

# Séance 10

## XML: Architecture et Ingénierie

**Prof. Yassin Aziz REKIK**  
[Yassin.rekik@he-arc.ch](mailto:Yassin.rekik@he-arc.ch)



- **Cours de : [Philippe.Poulard@sophia.inria.fr](mailto:Philippe.Poulard@sophia.inria.fr)**
  - INRIA
- **Cours : Stéphane Crozat et Bruno Bachimont**
  - UTC Formation
- **Cours : Christine Vanoirbeek**
  - EPFL
- **Divers supports en ligne**



- **Ingénierie Documentaire XML**
  - Chaînes d'édition
  - Gestion de contenu
  - Recherche d'information
  - Exemple : édition Web basée XML
    - *Cocoon*
- **Application et Architectures Web basées XML**
  - Les applications métiers
  - Les formats pivots
  - Les Web Services
  - Exemple :
    - *Multimédia*
    - *Protocoles*

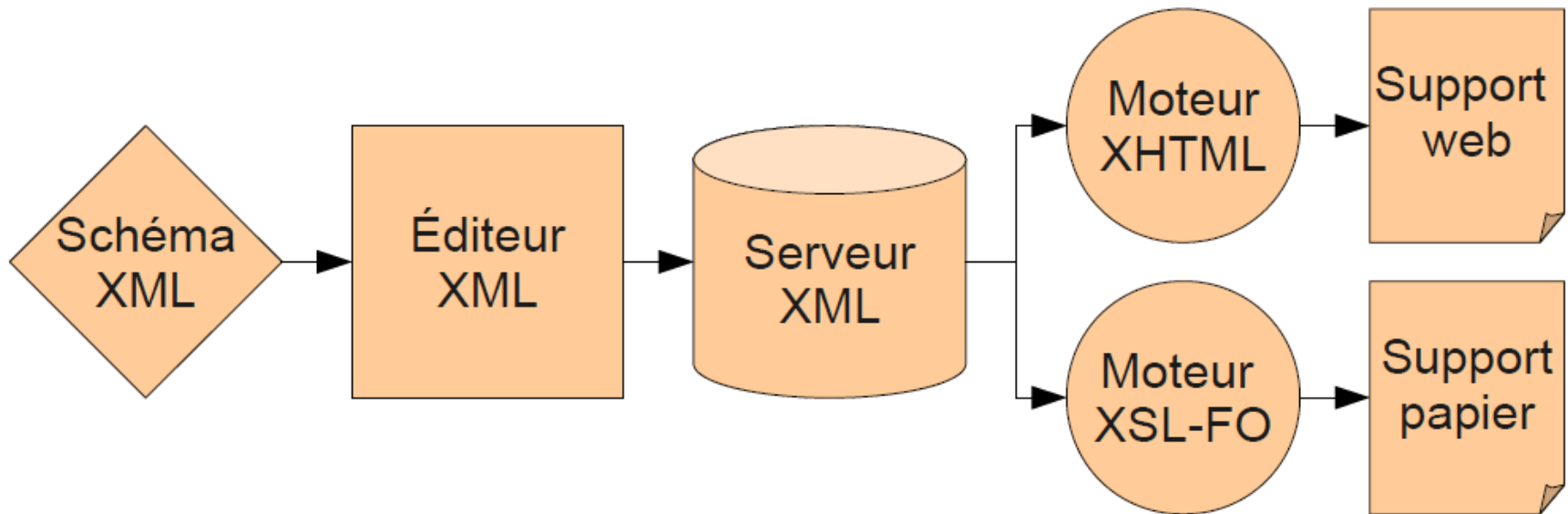


# Edition électronique

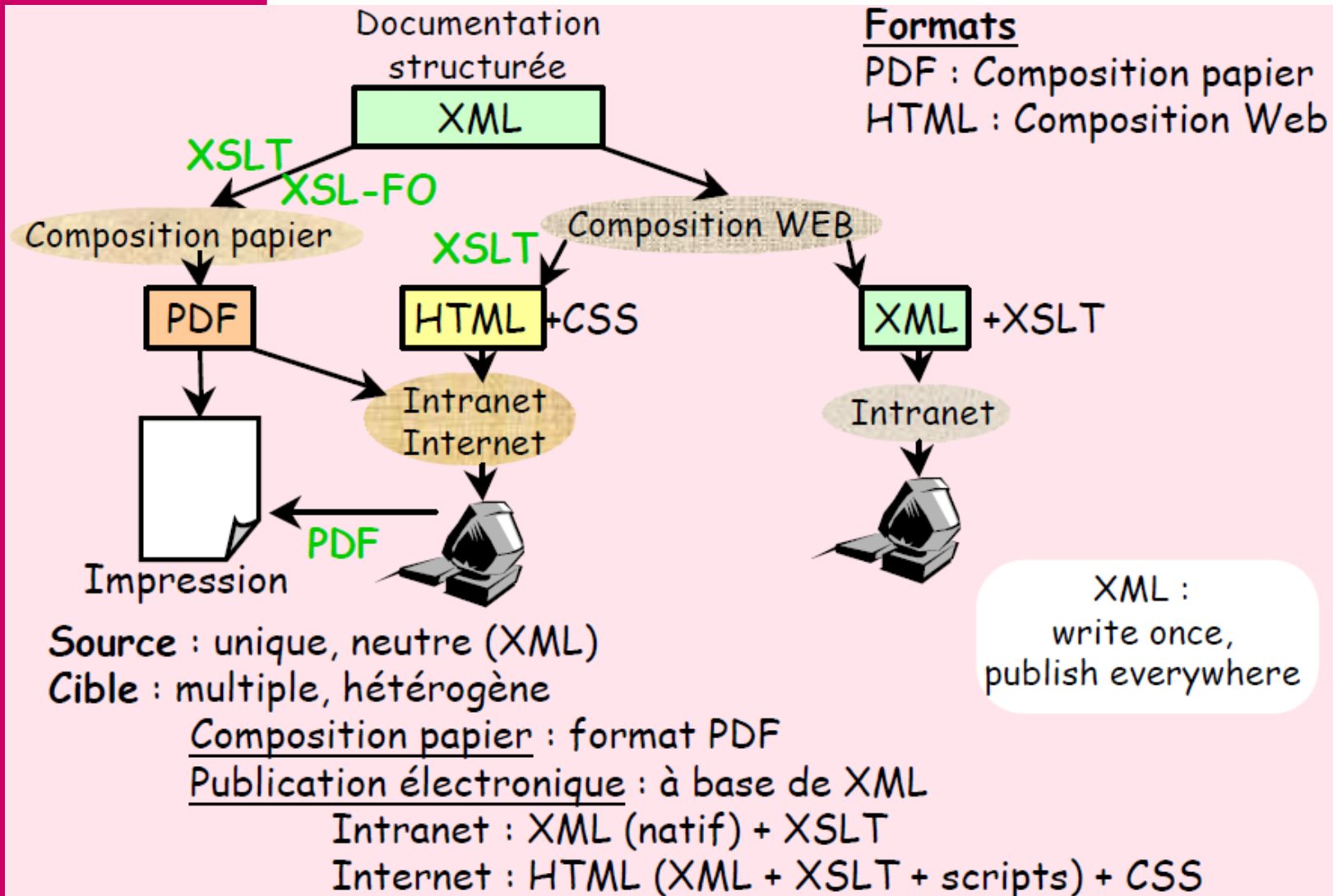


- **Une chaîne éditoriale numérique est un procédé de création de contenu multimédia qui, à la différence des outils bureautiques typiquement, ne fixe pas la mise en forme graphique du contenu a priori.**
