

# Présentation XML

Voici à quoi ressemble un document XML :

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Cours SYSTEM "Cours.dtd">
<Cours>
  <Titre>Cours XML</Titre>
  <Auteur>
    <Nom>Poulard</Nom>
    <Prénom>Philippe</Prénom>
  </Auteur>
  <Description>Ce cours aborde les concepts de base
    mis en œuvre dans XML.
  </Description>
</Cours>
```

C'est presque comme du HTML !

Sauf que :

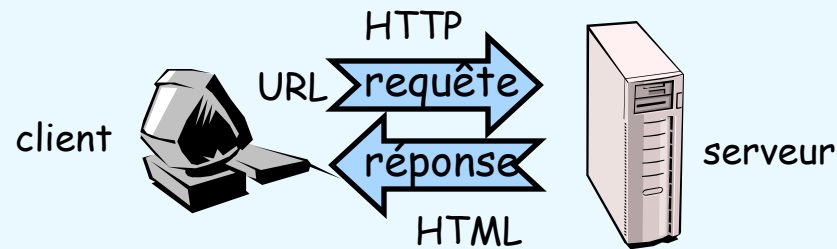
- j'ai créé **mes propres balises** pour mes propres besoins
- je suis **contraint** par une **syntaxe rigoureuse** qui n'accepte pas les erreurs

Et que :

- ce document **ne sait pas comment s'afficher** dans un navigateur

Mais que :

- je dispose d'une **boîte à outil** pour manipuler les **informations** de ce document



- Le Web peut transmettre n'importe quel type de document
- HTML s'est imposé comme le "format natif" du Web
- HTML a contribué au succès du Web
- HTML n'est qu'un **format de présentation**
- XML est un **format de structuration de l'information**

## Rappels HTML

- Format ASCII-7 bits
- Une application SGML
- Simple
- Editable à la main
- Inclusion d'objets
- Liens hypertextes
- Structure documentaire
- Certains mauvais choix de conception : h1, h2, h3, h4, h5, h6, br
- Syntaxe laxiste → lourdeur du code des navigateurs
- Appropriation par les éditeurs (Netscape, Microsoft) → compatibilité difficile

```
<html>
  <head><title>Cours XML</title></head>
  <body>
    <p><b>Poulard</b> Philippe
    <p>Ce cours aborde les concepts
    de base mis en &oeelig;uvre dans XML.
  </body>
</html>
```

Une application sur le poste utilisateur,  
le **navigateur** (Netscape, Internet Explorer)

Un client chez lui

Demande des documents à une machine distante,  
le **serveur**

Commande un meuble

À une adresse (**URL**) `http://www.inria.fr`

Chez ACME

Via le protocole de transport **HTTP**

Par la poste

Elle reçoit un document,  
en **HTML** (un "mode d'emploi")

En kit, avec le mode d'emploi

Pour s'afficher sur l'écran  
grâce à des "balises" ou "tag"

Exprimé dans un langage connu (le français)  
et qu'il peut interpréter  
avec ses outils (tournevis, marteau)

Les balises décrivent l'effet  
"à peu près" souhaité



Rendus différents selon les navigateurs  
Rendus contrôlés par les navigateurs

Il y en a un nombre défini et fixe : on sait de quoi on parle

Marquage inclus dans le texte (*embedded text markup*)

```
foo <b>bar</b>
```

**Données  
auto-descriptives**

## Autres solutions

• 2 fichiers séparés

