le réseau par des liens hypertexte

— [ Pourquoi un nouveau protocole ? Pour la simplicité et pour tirer parti des spécificités de l'hypertexte

— [ Nommage des serveurs HTTP en 'www.nom.nom'

# HTTP+HTML



# Démonstration

```
Terminal — bash — 80x18
champagny:~ Silber$ telnet www.cri.ensmp.fr 80
Trying 193.48.171.253...
Connected to www.cri.ensmp.fr.
Escape character is '^]'.
GET /people/silber/bonjour.html HTTP/1.0

HTTP/1.1 200 OK
Date: Mon, 12 Sep 2005 13:56:13 GMT
Server: Apache
Last-Modified: Mon, 12 Sep 2005 13:54:04 GMT
ETag: "4fc054-12-4325887c"
Accept-Ranges: bytes
Content-Length: 18
Connection: close
Content-Type: text/html

<h1>Bonjour !</h1>Connection closed by foreign host.
champagny:~ Silber$ █
```

# En 1991: mailing-list

— [ Création de la mailing-list WWW-talk

— [ Volonté de discuter des différents choix de manière ouverte  
(naturel pour un chercheur)

— [ Création de HTML+ par Dave Raggett (HP), pour améliorer la  
richesse de l'affichage



# 1992: NCSA Mosaic

— [ Développement de NCSA Mosaic par le National Center for Supercomputing Applications (USA)

— [ Dans l'équipe de développement: Marc Andreessen, le futur fondateur de Netscape (qui donna ensuite naissance à Mozilla et à Firefox)

— [ Introduction de la balise IMG par l'équipe de Mosaic

— [ Internet Explorer a repris une partie du code de Mosaic





Document Title: NCSA Mosaic Home Page

Document URL: http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/NCSAMosaicHome.html



Welcome to NCSA Mosaic, an Internet information browser and [World Wide Web](#) client. NCSA Mosaic was developed at the [National Center for Supercomputing Applications](#) at the [University of Illinois](#) in --> Urbana-Champaign. NCSA Mosaic software is [copyrighted](#) by The Board of Trustees of the University of Illinois (UI), and ownership remains with the UI.

---

Jan `97

The Software Development Group at NCSA has worked on NCSA Mosaic for nearly four years and we've learned a lot in the process. We are honored that we were able to help bring this technology to the masses and appreciated all the support and feedback we have received in return. However, the time has come for us to concentrate our limited resources in other areas of interest and development on Mosaic is complete.

*All information about the Mosaic project is available from the homepages.*

NCSA Mosaic Platforms:

- [NCSA Mosaic for the X Window System](#)
- [NCSA Mosaic for the Apple Macintosh](#)
- [NCSA Mosaic for Microsoft Windows](#)

World Wide Web Resources The following resources are available to help introduce you to cyberspace and keep track of its growth:

- A [glossary](#) of World Wide Web terms and acronyms
- An [INDEX](#) to Mosaic related documents
- [NCSA Mosaic Access Page](#) for persons with disabilities
- Mosaic and WWW related [Tutorials](#)
- [Internet Resources Meta-Index](#) at NCSA
- [Suggested Starting Points for Internet Exploration](#)

# 1994: W3C


— [ Formation du W3C en 1994

— [ Président: Tim Berners-Lee

— [ Développer et promouvoir des standards ouverts pour le Web



The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential. W3C is a forum for information, commerce, communication, and collective understanding. On this page, you'll find [W3C news](#), links to [W3C technologies](#) and ways to [get involved](#). New visitors can find help in [Finding Your Way at W3C](#). We encourage organizations to learn more [about W3C](#) and [about W3C Membership](#).

<h3>Mobile Web Initiative</h3> <p>W3C has launched the <a href="#">Mobile Web Initiative</a> (MWI) to make Web access from a mobile device as simple, easy, and convenient as Web access from a desktop device. Read <a href="#">about MWI</a> and <a href="#">how to sponsor MWI</a>.</p>	<h3>News</h3> <p>► <b>xml:id Is a W3C Recommendation</b></p> <p>2005-09-09: The World Wide Web Consortium today released <a href="#">xml:id Version 1.0</a> as a W3C Recommendation. The specification defines an attribute name, <code>xml:id</code>, that can always be treated as an identifier and hence can always be recognized, without fetching external resources, and without relying on an internal subset. The Recommendation is the latest deliverable of the <a href="#">XML Core Working Group</a>, part of the <a href="#">W3C XML Activity</a>. (<a href="#">News archive</a>)</p> <p>► <b>Working Draft: EARL 1.0 Schema</b></p> <p>2005-09-12: The Evaluation and Repair Tools Working Group has released a Working Draft of the <a href="#">Evaluation and Report Language (EARL) 1.0 Schema</a>. EARL is a flexible format used to exchange, combine and compare test results including bug reports, test suite evaluations and conformance claims. The test subjects might be Web sites, authoring tools, user agents or other entities. The group welcomes feedback from Web developers and researchers. Read about the <a href="#">Web Accessibility Initiative</a>. (<a href="#">News archive</a>)</p> <p>► <b>Position Papers Due 23 September for W3C Workshop on</b></p>	<h3>Search</h3> <p>Google™</p> <p>Search W3C</p> <input type="text"/> <input type="button" value="Go"/> <p><a href="#">Search W3C Mailing Lists</a></p> <h3>Members</h3> <p><b>CWI</b></p>  <p>The Centrum voor Wiskunde en Informatica (CWI) is a leading research institute in mathematics and computer science. CWI is active in several W3C Working Groups and hosts the W3C Benelux Office. As the starting point for the European Internet, we are proud to collaborate with W3C. (<a href="#">Member testimonials</a>)</p> <ul style="list-style-type: none"><li>• <a href="#">Member Home Page</a></li><li>• <a href="#">Member Submissions</a></li><li>• <a href="#">Current Members</a></li><li>• <a href="#">W3C Fellows</a></li></ul>
<h3>W3C Membership News</h3> <p>W3C has lowered fees for non-profit and small commercial organizations based in developing countries (see the <a href="#">press release</a>) and for intermediate-sized companies. Read <a href="#">about W3C fees</a>.</p>		
<h3>W3C A to Z</h3> <ul style="list-style-type: none"><li>• <a href="#">Accessibility</a></li><li>• <a href="#">Amaya</a></li><li>• <a href="#">Annotea</a></li><li>• <a href="#">CC/PP</a></li><li>• <a href="#">Compound Document Formats</a></li><li>• <a href="#">CSS</a></li><li>• <a href="#">CSS Validator</a></li><li>• <a href="#">Device Independence</a></li><li>• <a href="#">DOM</a></li><li>• <a href="#">HTML</a></li><li>• <a href="#">HTML Tidy</a></li><li>• <a href="#">HTML Validator</a></li></ul>		