- **La chaîne éditoriale permet en effet d'éditer une forme neutre par rapport à la mise en forme, et d'ajouter au contenu :**
  - des descripteurs de structure (la partie B est incluse dans la partie A)
  - des descripteurs sémantiques (la partie A décrit une procédure, la partie B est une définition, ...).
- **Avantages :**
  - Traitement automatiquement de forme neutre
  - Adaptation au contexte : support, public, modalité, ...

- **Architecture classique**



# Publication polymorphe

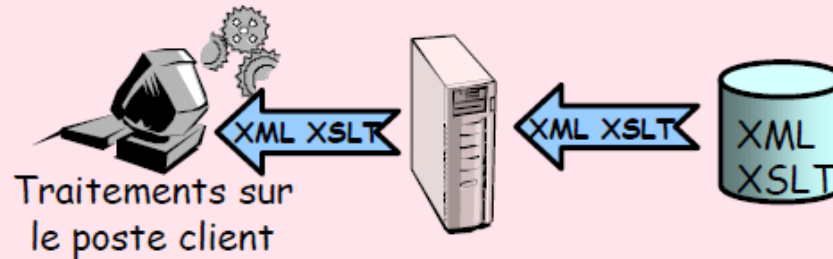


XML est orienté **structure**.  
Pour le publier on utilise un  
procédé de transformation  
approprié au média de  
diffusion : **XSLT**  
(eXtensible Stylesheet  
Language - Tranformations)

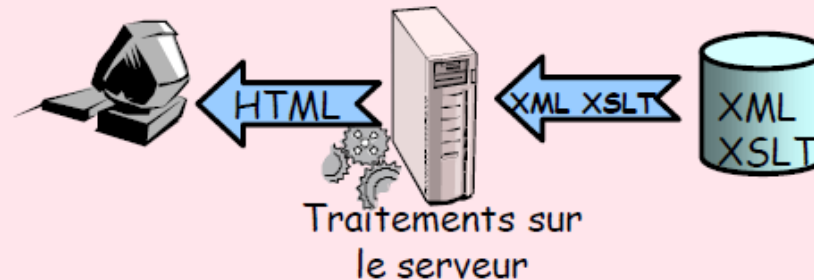
XSLT est un langage de  
programmation qui utilise  
une grammaire XML.