```
bold: de 5 à 7
```

```
1234567
```

```
foo bar
```

• Fichier binaire

```
foo *■bar
```

ou

```
"foo " 27 3 "bar"
```

• Descriptif

```
(normal "foo ") (bold "bar")
```

ou

```
normal:foo ↵  
bold:bar
```

• Etc...

Vocabulaire de balises fixes : les utilisateurs font du meccano pour exprimer ce qu'ils veulent

Navigateurs : course aux tags

- balises hétéroclites `center`, `blink`, `frames`...
- non-HTML : plug-in (flash, son...)

De plus en plus d'information disponible en HTML qu'on essaye de retraiter pour en extraire les informations

→ ça casse dès que les sites changent leur code HTML

Le Web en environnement professionnel, ce n'est plus du bricolage

1960	GML : Généralized Markup Language, IBM
1986	SGML : Standard Generalized Markup Language, ISO 8879, 1986 Meta-langage (permet de définir des langages de balisage) Utilisation industrielles lourdes : catalogues, documentations techniques
1992	HTML : HyperText Markup Language, CERN (Tim Berners-Lee) Une application de SGML, base du Web
Oct. 1994	Création du W3C ←
Juin 1996	Groupe de travail "SGML-light"
Fév. 1998	<b>XML 1.0</b> , recommandation du W3C
Sept. 1998	Groupe de travail W3C XML divisé en 5 groupes  Toutes les nouvelles normes seront en XML
1999-2000	Diversification des usages (web, database...), XML partout là où on ne l'attendait pas  Augmentation de la complexité (toutes les normes prennent du retard)
2001	28 groupes de travail au W3C

SGML

Niche technologique

La communauté SGML a peur de ne pas profiter de l'essor de HTML

W3C

Le Web est un mélange de présentation et de structure

But commun :  
Re-aiguiller le Web vers un  
successeur de SGML:  
XML = SGML --

Groupe de travail constitué au sein du W3C par des gens de la communauté SGML

Conception rapide et homogène

XML, c'est SGML arrivé à maturité

XML n'est donc pas une techno JEUNE



- utilisable simplement sur internet
- internationalisation (ASCII + Unicode)
- applicable à un maximum de domaines
- compatible avec SGML
- facilité de traitement
- sans options (contrairement à SGML)
- document lisible par un humain
- spécification XML disponible rapidement
- spécification formelle et concise
- documents XML faciles à créer, même à la main
- ne cherche pas la concision du document

Pas de codage binaire  
Se charge dans un bloc-note

(les documents sont d'ailleurs très verbeux)

