

# Intégration de données XML hétérogènes

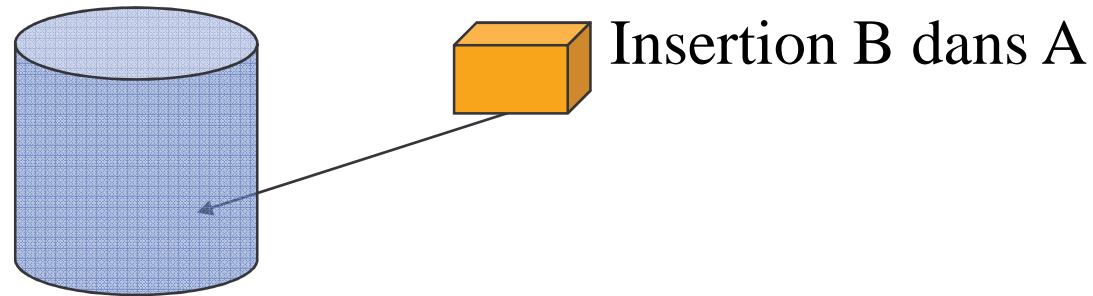
Les « marques schématiques »  
comme lien entre les niveaux  
syntaxiques et sémantiques



## Problématique Intégration de données

Données A

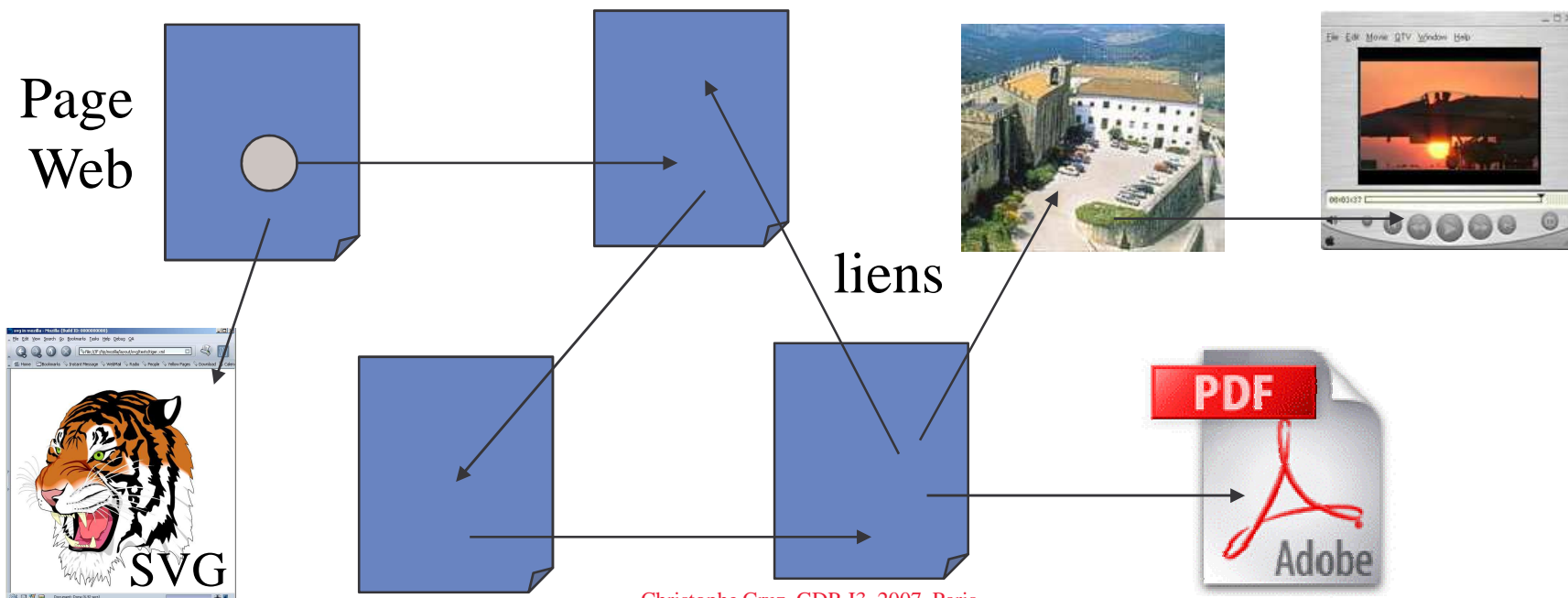
Données B



Quelle est la signification d'insérer ?