3 méthodes pour  
disposer d'un format  
consultable

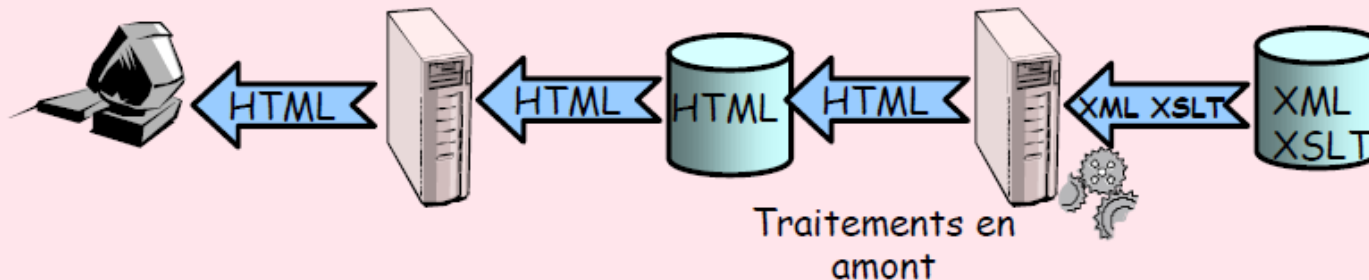
## Transformation sur le poste client



## Transformation à la volée sur le serveur



## Transformation et stockage sur le serveur







## Transformation statique

### Avantages :

- Utilisation d'un serveur Web classique
- Publication à emporter (sur cédérom)
- Snapshot du fonds documentaire
- Production des publications en Batch
- Robustesse

### Inconvénients :

- Gestion des publications
- Stockage des publications

## Transformation à la volée

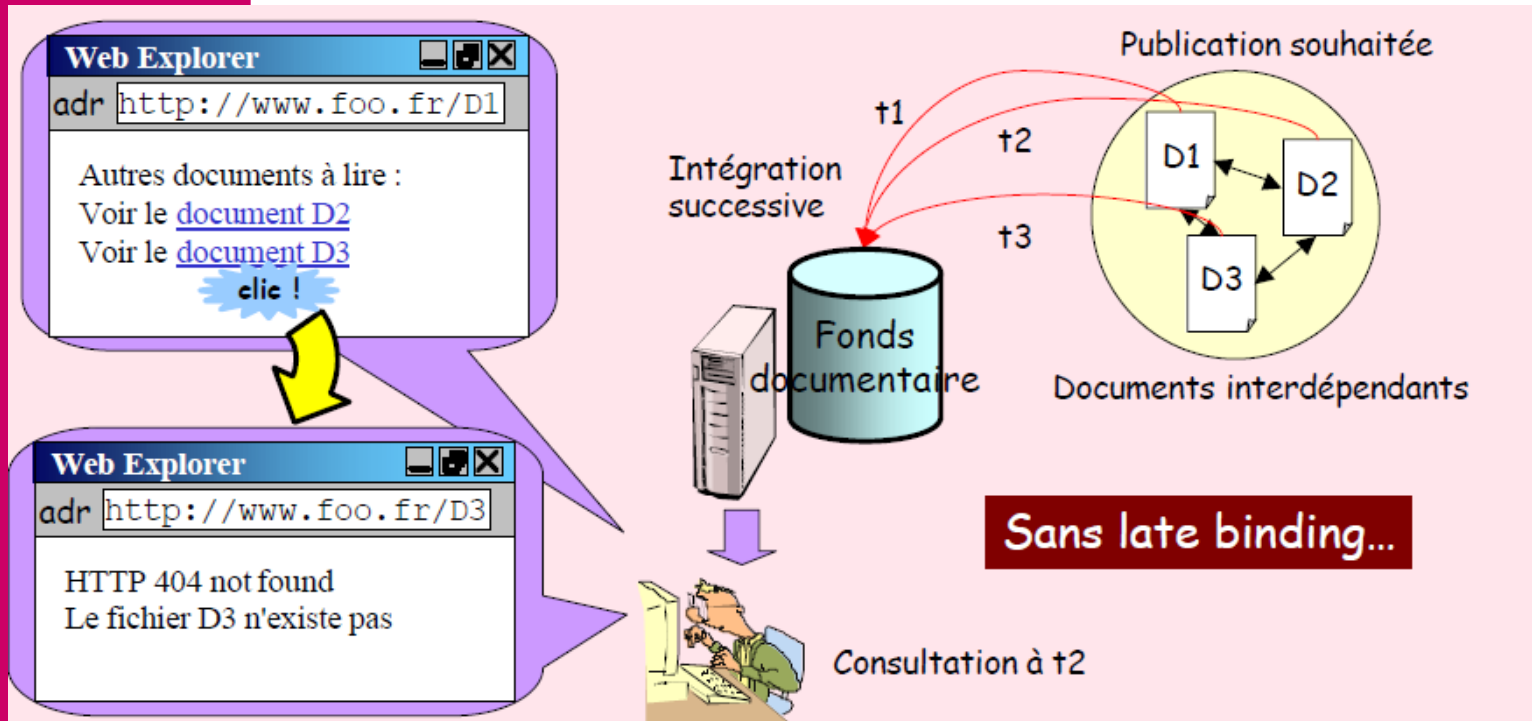
### Avantages :