```
<Titre>Cours XML</>
<Auteur>
  <Nom>Poulard</>
  <Prénom>Philippe</>
</>
```

et non !

On aurait pu être plus concis...

<http://w3.org>



Organisme à but non lucratif, 500 membres, 60 collaborateurs  
3 centres :

MIT (Boston, USA)  
INRIA (Sophia-Antipolis, France)  
Keio University (Keio, Japon)



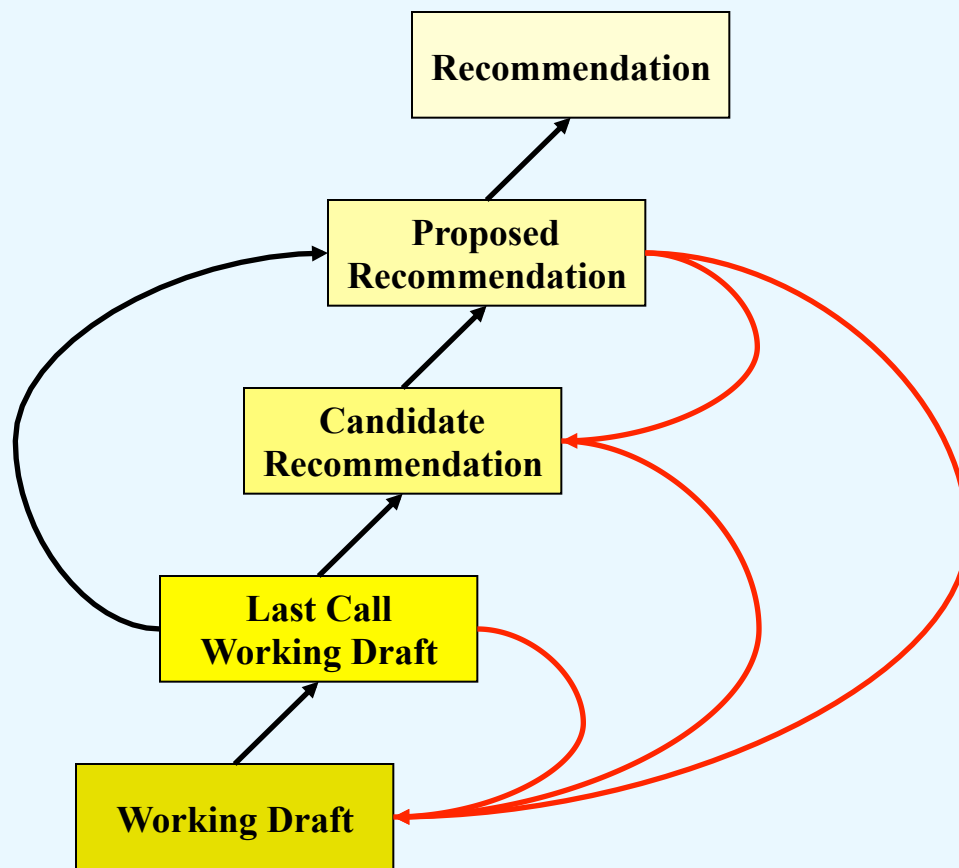
Ce qu'apporte le W3C :

- Arbitrage neutre (éditeurs de logiciel, utilisateurs)
- Expertise technique
- Processus de compromission
- Producteur de recommandations
- Producteur d'implémentations de référence (jigsaw, amaya...)

4 domaines :

- Architecture : XML, URL/URI, DOM, SOAP...
- User interface : HTML, SVG, PNG, device independance, MathML, CSS, Multimedia (SMIL), VoiceML...
- Technology & society : commerce électronique, confidentialité, web sémantique, XML encryption et signature
- WAI : Web Accessibility Initiative

### Cycle de production des standards au W3C

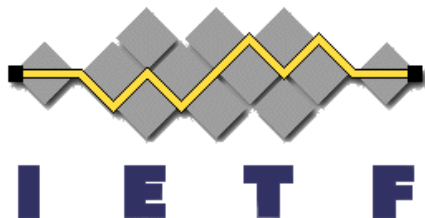


<http://www.w3.org/TR>

(standardisation)

(liste non exhaustive !)

- W3C Standards XML World Wide Web Consortium <http://w3.org>