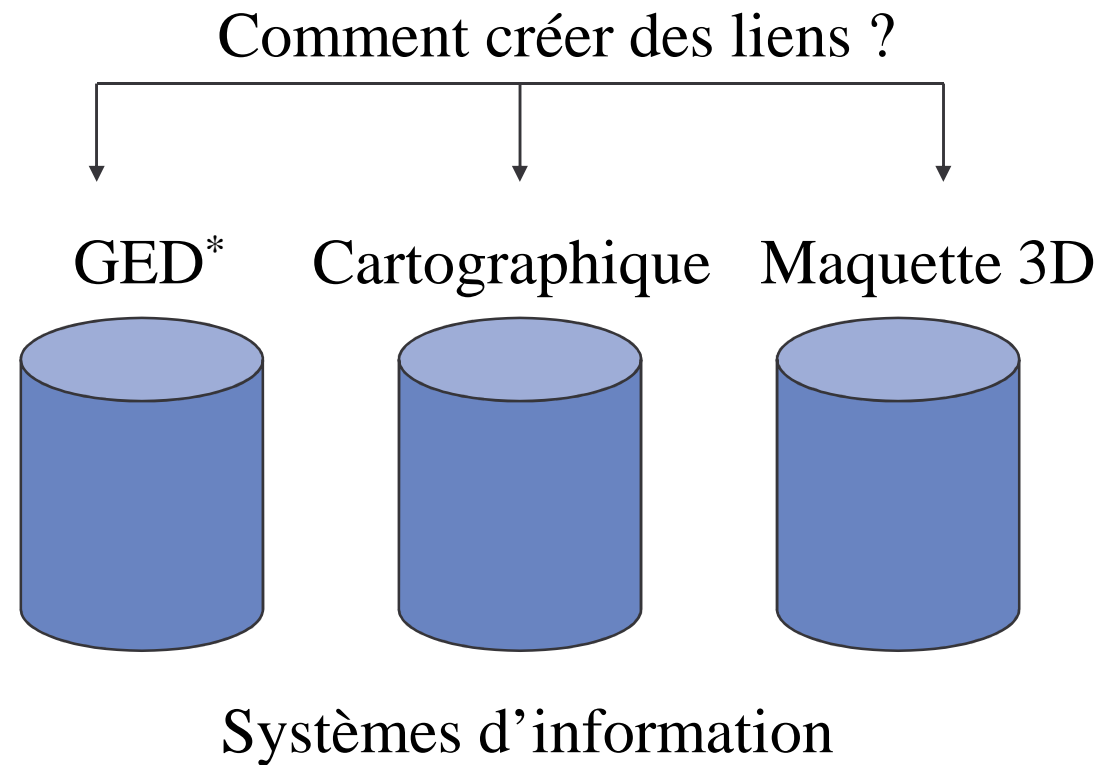
## Contexte du Web

### Hypertexte et Hypermédia



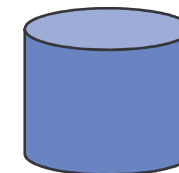
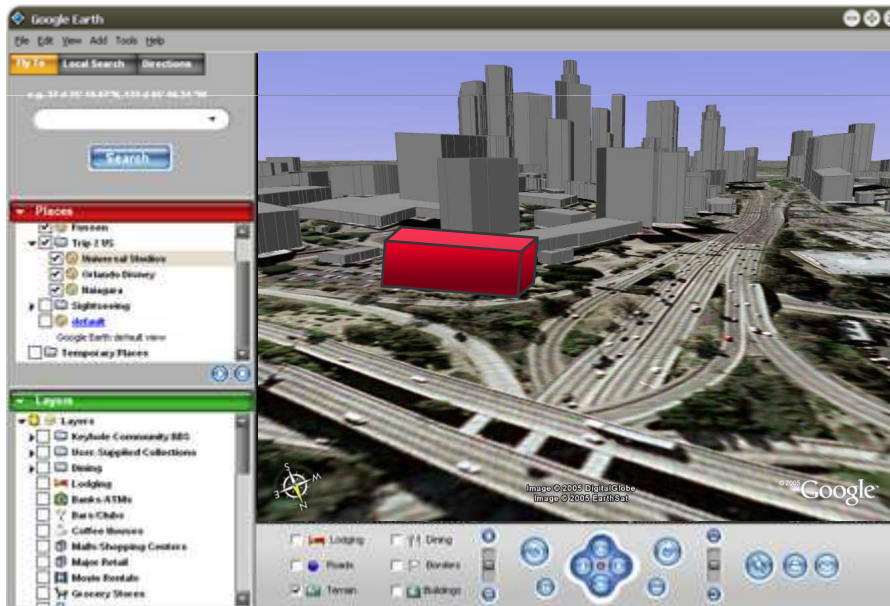
Christophe Cruz, GDR I3, 2007, Paris