# 1994-200x

— [ HTML 1.0, 2.0, 3.0, 4.0, 4.1

— [ CSS (Cascading Style Sheets) 1.0, 2.0

— [ JavaScript

— [ Volonté du W3C depuis 1996 de promouvoir un “meilleur HTML”

— [ Création de XML en 1998 (1.0)

— [ XHTML 1.0 (HTML 4.01 “XMLisé”)

**XML**

# La norme XML

- [ XML : eXtensible Markup Langage
- [ Langage de balisage extensible
- [ Norme du W3C depuis 1998
- [ Version 1.0 (février 1998)
- [ Version 1.1 (février 2004)

# Origine et buts de XML

— [ HTML 1.0, 2.0, 3.0, 4.0, 4.1

— Volonté de “stabiliser” le langage du Web

— [ Comment : en créant un meilleur HTML

— [ Inspiration : SGML (Norme ISO 1986)

— [ Séparation données / présentation

— Données : XML (ou XHTML)

— Présentation: CSS, XSLT, etc...



# Extensible Markup Language (XML) 1.1

W3C Recommendation 04 February 2004, edited in place 15 April 2004

**This version:**

<http://www.w3.org/TR/2004/REC-xml11-20040204/>

**Latest version:**

<http://www.w3.org/TR/xml11>

**Previous version:**

<http://www.w3.org/TR/2003/PR-xml11-20031105/>

**Editors:**

Tim Bray, Textuality and Netscape <[tbray@textuality.com](mailto:tbray@textuality.com)>

Jean Paoli, Microsoft <[jeanpa@microsoft.com](mailto:jeanpa@microsoft.com)>

C. M. Sperberg-McQueen, W3C <[cmsmcq@w3.org](mailto:cmsmcq@w3.org)>

Eve Maler, Sun Microsystems, Inc. <[eve.maler@east.sun.com](mailto:eve.maler@east.sun.com)>

François Yergeau <[fyergeau@alis.com](mailto:fyergeau@alis.com)>

John Cowan <[cowan@ccil.org](mailto:cowan@ccil.org)>

Please refer to the [errata](#) for this document, which may include some normative corrections.

This document is also available in these non-normative formats: [XML](#) and [XHTML with color-coded revision indicators](#).

See also [translations](#).

Copyright © 2004 W3C® (MIT, ERCIM, Keio), All Rights Reserved. W3C [liability](#), [trademark](#), [document use](#) and [software licensing](#) rules apply.

---

## Abstract

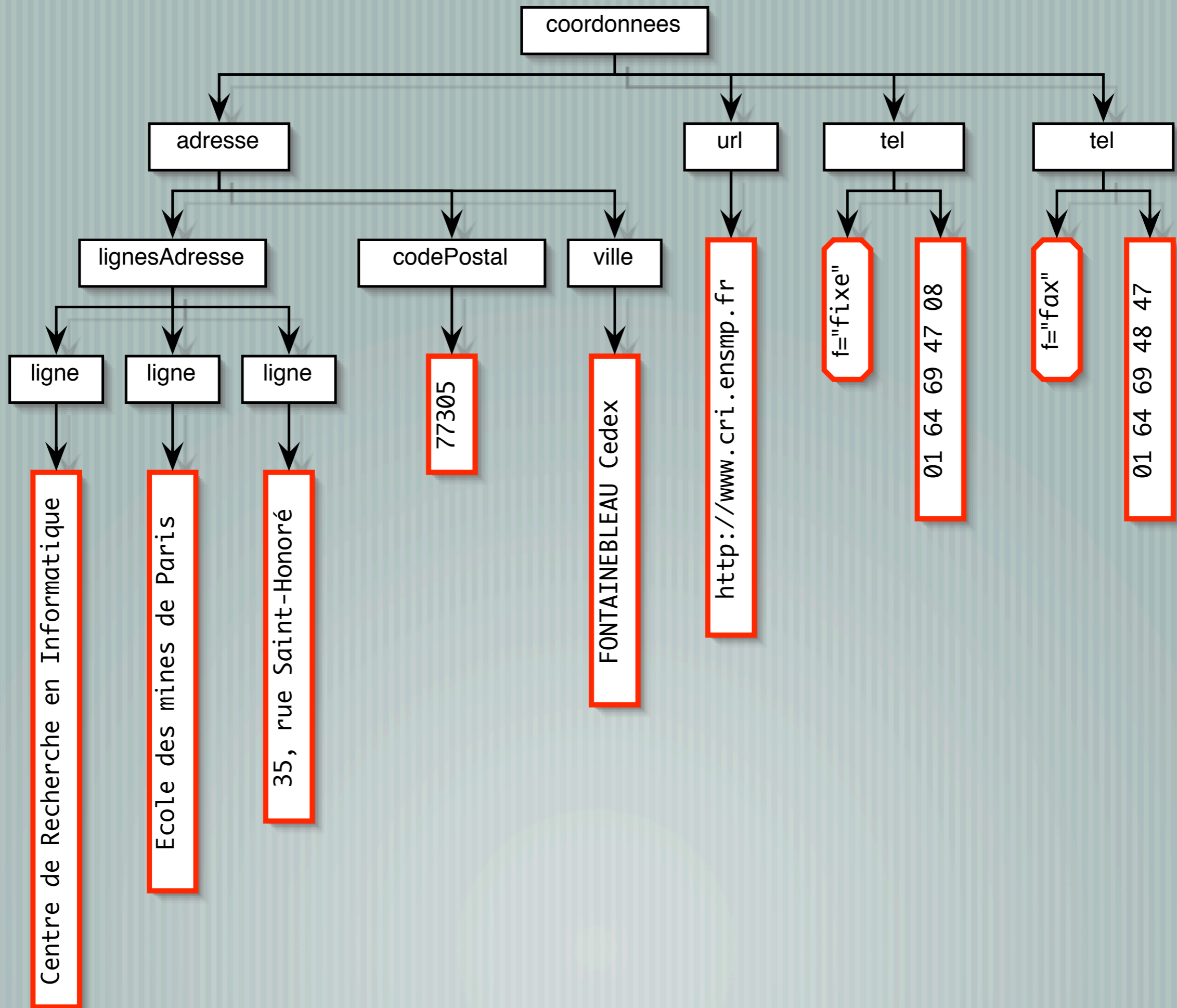
The Extensible Markup Language (XML) is a subset of SGML that is completely described in this document. Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