- Late binding
- Profiling possible
- Interactivité plus grande
- Pas de gestion des fichiers produits
- Qui peut le plus peut le moins

### Inconvénients :

- Publication non portable
- Attention aux temps de réponse
- Requiert un serveur robuste

# Late binding



Si les documents sont intégrés les uns après les autres, il y aura, selon la stratégie de publication :

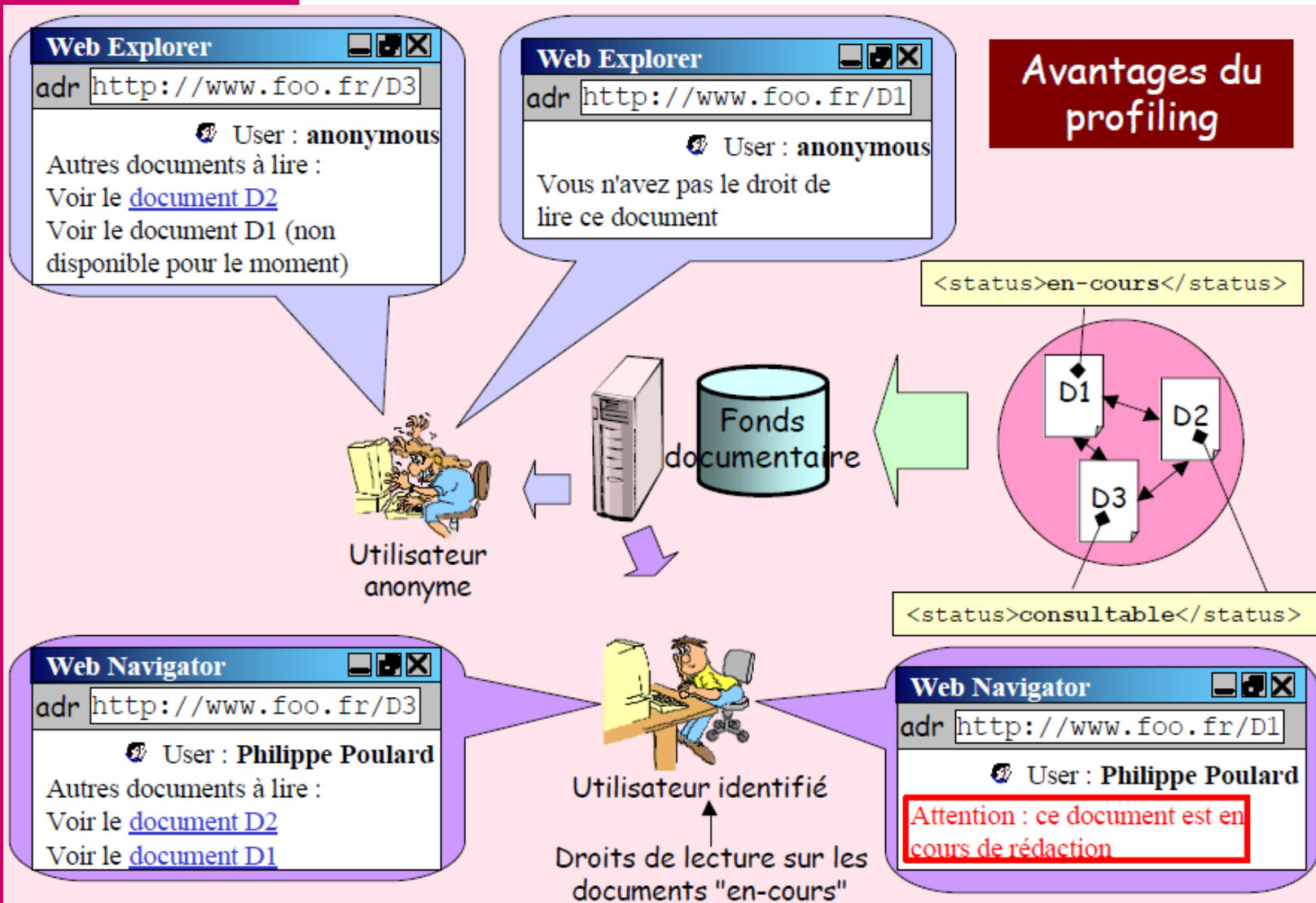
- des liens non résolus
- ou des liens cassés