## Contexte des bases de données

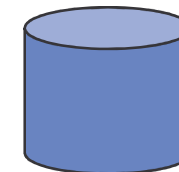


\* Gestion Électronique de Documents Christophe Cruz, GDR I3, 2007, Paris

## Contexte des bases de données

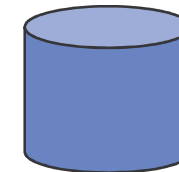
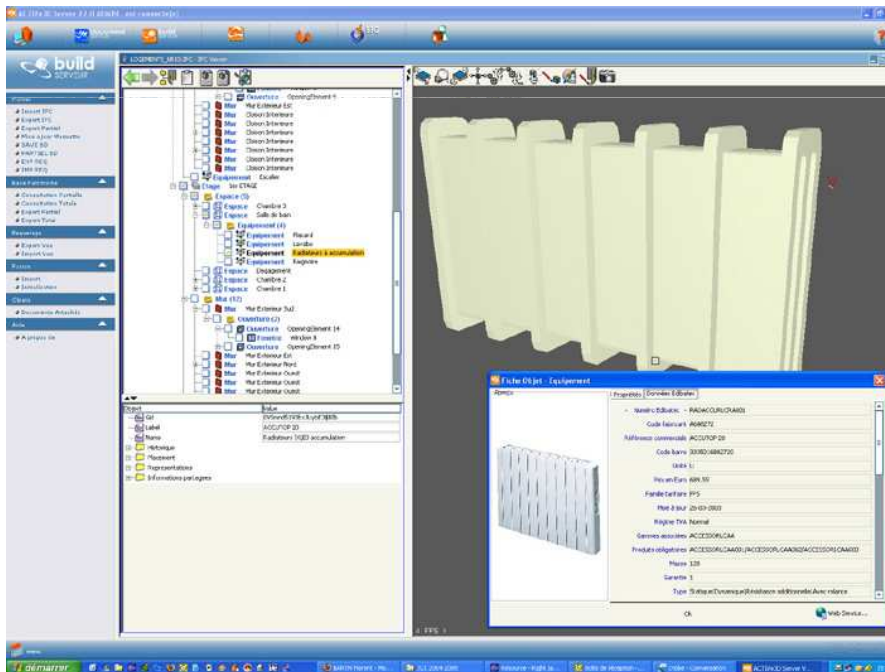