# Qu'est-ce que XML ?

- [ Un langage de description d'une classe d'objets de données appelés documents XML
- [ La norme décrit partiellement le comportement de programmes les manipulant
- [ XML est une forme restreinte de SGML (1986)

# Exemple de document XML

```
<?xml version="1.0" encoding="iso-8859-1"?>
<coordonnees>
  <adresse>
    <lignesAdresse>
      <ligne>Centre de Recherche en Informatique</ligne>
      <ligne>Ecole des mines de Paris</ligne>
      <ligne>35, rue Saint-Honoré</ligne>
    </lignesAdresse>
    <codePostal>77305</codePostal>
    <ville>FONTAINEBLEAU Cedex</ville>
  </adresse>
  <url>http://www.cri.ensmp.fr</url>
  <tel t="fixe">01 64 69 47 08</tel>
  <tel t="fax">01 64 69 48 47</tel>
</coordonnees>
```



# Documents XML

— [ Composés d'unités de stockage appelées entités, contenant des données analysées syntaxiquement (parsées) ou non.

— [ Ces données sont des caractères qui sont soit des données simples soit des données concourant au marquage (balisage).

# Entités, etc...

```
<?xml version="1.0" encoding="iso-8859-1"?>
<coordonnees> balise ouvrante
  <adresse>
    <lignesAdresse>
      <ligne>Centre de Recherche en Informatique</ligne>
      <ligne>Ecole des mines de Paris</ligne>
      <ligne>35, rue Saint-Honoré</ligne>
    </lignesAdresse>
    <codePostal>77305</codePostal> élément
    <ville>FONTAINEBLEAU Cedex</ville> élément
  </adresse> données simples
  <url>http://www.cri.ensmp.fr</url>
  <tel t="fixe">01 64 69 47 08</tel>
  <tel t="fax">01 64 69 48 47</tel> balise fermante
</coordonnees> attribut
```

# Documents XML

— [ Le balisage décrit les structures logiques et de stockage du document

— [ Un document XML est BIEN FORME il respecte les règles de XML

— [ XML fournit un mécanisme pour contraindre ces structures, les DTD (Définition de Type de Document)

— [ Un document XML peut être VALIDE par rapport à une ou plusieurs DTD

# Exemple de DTD

```
<!ELEMENT coordonnees (adresse, url?, tel*)>
<!ELEMENT adresse (lignesAdresse, codePostal, ville)>
<!ELEMENT lignesAdresse (ligne+)>
<!ELEMENT ligne (#PCDATA)>
<!ELEMENT codePostal (#PCDATA)>
<!ELEMENT ville (#PCDATA)>
<!ELEMENT url (#PCDATA)>
<!ELEMENT tel (#PCDATA)>
<!ATTLIST tel t (fixe|fax|mob) "fixe">
```

# Utilisation d'une DTD locale

```
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE coordonnees SYSTEM "adresse.dtd">
<coordonnees>
    <adresse>
        <lignesAdresse>
            <ligne>Centre de Recherche en Informatique</ligne>
            <ligne>Ecole des mines de Paris</ligne>
            <ligne>35, rue Saint-Honoré</ligne>
        </lignesAdresse>
        <codePostal>77305</codePostal>
        <ville>FONTAINEBLEAU Cedex</ville>
    </adresse>
    <url>http://www.cri.ensmp.fr</url>
    <tel t="fixe">01 64 69 47 08</tel>
    <tel t="fax">01 64 69 48 47</tel>
</coordonnees>
```