ou

Nécessité de générer la publication :

- à chaque fois qu'un document est intégré
- ou à la fin quand tous les documents ont été intégrés le fonds documentaire

# Profiling



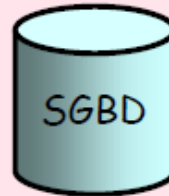
Permet de distribuer une vue du document en fonction du profil de l'utilisateur

# Agrégation des données



Sources :  
•XML  
•non XML

Données stockées



Flux



Application



PGI

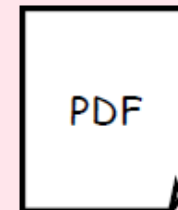
## Intérêt

- Séparation des points de vues (les préoccupations des web-designers ne sont pas les mêmes que celles des programmeurs)
- Découpler la logique métier de la logique de présentation
- Intéropérabilité
- Universalité

XML

```
<?xml version="1.0"
encoding="ISO-8859-1" ?>
<!DOCTYPE rapport
SYSTEM "rapport.dtd" [
<!ENTITY ventes
SYSTEM "ventes.xml">
]>
<rapport>
  <ventes>
    <analyse>
      <p>Cette année, la croissance
est au rendez-vous,
...</p>
    </analyse>
  </rapport>
```

Application



WML  
WAP





# Gestion des contenus



- **Gestion électroniques de documents (GED)**
  - Les fonctions de gestion concernent typiquement les droits des utilisateurs, la formalisation de processus (workflow) éditoriaux (création, relecture, validation, diffusion, etc.), la centralisation et la sécurisation des documents, la gestion de la concurrence.
  - Initialement ces systèmes misent massivement sur la bureautique pour la création et la génération PDF pour la diffusion.
  - Mais la généralisation du Web et la convergence des systèmes, conduit de plus en plus les GED à se rapprocher des Web CMS, et par la même des Wiki.



- **Gestion électroniques de documents (GED)**
  - Le domaine est historiquement dominé par de gros éditeurs :
    - *Documentum (ECM)*
    - *FileNet (IBM)*
    - *SharePoint (Microsoft)*
    - *Open Text*
    - *Stellent (Oracle)*
  - Mais de nouveaux acteurs émergent, notamment issus du monde Open Source :
    - *Alfresco : éditeur Open Source le plus actif actuellement au niveau international*
    - *Nuxéo : éditeur Open Source français en croissance*



- **CMS et WCMS**

- Un Web CMS (WCMS ou CMS ou SGC) est un outil de production de pages HTML, généralement adossé à une base de données relationnelle, à des logiques de gabarits de pages (*templates*) et des feuilles de style CSS.
- Cette classe d'outils née au tout début des années 2000 a permis de démocratiser la réalisation de sites Web et d'Intranets d'entreprise, en délestant les auteurs de contenus des tâches techniques d'encodage HTML et de programmation client (JavaScript) ou serveur (PHP généralement)





- **CMS et WCMS**

- La notion de « chaîne d'édition » :  
De nouveau on retrouve des rôles éditoriaux  
Rapprochement avec les Wiki
- La notion de groupeware  
Applications et documents partagés  
Importance de la définition du Workflow

# Web Sémantique

# La vision du web sémantique



- **"Avoir des données et des informations décrites d'une manière qui facilite leur réutilisation et leur composition par les machines à travers différentes applications." [Laublet, Ontosaurus, 2002]**