Google Earth,  
Système d'infor.  
Cartographique

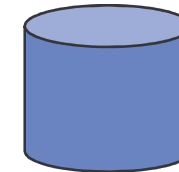


Maquette  
numérique 3D

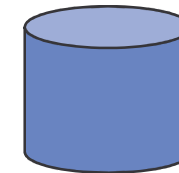
## Contexte des bases de données



Google Earth,  
Système d'infor.  
Cartographique



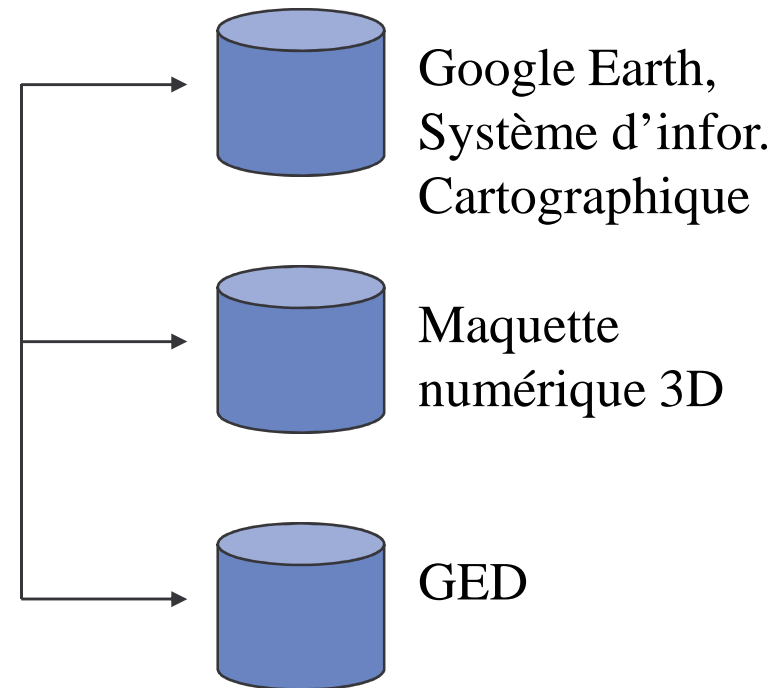
Maquette  
numérique 3D



GED

## Contexte des bases de données

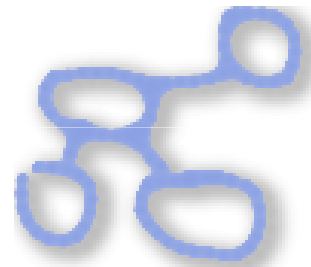
Intégration de données XML  