déclaration de type de document

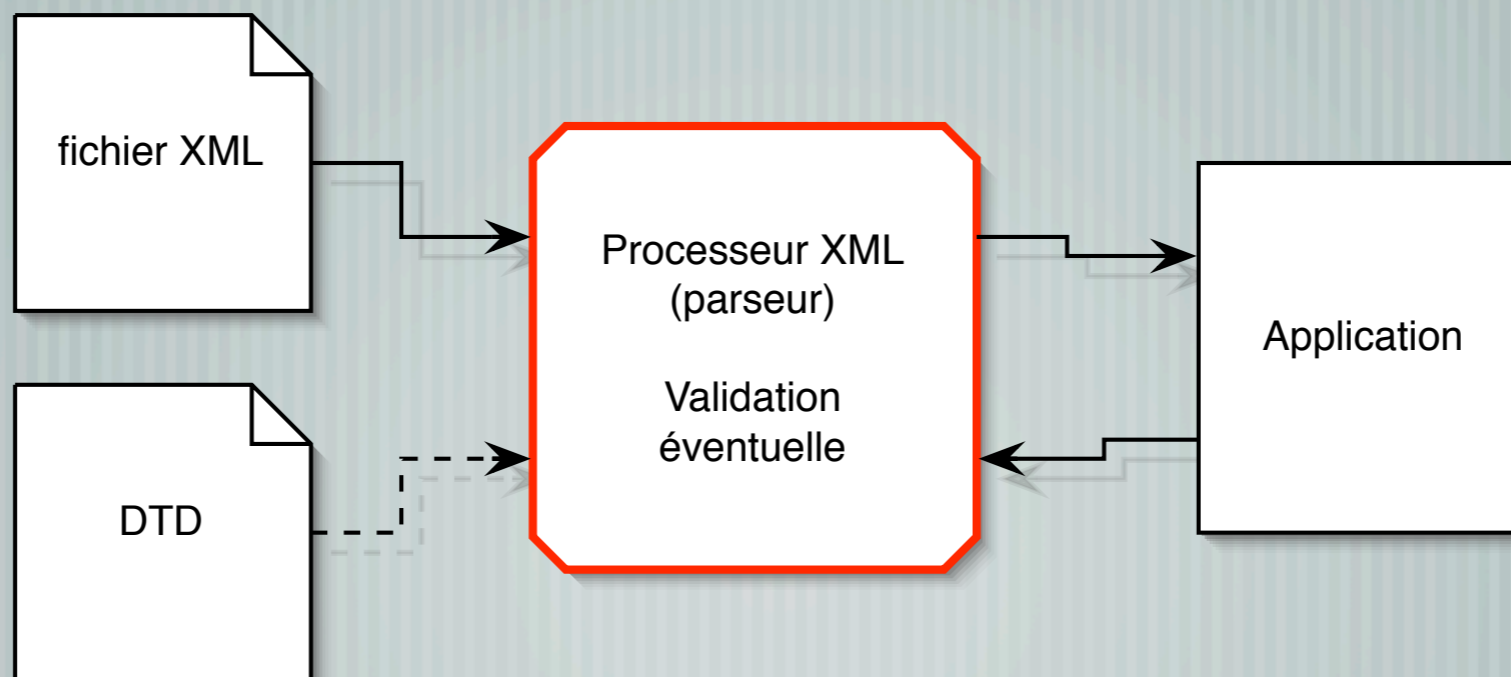
# Utilisation d'une DTD distante

```
<?xml version="1.0" encoding="iso-8859-1"?> Formal Public Identifier (FPI)  
<!DOCTYPE coordonnees PUBLIC ("-//silber//DTD adresse//FR")  
  "http://www.cri.ensmp.fr/people/silber/xml/adresse.dtd">  
  URL de secours  
<coordonnees>  
  <adresse>  
    <lignesAdresse>  
      <ligne>Centre de Recherche en Informatique</ligne>  
      <ligne>Ecole des mines de Paris</ligne>  
      <ligne>35, rue Saint-Honoré</ligne>  
    </lignesAdresse>  
    <codePostal>77305</codePostal>  
    <ville>FONTAINEBLEAU Cedex</ville>  
  </adresse>  
  <url>http://www.cri.ensmp.fr</url>  
  <tel t="fixe">01 64 69 47 08</tel>  
  <tel t="fax">01 64 69 48 47</tel>  
</coordonnees>
```

# Processeur XML

Un module logiciel appelé processeur XML est utilisé pour lire les documents XML et fournir un accès à son contenu (avec validation éventuelle).

Un processeur XML effectue le travail pour un autre module, l'application.



# Exemple d'outil: xmllint

— [ 'xmllint': un simple parseur validant ([xmlsoft.org](http://xmlsoft.org))

```
champagny:~$ xmllint --valid adresse-avec-dtd.xml
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE coordonnees SYSTEM "adresse.dtd">
<coordonnees>
  <adresse>
    <lignesAdresse>
      <ligne>Centre de Recherche en Informatique</ligne>
      <ligne>Ecole des mines de Paris</ligne>
      <ligne>35, rue Saint-Honoré</ligne>
    </lignesAdresse>
    <codePostal>77305</codePostal>
    <ville>FONTAINEBLEAU Cedex</ville>
  </adresse>
  <url>http://www.cri.ensmp.fr</url>
  <tel t="fixe">01 64 69 47 08</tel>
  <tel t="fax">01 64 69 48 47</tel>
</coordonnees>
champagny:~$
```

# Objectifs de conception (1)

- [ Utilisation facile sur Internet
- [ Permettre de nombreuses applications
- [ Compatible avec SGML

# Objectifs de conception (2)

- [ Nombre d'options dans XML réduit au minimum, idéalement aucune
- [ Documents XML lisibles par l'homme et raisonnablement clairs
- [ Description de XML formelle et concise