- ISO SGML International Organization for Standardization <http://www.iso.ch>
- IETF RFCs The Internet Engineering Task Force <http://www.ietf.org/>
- Unicode Caractères <http://www.unicode.org>
- IANA Registres Internet Assigned Numbers Authority <http://www.iana.org/>
- OASIS Catalogues SGML <http://www.oasis-open.org/>
- ECMA ECMAScript European Computer Manufacturers Association <http://www.ecma.ch>



ECMA



XML ne sait pas faire grand chose tout seul, à part servir pour les échanges

Standards complémentaires

Présentation XML

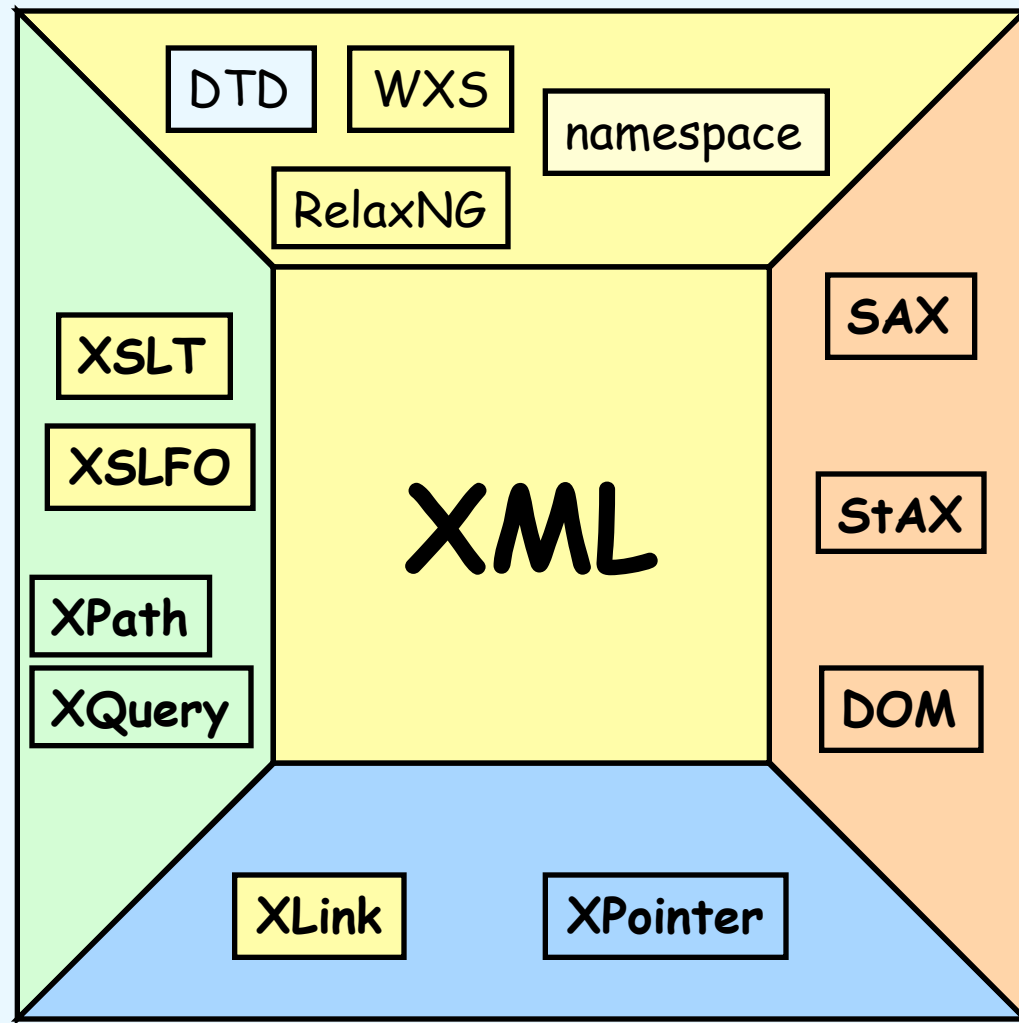
transformation et formattage

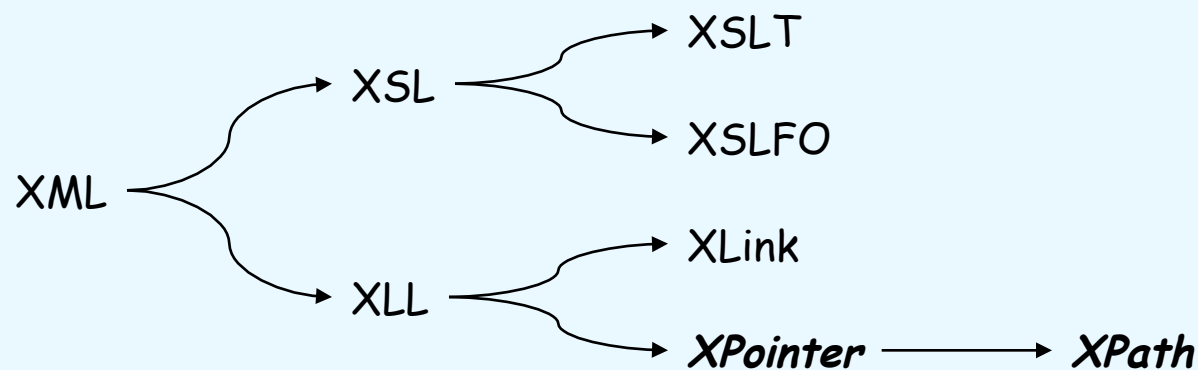
requête

definition

manipulation

liaison





*DOM*  
*SAX*

*Namespace*

*XML Infoset*

*DTD*

*WXS*

*Relax-NG*

*XForms*

*XQuery*

...

Applications de présentation :

- XHTML
- WML
- XSLFO
- SVG
- SMIL

Applications d'échange :

- SOAP
- UDDI

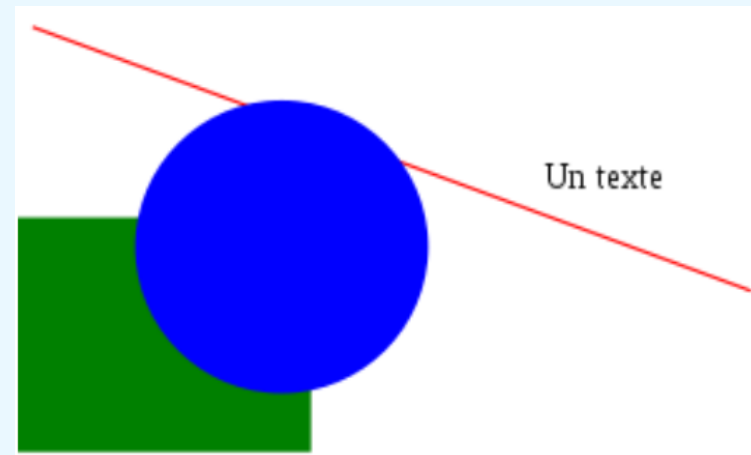
Applications sectorielles :

- MathML
- CML
- DocBook

Applications de description :

- RDF
- RSS
- WSDL