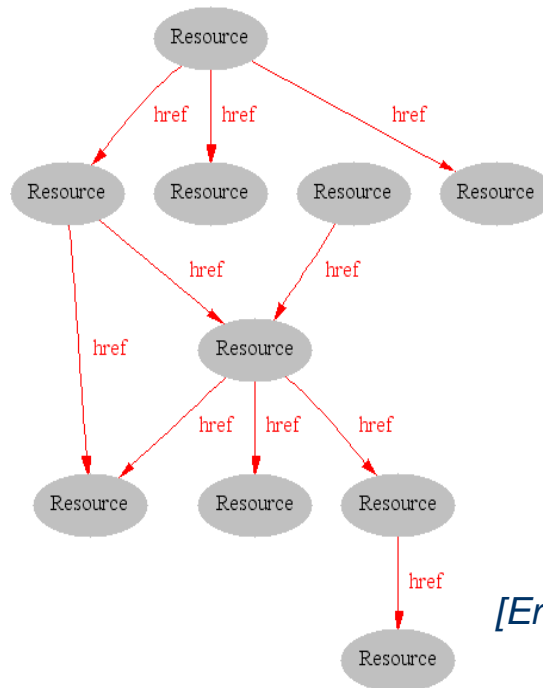


*[Berners-lee, Hendler and Lassila, The semantic web, Scientific American, 17 Mai 2001]*

# Le web d'aujourd'hui



- Un ensemble de ressources et de liens
- Les ressources sont désignées par des URIs
- Les documents sont adaptés aux humains, pas aux machines
- La structure des documents est explicite pour un anglophone (header, body, paragraph, etc.)
- La signification des liens dépend du contexte



*[Eric Miller, W3C]*

# Le web demain



- **Un espace d'échange de ressources entre machines permettant l'exploitation de grands volumes de données et de services variés.**
  - **Simplifier le rôle de l'utilisateur dans la recherche et l'exploitation de l'information grâce à :**
    - de meilleurs moteurs de recherche
    - l'intégration facilitée des sources de données
    - la découverte, l'exploitation et la combinaison de services
    - des possibilités accrues de raisonnement sur les données
- [Laublet,2002]

# Comment y arriver ?

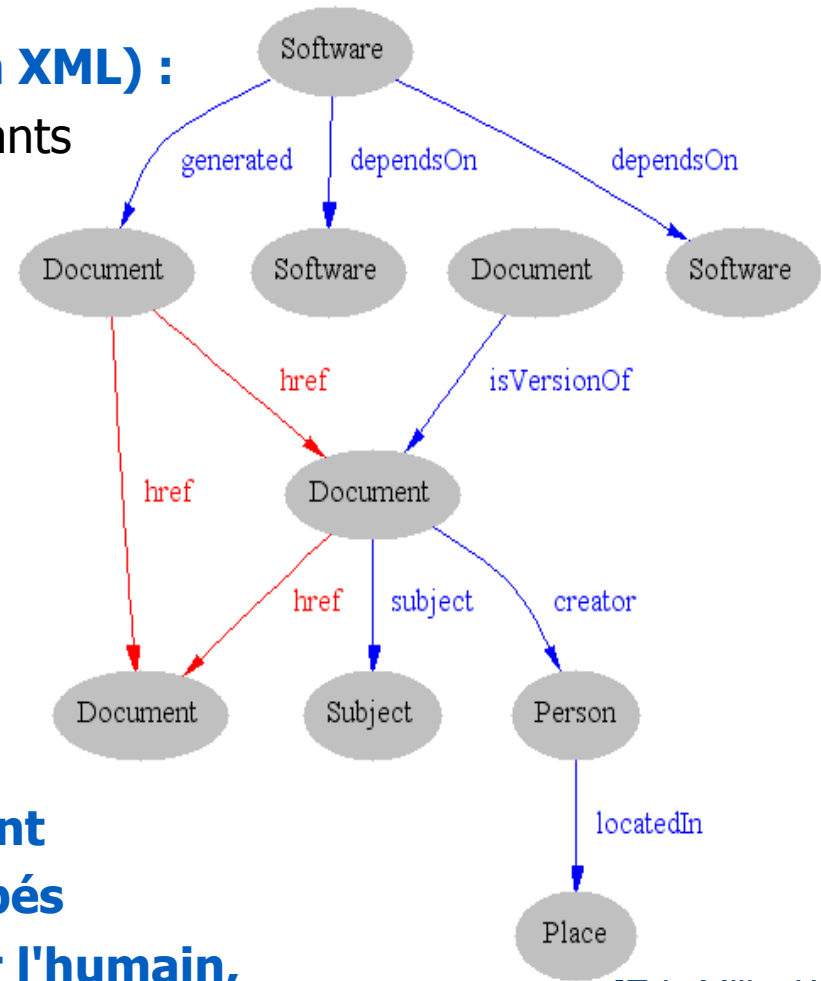


- **Apprendre la sémantique à la machine (IA/KR)**
  - structure du document
  - dictionnaires
  - connaissances sur la langue
  - etc.
- **Ajouter des informations sémantiques destinées aux machines (SW)**
  - Ajouter des concepts, relations, prédicats.
  - Ajouter des règles d'inférence