# Objectifs de conception (3)

- [ Facilité d'écriture des programmes traitant les documents XML
- [ Facilité de création de documents XML
- [ Concision dans le balisage importe peu

# Normes associées

— [ Unicode et ISO/CEI 10646 pour les caractères (95000 caractères possibles)

— [ La RFC Internet 1766 pour les noms de langue

— [ La norme ISO 3166 pour les noms de pays

# Format ouvert

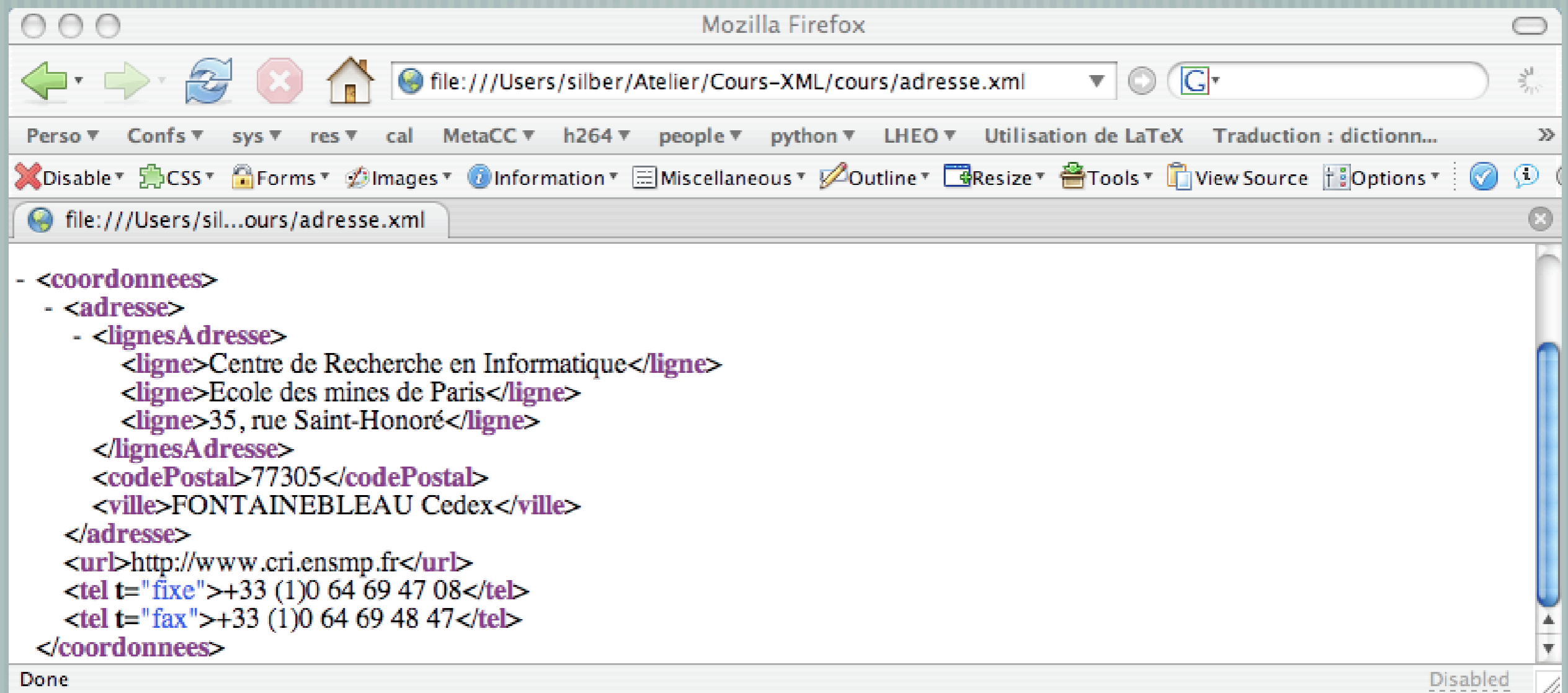
- [ Pas de droits à payer pour l'utilisation de XML
- [ La spécification de XML peut être distribuée librement, à condition que tout le texte et les notices juridiques demeurent intacts

# Outils d'édition

# Voir du XML

— [ Editeur de texte (Emacs, notepad, ...)

— [ Navigateur Web (Firefox, Explorer, ...)



# Edition de XML

— [ Un document XML est du texte, donc n'importe quel éditeur de texte peut être utilisé pour créer et éditer un document XML

— GNU emacs ([www.thaiopensource.com/nxml-mode](http://www.thaiopensource.com/nxml-mode))

— [ Il existe néanmoins des éditeurs XML graphiques pour aider à l'édition

— Xerlin ([www.xerlin.org](http://www.xerlin.org)), open source, écrit en Java

# Xerlin

The screenshot shows the Xerlin XML editor interface. The main window displays a tree view of an XML document named 'adresse-avec-dtd.xml'. The tree structure is as follows:

- coordonnees
  - adresse
    - lignesAdresse
      - ligne
        - #text: Centre de Recherche en Informatique
      - ligne
        - #text: Ecole des mines de Paris
      - ligne
        - #text: 35, rue Saint-Honoré
    - codePostal
      - #text: 77305
    - ville
      - #text: FONTAINEBLEAU Cedex
  - url
    - #text: http://www.cri.ensmp.fr
  - tel (selected)
    - #text: +33 (1)0 64 69 47 08
  - tel
    - #text: +33 (1)0 64 69 48 47

The right-hand pane shows the detailed view of the selected 'tel' element. It includes a type selector set to 'fixe' and a text area containing the value: '#text: +33 (1)0 64 69 47 08'.