A l'aide de liens hypermédias  
mais  
COMMENT ?



Aujourd'hui

XML

Résout le problème d'hétérogénéité syntaxique et structurel



<balise> </balise>

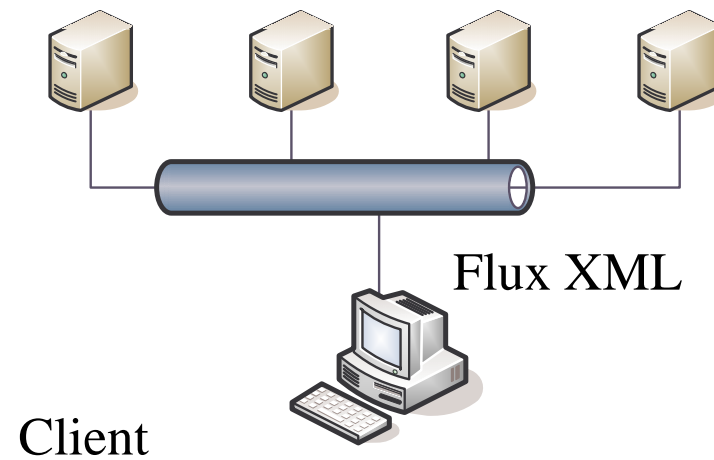
Technologies et outils génériques



## XML

### Intéropérabilité et SOA

#### Web Services



## Hétérogénéité sémantique

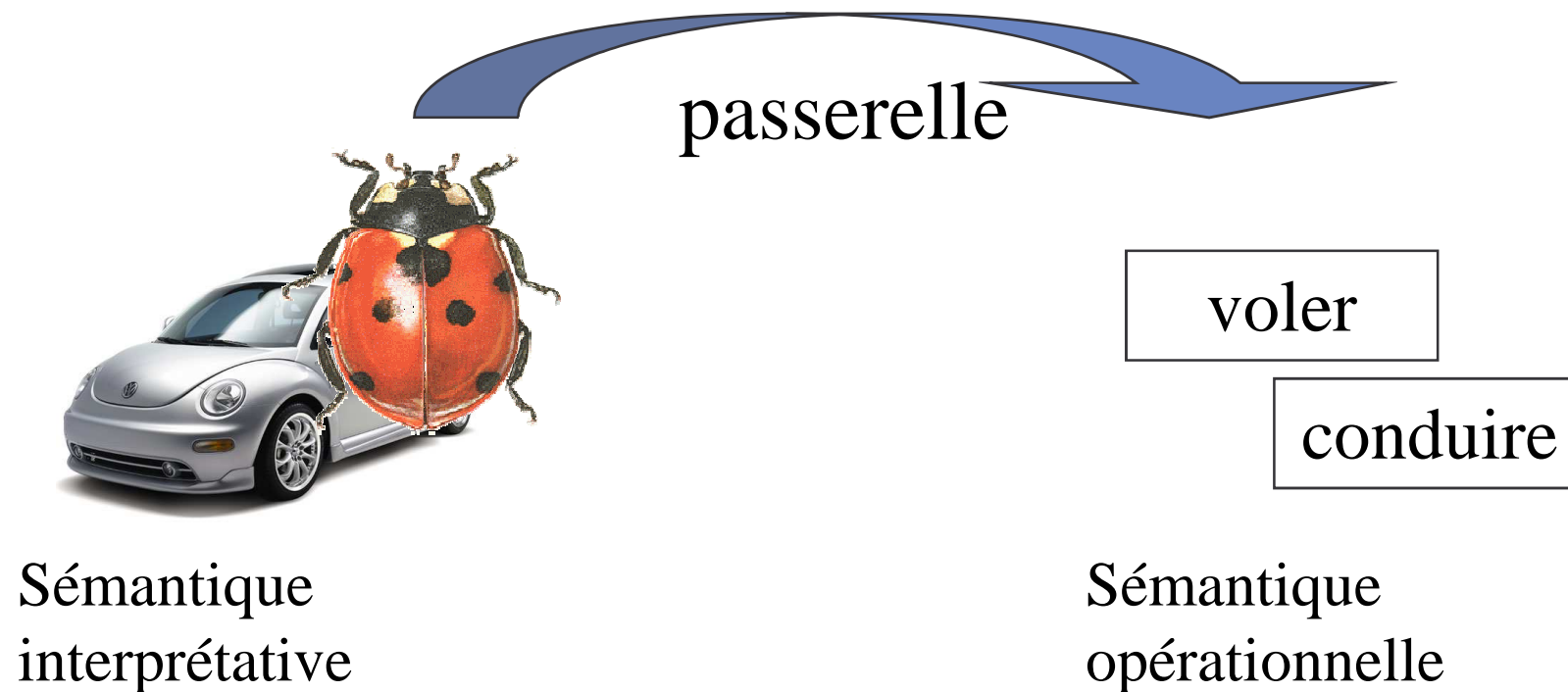
*La « Sémantique »* fait référence à la signification des données  
contrairement à la syntaxe