# Le web sémantique



- **Web +**
- **Méta-données (en XML) :**
  - RDF et descendants
  - Ontologie



[Eric Miller, W3C]

- **👉 Les liens peuvent être fortement typés**
- **👉 Une vision pour l'humain, une autre pour la machine**



- **Identification, exploitation et combinaison intelligente et automatique de ressources (documents, services)**
- **Traitements automatisés des données et des connaissances**
- **Gestion de l'hétérogénéité**
- **Gestion de l'interaction avec l'utilisateur**

[Charley, Laublet, Reynauld, AS 32 Web Sémantique, 18 Juin 2002]

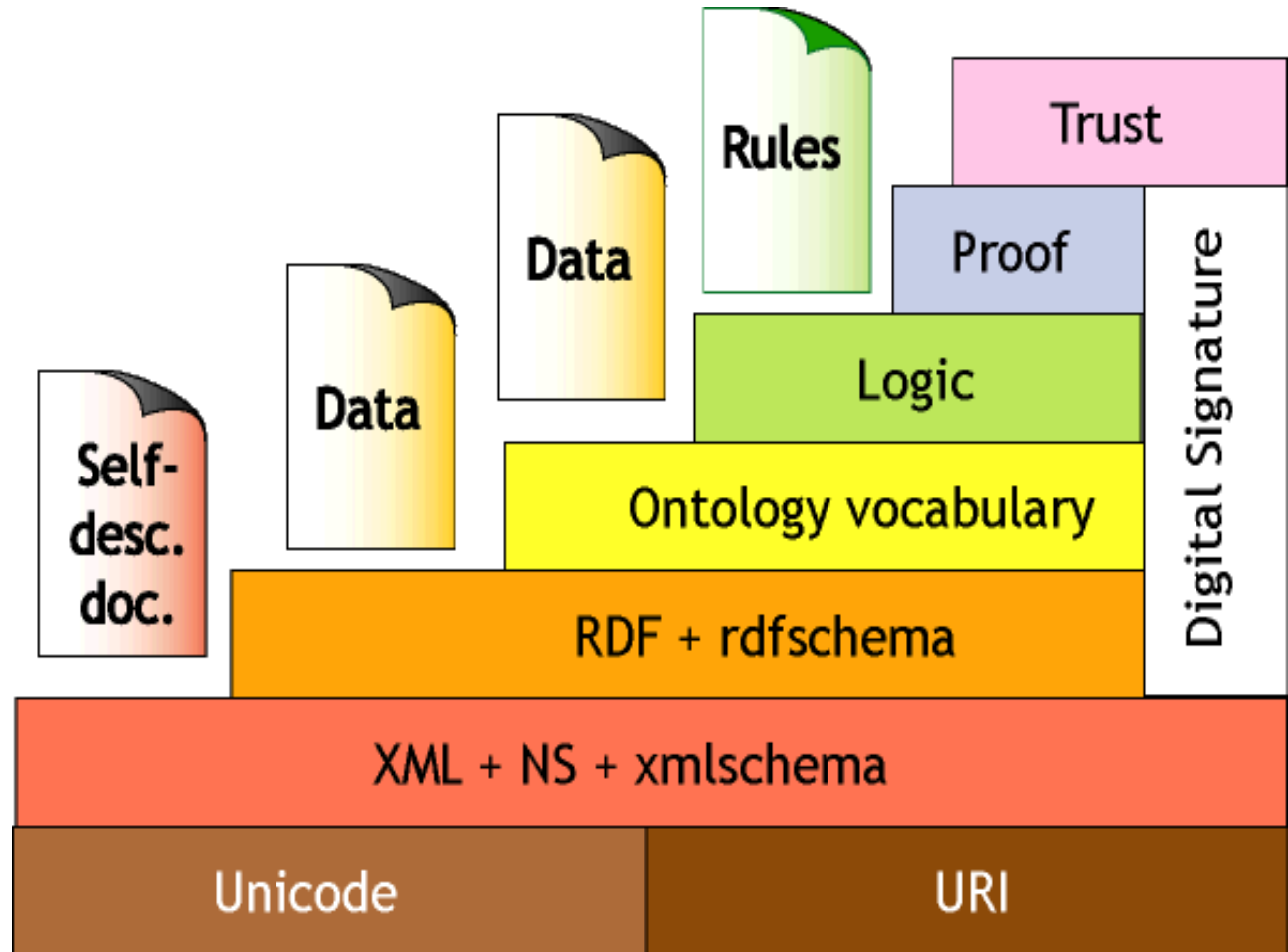


# Exemples d'application



- **Moteurs de recherche "intelligents"**
- **Comparaison de produits sur le web**
- **Portails rassemblant les commentaires sur des produits (livre, CD, vidéos, films) de divers sites (amazon.com, fnac.com, universal.com).**
- **Annotations (revue d'article, signets collaboratifs, ...)**