```
<?xml version="1.0" encoding="utf-8"?>
<svg xmlns="http://www.w3.org/2000/svg"
  version="1.1"
  width="300"
  height="200">
  <title>Exemple simple de figure SVG</title>
  <desc>
    Cette figure est constituée d'un rectangle,
    d'un segment de droite et d'un cercle.
  </desc>
  <!-- ceci est un commentaire-->
  <rect
    width="100" height="80"
    x="0" y="70"
    fill="green" />
  <line
    x1="5" y1="5"
    x2="250" y2="95"
    stroke="red" />
  <circle
    cx="90" cy="80"
    r="50"
    fill="blue" />
  <text x="180" y="60">
    Un texte
  </text>
</svg>
```



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

```
<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>
```

```
<mrow>
```

```
<mn>2</mn>
```

```
<mo>&InvisibleTimes;</mo>
```

```
<mi>a</mi>
```

```
</mrow>
```

```
</mfrac>
```

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

LaTeX

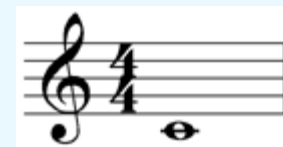
```
x = \frac {-b \pm \sqrt{b^2 - 4ac}}{2a}
```



```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE score-partwise PUBLIC
  "-//Recordare//DTD MusicXML 2.0 Partwise//EN"
  "http://www.musicxml.org/dtds/partwise.dtd">
<score-partwise version="2.0">
  <part-list>
    <score-part id="P1">
      <part-name>Music</part-name>
    </score-part>
  </part-list>
  <part id="P1">
    <measure number="1">
      <attributes>
        <divisions>1</divisions>
        <key>
          <fifths>0</fifths>
        </key>
        <time>
          <beats>4</beats>
          <beat-type>4</beat-type>
        </time>
        <clef>
          <sign>G</sign>
          <line>2</line>
        </clef>
      </attributes>
      <note>
        <pitch>
          <step>C</step>
          <octave>4</octave>
        </pitch>
        <duration>4</duration>
        <type>whole</type>
      </note>
    </measure>
  </part>
</score-partwise>

```



Affichage dans Internet Explorer → représentation arborescente du document  
(pas vraiment utile pour un utilisateur)



Association d'une feuille de style à un document XML

- CSS (très rudimentaire)
- XSLT (beaucoup plus puissant)

Les CSS (Cascading Style Sheets)

Permet d'associer une présentation à chaque élément d'un document

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Fable>
  <Titre>Le corbeau et le renard</Titre>
  - <Auteur>
    <Nom>De la Fontaine</Nom>
    <Prénom>Jean</Prénom>
  </Auteur>
  - <Contenu>
    <p>Maître corbeau sur un arbre
      perché, tenait dans son bec un
      fromage.</p>
  </Contenu>
</Fable>
```

Problème CSS :

- s'applique seulement aux éléments
- dans l'ordre du document

```
<html>
  <head>
    <title>Fable</title>
  </head>
  <body>
    <h1>Le corbeau et le renard</h1>
    <p><i>Jean De la Fontaine</i></p>
    <p>Maître corbeau sur un arbre perché,
tenait dans son bec un fromage.</p>
.../...
  </body>
</html>
```