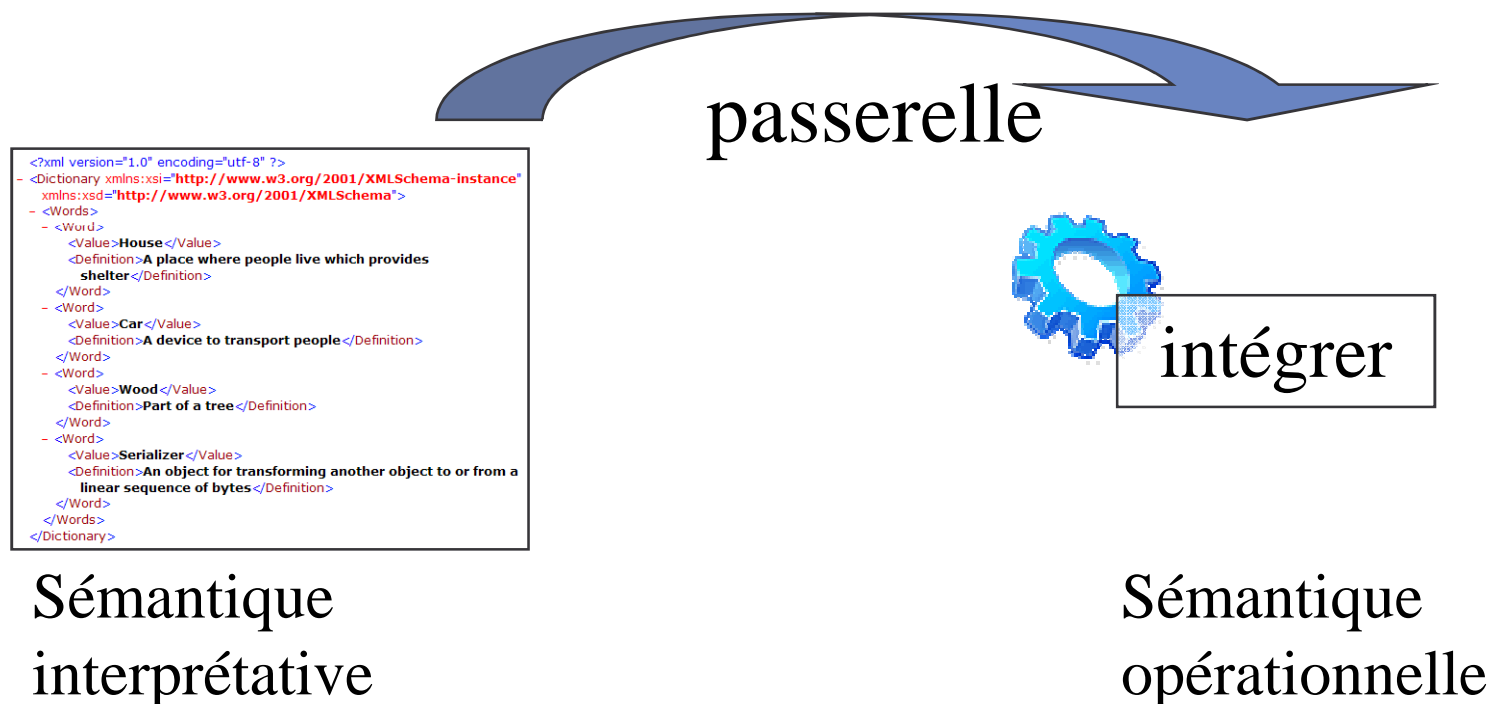
<Coccinelle>



## Ontologie (sémantique formelle)



## Ontologie (sémantique formelle)



## Schémas et documents

<balise>

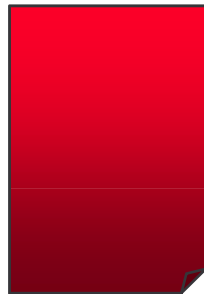
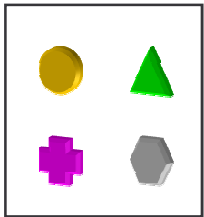


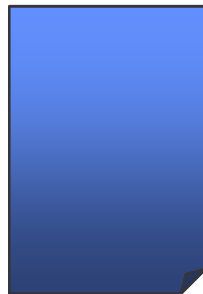
Schéma XML 1



Schéma XML 2



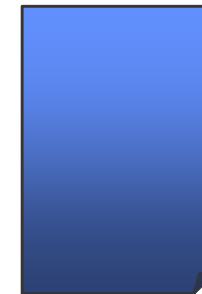
Document XML 1



Document XML 2



Document XML 3



Document XML 4

## Liens entre schémas

<balise>

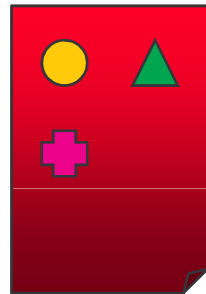
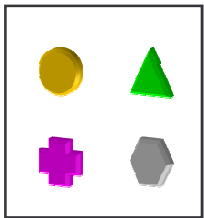


Schéma XML 1



Ontologie

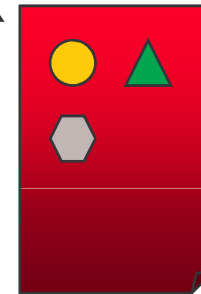
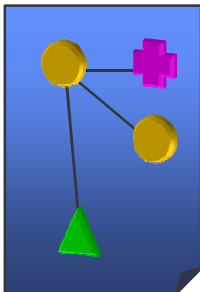
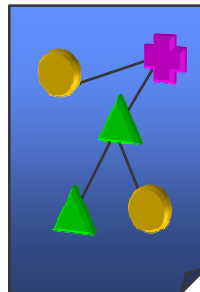