# XHTML et CSS

# XHTML et HTML

XHTML est une réécriture de HTML 4.01 sous la forme d'une grammaire XML (DTD)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" >
    ...
</html>
```

# CSS

— [ CSS (Cascading Style Sheets)

— [ Contrôle de la présentation visuelle des pages Web

— [ Support de XML et de HTML

— [ Selon le W3C

— données: XHTML 1.1 (strict)

— présentation: CSS

# Zen Garden

The Beauty of CSS Design

*A demonstration of what can be accomplished visually through CSS-based design. Select any style sheet from the list to load it into this page.*

*Download the sample [html file](#) and [css file](#)*

## *The Road to Enlightenment*

Littering a dark and dreary road lay the past relics of browser-specific tags, incompatible DOMs, and broken CSS support.

Today, we must clear the mind of past practices. Web enlightenment has been achieved thanks to the tireless efforts of folk like the W3C, WaSP and the major browser creators.

The css Zen Garden invites you to relax and meditate on the important lessons of the masters. Begin to see with clarity. Learn to use the (yet to be) time-honored techniques in new and invigorating fashion. Become one with the web.

## *So What is This About?*

There is clearly a need for CSS to be taken seriously by graphic artists. The Zen Garden aims to excite, inspire, and encourage participation. To begin, view some of the existing designs in the list.

Clicking on any one will load the style sheet into this very page. The code remains the same, the only thing that has changed is the external .css file. Yes, really.

CSS allows complete and total control over the style of a hypertext document. The only way this can be illustrated in a way that gets people excited is by demonstrating what it can truly be, once the reins are placed in the hands of those able to create beauty from structure. To date, most examples of neat tricks and hacks have been demonstrated by structurists and coders. Designers have yet to make their mark. This needs to change.

## *Participation*

Graphic artists only please. You are modifying this page, so strong CSS skills are necessary, but the example files are commented well enough that even CSS novices can use them as starting points. Please see the [CSS Resource Guide](#) for advanced tutorials and tips on working with CSS.

You may modify the style sheet in any way you wish, but not the HTML. This may seem daunting at first if you've never worked this way before, but follow the listed links to learn more, and use the sample files as a guide.



select a design:

[45 RPM](#) by Thomas Michaud

[Pretty in Pink](#) by Jordi Romkema

[Vertigo](#) by Antonio Cella

[Vin Rouge](#) by Thorsten Bopp

[Pinups](#) by Emiliano Pennisi

[Zen City Morning](#) by Ray Henry

[Kelmescott](#) by Bronwen Hodgkinson

[Business Style](#) by Gunta Klavina

archives:

[next designs](#) »

[View All Designs](#)

resources:

[View This Design's CSS](#)

[CSS Resources](#)

[FAQ](#)