format de présentation  
HTML

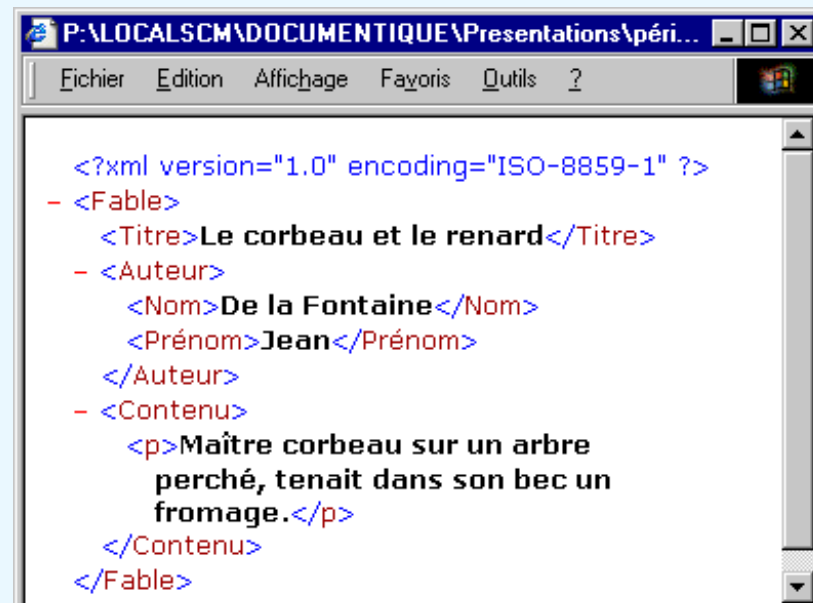
XML

format de structuration sémantique

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE Fable SYSTEM "Fable.dtd">
<Fable>
  <Titre>Le corbeau et le renard</Titre>
  <Auteur>
    <Nom>De la Fontaine</Nom>
    <Prénom>Jean</Prénom>
  </Auteur>
  <Contenu>
    <p>Maître corbeau sur un arbre perché,
tenait dans son bec un fromage.</p>
.../...
  </Contenu>
</Fable>
```



↑  
XSLT  
↑



## Document HTML

jeu de balises fini et connu

structure orienté présentation exclusivement

wysiwyg qui dépend du navigateur (!!)

réutilisation limitée (par la quantité de documents HTML)

cible : web exclusivement

écriture anarchique, pas de respect de charte graphique possible, temps perdu alloué à la présentation

stockage : SGF

## Document XML

jeu de balises extensible, à priori non connu

toutes structures (sémantique) données auto-descriptives

pas de présentation intrinsèque possible : obligation de feuille de style

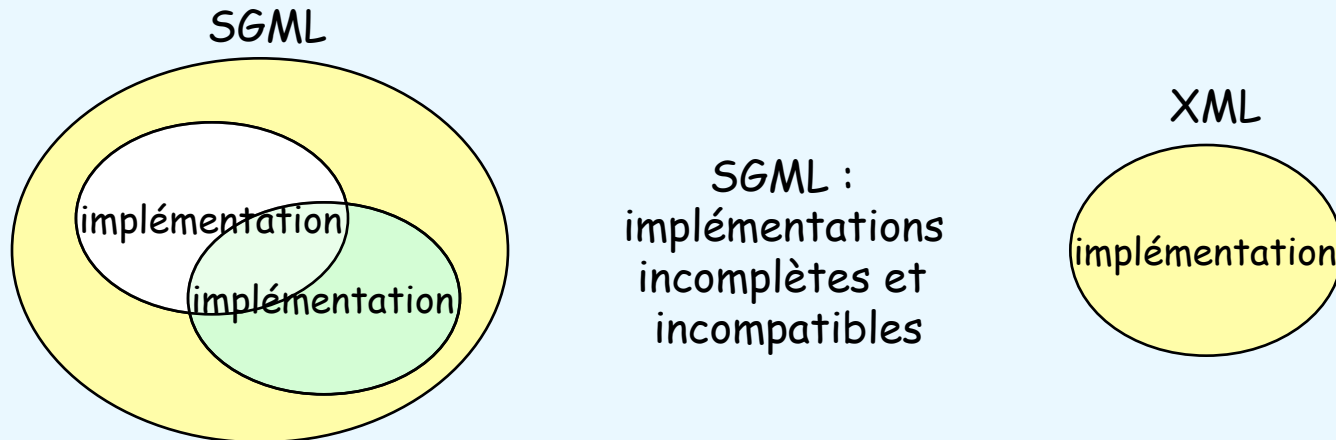
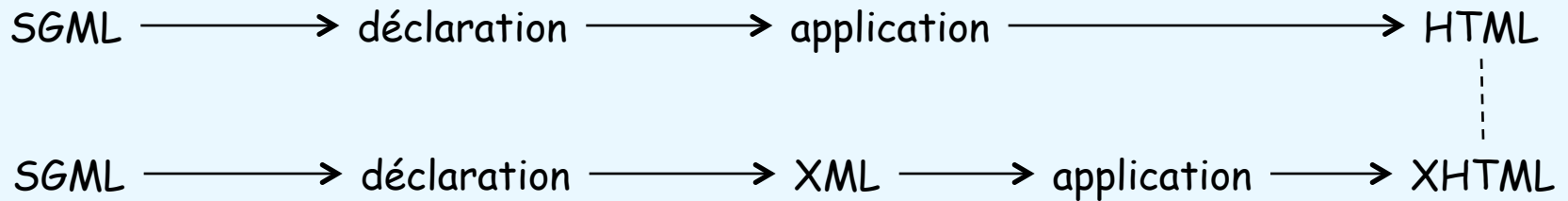
exploitation sémantique possible, transformation de structure

cible : Web, Wap, PDF, texte, autres échanges entre applications

rédaction de contenu exclusivement

stockage : SGF, BdDR, natif (XND)

XML est un sous-ensemble de SGML



Profils d'applications

Exemple pour les feuilles de style : DSSSL, FOSI os CALS-B

**XML :**

- Déclaration fixe (syntaxe fixe)
- Infiniment plus simple à analyser (*parser*) que SGML