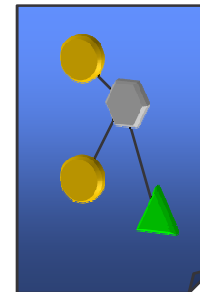
Schéma XML 2



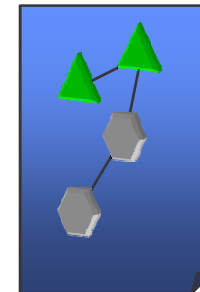
Document XML 1



Document XML 2

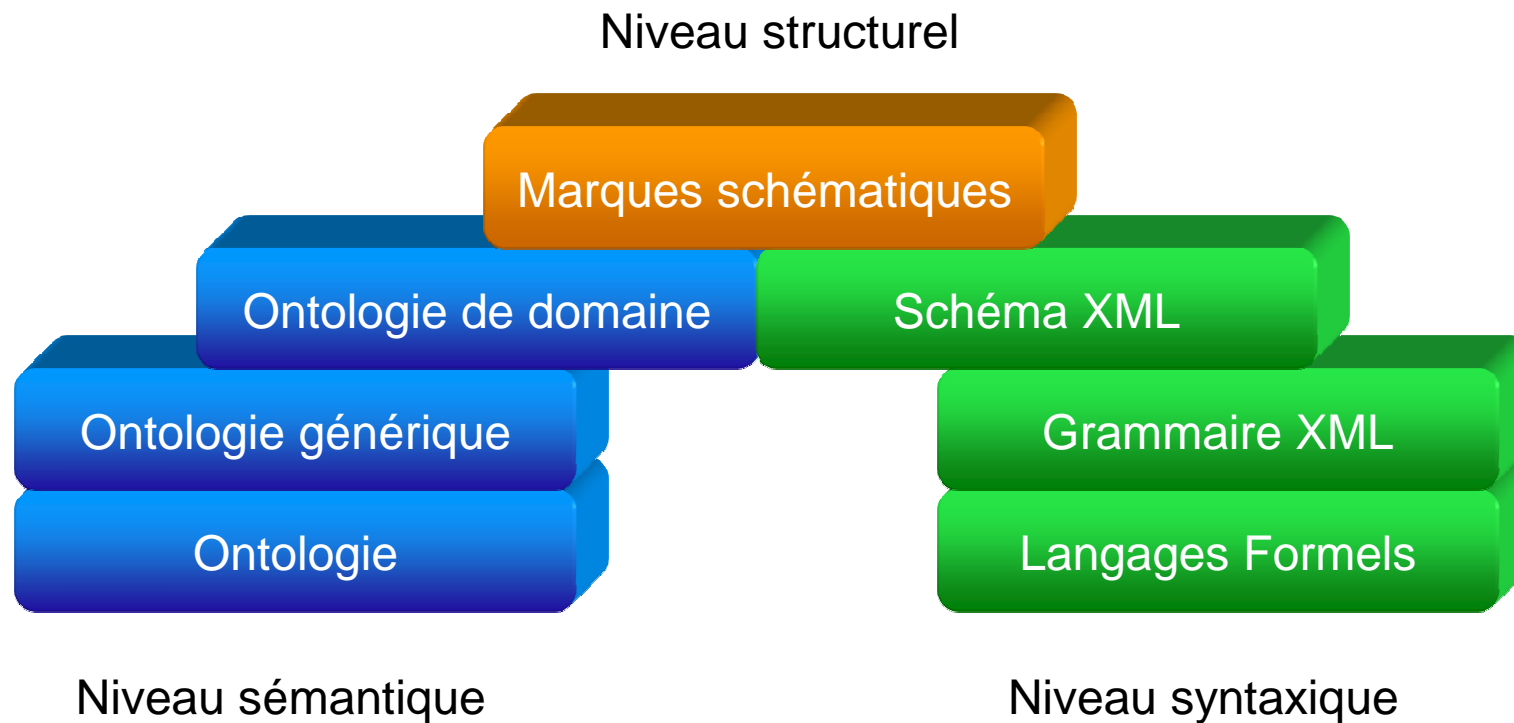


Document XML 3



Document XML 4

## Vue d'ensemble



## Règles de production

**Définition** : Un document  $x$  est bien formé si  $x$  est généré par des règles de production d'un langage de Dyck [LIE98] sur

$$T = A \cup \bar{A}$$

Un langage de Dyck est un langage engendré par une grammaire hors contexte de la forme suivante où

$$a_n \in A \quad \text{et son équivalent} \quad b_n \in \bar{A}$$

$$S \rightarrow SS \mid \varepsilon \mid a_1 S b_1 \mid a_2 S b_2 \mid \dots \mid a_n S b_n \text{ avec } n \geq 1$$



## Langage de Dyck (J. Berstel)

$$P = \{(S \rightarrow a(S \mid T)(S \mid T)\bar{a}), (T \rightarrow b), (T \rightarrow Tb)\}$$

```
<?xml version="1.0"«  
encoding="UTF-8"?>  
<!ELEMENT ROOT (A | (B, C))>
```

```
<?xml version="1.0" encoding="UTF-8"?>  
<xs:schema      xmlns:xs="http://www.w3.org/2001/XMLSchema"  
elementFormDefault="qualified">  
  <xs:element name="A"/>  
  <xs:element name="B"/>  
  <xs:element name="C"/>  
  <xs:element name="ROOT">  
    <xs:complexType>  
      <xs:choice>  
        <xs:element ref="A"/>  
        <xs:sequence>  
          <xs:element ref="B"/>  
          <xs:element ref="C"/>  
        </xs:sequence>  
      </xs:choice>  
    </xs:complexType>  
  </xs:element>  
</xs:schema>
```