[Submit a Design](#)

[Translations](#)

# eXtensible Stylesheet Language (XSL)

# Trois parties distinctes

- [ XSL Transformations (XSLT)

- un langage pour transformer du XML

- [ XML Path Language (XPath)

- un langage d'expression utilisé par XSLT pour accéder ou se référer à des morceaux de documents XML

- [ XSL Formatting Objects (XSL-FO)

- une grammaire XML pour spécifier le formatage des documents

# Langage XSLT

— [ XSLT (XML Stylesheet Language Transformations)

— [ Langage permettant de transformer des documents XML en d'autres documents XML

— [ Peut-être vu comme un langage de feuilles de style évolué

— [ Un document XSLT est lui-même en XML



# XSL Transformations (XSLT) Version 1.0

W3C Recommendation 16 November 1999

**This version:**

<http://www.w3.org/TR/1999/REC-xslt-19991116>  
(available in [XML](#) or [HTML](#))

**Latest version:**

<http://www.w3.org/TR/xslt>

**Previous versions:**

<http://www.w3.org/TR/1999/PR-xslt-19991008>  
<http://www.w3.org/1999/08/WD-xslt-19990813>  
<http://www.w3.org/1999/07/WD-xslt-19990709>  
<http://www.w3.org/TR/1999/WD-xslt-19990421>  
<http://www.w3.org/TR/1998/WD-xsl-19981216>  
<http://www.w3.org/TR/1998/WD-xsl-19980818>

**Editor:**

James Clark [<jjc@jclark.com>](mailto:jjc@jclark.com)

Copyright © 1999 W3C® (MIT, INRIA, Keio), All Rights Reserved. W3C [liability](#), [trademark](#), [document use](#) and [software licensing](#) rules apply.

---

## Abstract

This specification defines the syntax and semantics of XSLT, which is a language for transforming XML documents into other XML documents.

XSLT is designed for use as part of XSL, which is a stylesheet language for XML. In addition to XSLT, XSL includes an XML vocabulary for specifying formatting. XSL specifies the styling of an XML document by using XSLT to describe how the document is transformed into another XML document that uses the formatting vocabulary.

XSLT is also designed to be used independently of XSL. However, XSLT is not intended as a complete replacement for XML.

# Principe de XSLT

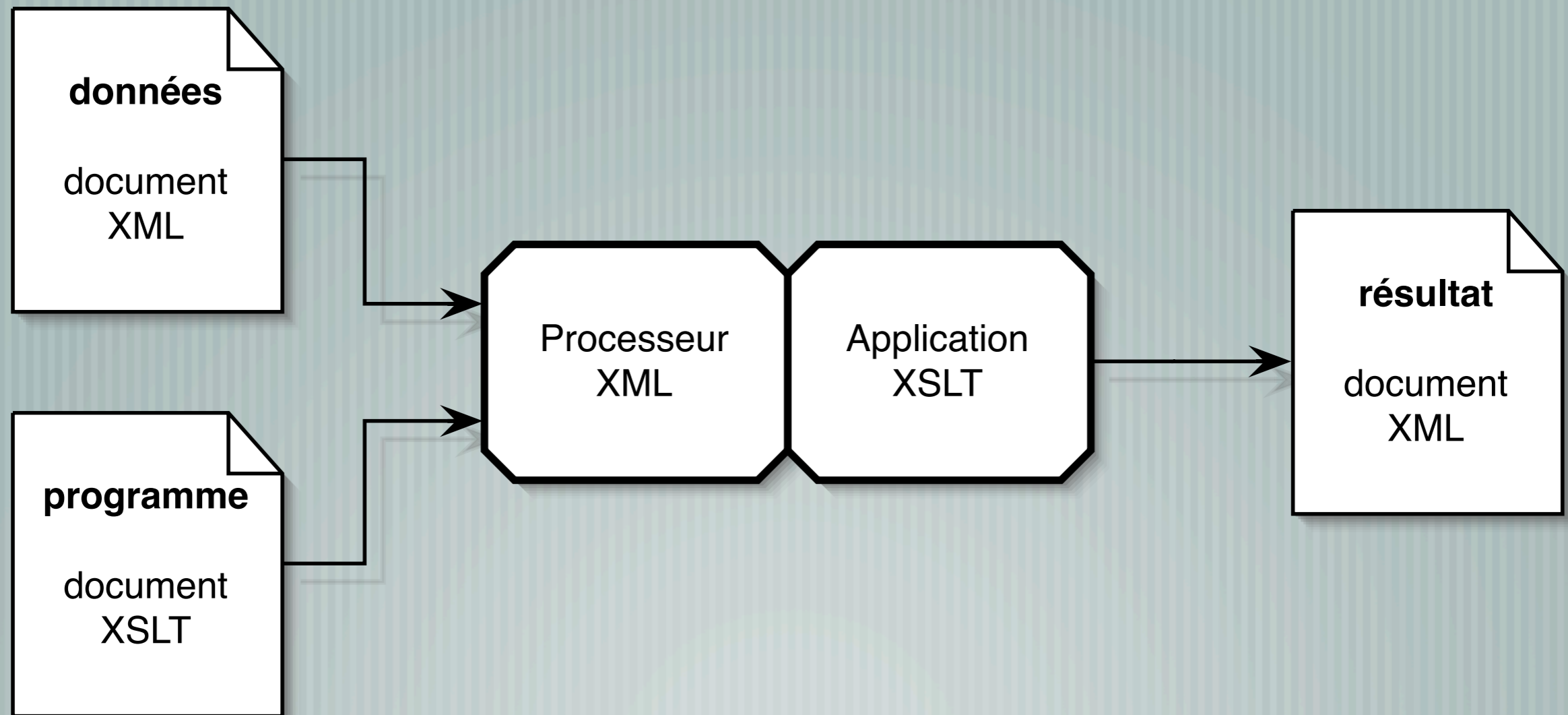
— [ Lecture du document XML source en mémoire

— [ Lecture du document XSLT en mémoire

— [ Application des règles XSLT sur le document source

— [ Obtention d'un résultat

# Chaîne de traitement



# Applications XSLT

— [ Il existe de nombreuses implémentations de XSLT, sous la forme de bibliothèques ou d'outils

— [ Bibliothèques/outils

— libxslt2 ([xmlsoft.org/XSLT](http://xmlsoft.org/XSLT)), outil 'xsltproc'

— xalan ([xml.apache.org/xalan-j](http://xml.apache.org/xalan-j))

— saxon ([saxon.sourceforge.net](http://saxon.sourceforge.net))

— Navigateurs Web modernes

# Utilisation de XSLT

— [ Statiquement, sur le serveur, pour les traitements lourds de documents XML (xsltproc, saxon)

— [ Dynamiquement, sur le serveur pour le XML généré dynamiquement (xalan)

— [ Dynamiquement, sur le navigateur client, avec une instruction de mise en page rajoutée dans le XML