# Structure du web sémantique



[Berners-Lee, W3C 99]



- **XML : format d'échange de données, pas de sémantique**
- **RDF : méta-données, un peu de sémantique et d'inférence.**
- **RDF Schema : modélisation de concept, plus de sémantique et d'inférence.**
- **DAML+OIL : langage d'ontologie plus expressif, très expressif => inférence plus coûteuse.**
- **OWL : langage d'ontologie standard, avec différents niveaux d'expressivité (Lite, Descr. Logic, Full)**

# Publication Web Basée XML

## Le Framework Cocoon

# Exemple : Cocoon



- **Framework de publication Web basé XML**
  - (voir [présentation](#) de Sylvain Wallez,  
AnyWare Technologies)

# Applications et Architectures basées XML

# Les classes de documents Métiers



- **Utilisation de XML pour des classes d'information utiles à un domaine métier donné**
- **Exemples**
  - Business
  - Health
  - Engineering
  - ...

# Exemples



- **ebXML**
  - ISO 15000-1:  
ebXML Collaborative Partner Profile Agreement
  - ISO 15000-2:  
ebXML Messaging Service Specification
  - ISO 15000-3:  
ebXML Registry Information Model
  - ISO 15000-4:  
ebXML Registry Services Specification
  - ISO 15000-5:  
ebXML Core Components Technical Specification, Version 2.01.



# Exemples



- **xBML**
  - **e**xtended **B**usiness **M**odeling **L**anguage
  - Permet de définir des « business Process Models » et des « Organisation Models »

# Exemples



- **Health Level 7 (HL7) Healthcare XML Format**
  - XML specification for the exchange of clinical data and information.
  - The purpose of the exchange of clinical data includes, but is not limited to: provision of clinical care, support of clinical and administrative research, execution of automated transaction oriented decision logic (medical logic modules), support of outcomes research, support of clinical trials, and to support data reporting to government and other authorized third parties.

# Exemples



- **PDML :Product Data Markup Language**
- **FpML :Financial products Markup Language**
- **aecXML :Architecture-Engineering-Construction  
Extensible Markup Language**
- ...



- **Utilisation de XML comme format d'échange et d'interopérabilité entre applications**
- **Protocole**
  - SOAP
  - REST
  - ...
- **Format de données**
  - SVG
  - MathML
  - ...

# Exemples



- **SVG**
- <http://svgmap.free.fr/carte.htm>



- **MathML**

```
<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>
  </mrow>
</math>
```

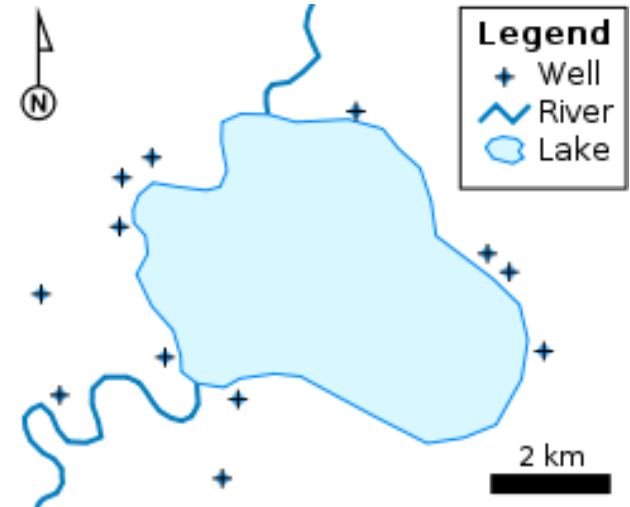
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

# Exemples



## • GML

```
<abc:Building gml:id="SearsTower">
  <gml:name>Sears Tower</gml:name>
  <abc:height>52</abc:height>
  <abc:position>
    <gml:Point>
      <gml:posList>100,200</gml:posList>
    </gml:Point>
  </abc:position>
  <app:extent>
    <gml:Polygon>
      <gml:exterior>
        <gml:LinearRing>
          <gml:posList>100,200</gml:posList>
        </gml:LinearRing>
      </gml:exterior>
    </gml:Polygon>
  </app:extent>
</abc:Building>
<abc:Building gml:id="SearsTower">
  <abc:position xlink:type="Simple" xlink:href="#p21"/>
</abc:Building>
<abc:SurveyMonument gml:id="g234">
  <abc:position>
    <gml:Point gml:id="p21">
      <gml:posList>100,200</gml:posList>
    </gml:Point>
  </abc:position>
</abc:SurveyMonument>
```





- Voir le cours : Rémi Badonnel - LORIA , INRIA