```
<?xml-stylesheet href="style.xsl" type="text/xsl"?>
```

# Exemple: XML vers texte

```
<?xml version="1.0" encoding="iso-8859-1"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  version="1.0">

  <xsl:output method="text"/>

  <xsl:template match="/">
    <xsl:apply-templates select="*" />
  </xsl:template>

  <xsl:template match="tel">
    <xsl:value-of select="@t" />
    <xsl:text>: </xsl:text>
    <xsl:value-of select="text()" />
  </xsl:template>

</xsl:stylesheet>
```

# Exemple: XML vers HTML

```
<?xml version="1.0" encoding="iso-8859-1"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
  <xsl:output method="html" encoding="iso-8859-1"/>
  <xsl:template match="/">
    <html><body><xsl:apply-templates /></body></html>
  </xsl:template>
  <xsl:template match="ligne | ville">
    <code><xsl:value-of select="text()"/></code><br/>
  </xsl:template>
  <xsl:template match="codePostal">
    <code><xsl:value-of select="text()"/></code>
  </xsl:template>
  <xsl:template match="tel">
    <b><xsl:value-of select="@t"/>: </b>
    <code><xsl:value-of select="text()"/></code><br/>
  </xsl:template>
  <xsl:template match="url">
    <a href="{text()}"><xsl:value-of select="text()"/></a><br/>
  </xsl:template>
</xsl:stylesheet>
```

# XSL-FO

- [ Hypothèse: on veut obtenir plusieurs types de documents affichables depuis un document XML (HTML, WML, PDF, texte, RTF, ...)
- [ Plutôt que d'écrire une transformation XSLT par type de document affichable que l'on veut obtenir, il suffit d'écrire une transformation XSLT qui génère du XSL-FO
- [ Il existe ensuite des transformations XSLT toutes faites pour convertir le XSL-FO vers d'autres formats

# Interfaces de programmation XML

# DOM

— [ Document Object Model

— [ Héritage de HTML

— [ Permet de construire un arbre en mémoire depuis un document XML

— [ Méthodes d'accès et de modification de cet arbre

— [ <http://www.w3.org/DOM>

# SAX

— [ Simple API for XML

— [ Lecture séquentielle du document, pas de création d'arbre en mémoire

— [ Association d'actions (fonctions utilisateur) à la lecture des éléments XML, déclenchées au fur et à mesure de la lecture du document

— [ <http://www.saxproject.org>

# Implémentation

— [ Il existe une implémentation de SAX et de DOM dans tous les langages courants

— [ En standard dans JavaEE et dans .NET

# Schémas XML

# XML Schema

— [ Norme du W3C

— [ Permet de représenter une grammaire en XML

— [ Typage des éléments

— [ Contrôle du contenu des éléments et des attributs

— [ <http://www.cri.ensmp.fr/people/silber/xml/adresse.xsd>

# Relax NG

— [ Norme indépendante du W3C

— [ Permet de représenter une grammaire en XML

— [ Contrôle du contenu des éléments et des attributs

— [ Plus simple que les schémas XML

— [ <http://www.cri.ensmp.fr/people/silber/xml/adresse.rng>

— [ <http://www.cri.ensmp.fr/people/silber/xml/adresse.rnc>

# **XML et les bases de données**

# Stockage du XML

— [ Système de fichiers

— [ SGBD Relationnel

— [ XML Natif

# Systeme de fichiers

— [ Fichiers stockés directement sur le système de fichiers

— [ Fichiers stockés dans un SGBD sous la forme de Blobs (Binary large objects) ou de Clogs (Character large objects)

— [ Support de XPath dans certains SGBD comme MySQL

— [ Avantages: implémentation aisée, utile pour un petit volume de données

— [ Inconvénients: l'accès et la mise à jour sont difficiles

# SGBDR

— [ Format de stockage: tables

— [ Avantages: passage à l'échelle, fiabilité et implémentation facile

— [ Inconvénients: requiert beaucoup de jointures à cause de l'éclatement des documents, nécessité de créer des clés artificielles

# SGBDR et polémiques

— [ *Jean-Jacques Thomasson* dans l'Informatique Professionnelle de mars 2005

— [ “L'on voit ainsi apparaître des livres blancs [...] dans lesquels transpirent les velléités des directions qui souhaitent montrer leur modernisme, tout en protégeant leur fond de commerce, à savoir : le marché des bases et entrepôts de données, tas d'or sur lequel ces sociétés sont assises.”

# SGBDR et Polémiques 2

- [ “Ce n’est pas de la science, mais de la protection d’intérêts financiers. L’un et l’autre sont respectables, mais noyer l’un dans l’autre est détestable. Ces consultants ramènent le rôle de XML à celui de format d’échange des données, alors que sa vraie nature est d’être un modèle de structuration de l’information.”
- [ “Le modèle relationnel est intrinsèquement inapte à résoudre la question du système d’information, parce qu’on ne réduit pas la complexité de l’information à la simplicité de tableaux.”

# XML Natif

— [ Base de données supportant de XML natif

— [ Utilisation de modèles de données XML (Schémas)

— [ Avantages: flexibilité, amélioration des performances d'accès

— [ Inconvénients: moins mature que les SGBD classiques

# Produits XML Natif

[www.ipedo.com](http://www.ipedo.com)

[xml.apache.org/xindice](http://xml.apache.org/xindice)

[exist.sourceforge.net](http://exist.sourceforge.net)

[www.softwareag.com/corporate/products/tamino](http://www.softwareag.com/corporate/products/tamino)

[www.dbxml.com](http://www.dbxml.com)

[www.cs.toronto.edu/tox](http://www.cs.toronto.edu/tox)

[www.sleepycat.com/products/xml.html](http://www.sleepycat.com/products/xml.html) (BerkeleyDB XML)

# Web Services

# Définition en 5 points

- [ Service disponible sur Internet (ou sur un {ex,in}tranet)
- [ Utilise un système de messages XML standardisé
- [ N'est lié à aucun système d'exploitation/matériel
- [ S'auto-décrit grâce à une grammaire XML
- [ Peut-être découvert grâce à un mécanisme de recherche

# Technologies

— [ Technologies internet (TCP/IP, HTTP, SMTP, FTP) pour le service disponible sur Internet

— [ Messagerie XML: XML-RPC, SOAP

— [ Description du service: WSDL

— [ Recherche du service: UDDI

# Messagerie

— [ XML-RPC ([www.xmlrpc.com](http://www.xmlrpc.com))

— [ Protocole simple, types de données simples, pas un standard W3C

— [ SOAP (Simple Object Access Protocol)

— [ Norme du W3C, séparation enveloppe/données, types de données complexes

# Description

— [ WSDL (Web Service Description Language)

— [ Langage XML de description de Web Service

— [ Standard du W3C utilisant SOAP et les Schémas XML

# UDDI

— [ UDDI ([www.uddi.org](http://www.uddi.org))

— [ Universal Description, Discovery and Integration

— [ Service permettant de rechercher des services et de récupérer automatiquement leur description (par exemple en WSDL)

# Examples

— [ Google

— <http://api.google.com>

— [ NEOS Server for optimization

— <http://www-neos.mcs.anl.gov/neos>

— [ Amazon

— <http://www.amazon.com/gp/aws/landing.html>

# Bibliographie