## Langage de Dyck (J. Berstel)

**Proposition** : Pour chaque langage XML  $L$ , il existe une seule grammaire XML **réduite** générant  $L$ .

Soit le langage

$$L = \left\{ ca(b\bar{b})^{n_1} \bar{a}a(b\bar{b})^{n_2} \bar{a}K a(b\bar{b})^{n_k} \bar{a}\bar{c} \mid n_{1,2} K_k > 0 \right\}$$

Alors

$$F_c(L) = L, F_b(L) = \{b\bar{b}\}^*, F_a(L) = \{a(b\bar{b})^* \bar{a}\}$$

**Définition** : Une marque schématique est une marque sur un schéma XML permettant d'identifier une règle de production.

## Ontologies

- Représentation de la connaissance
- Approuvé par une communauté de personne
- Accessible par les humains et applications



Ontologie

tacite / explicit  
« un fenêtre est dans un mur »

## Ontologie Générique



 Concept Générique

 Relation Générique

 Attribut Générique

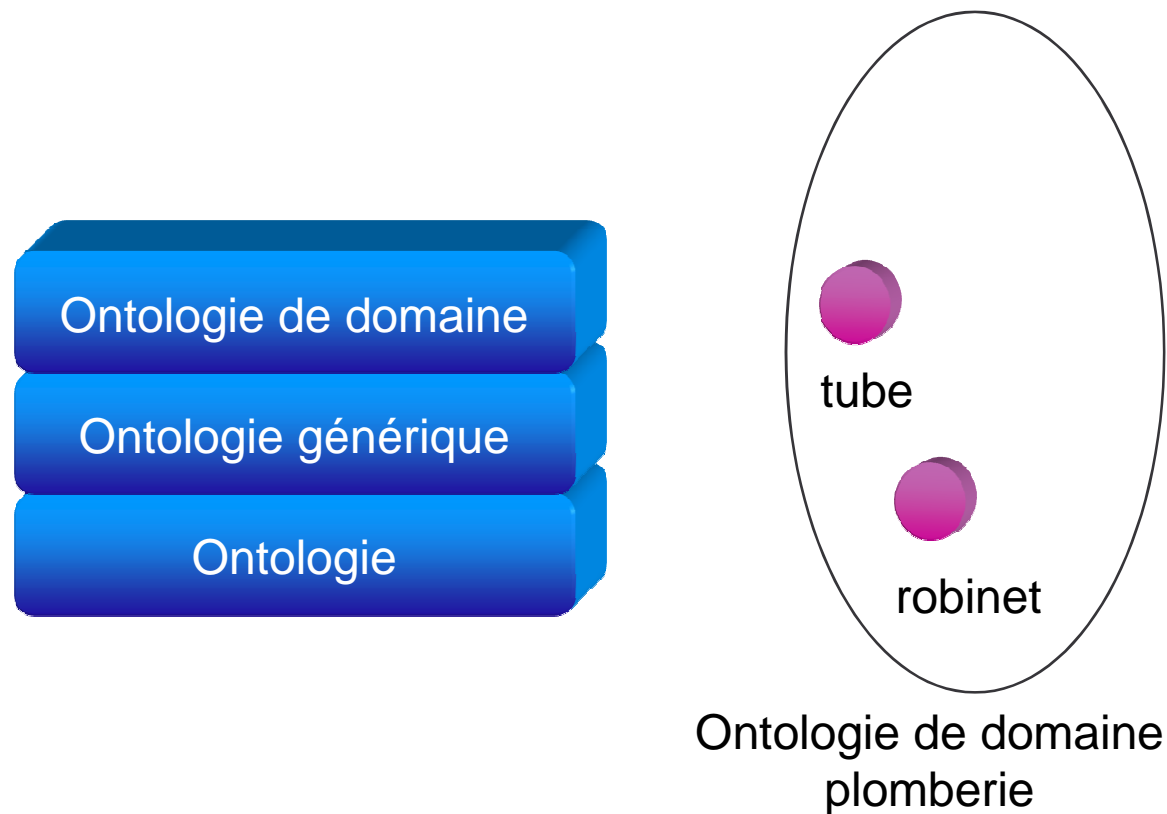
## Ontologie Générique

Ontologie de domaine

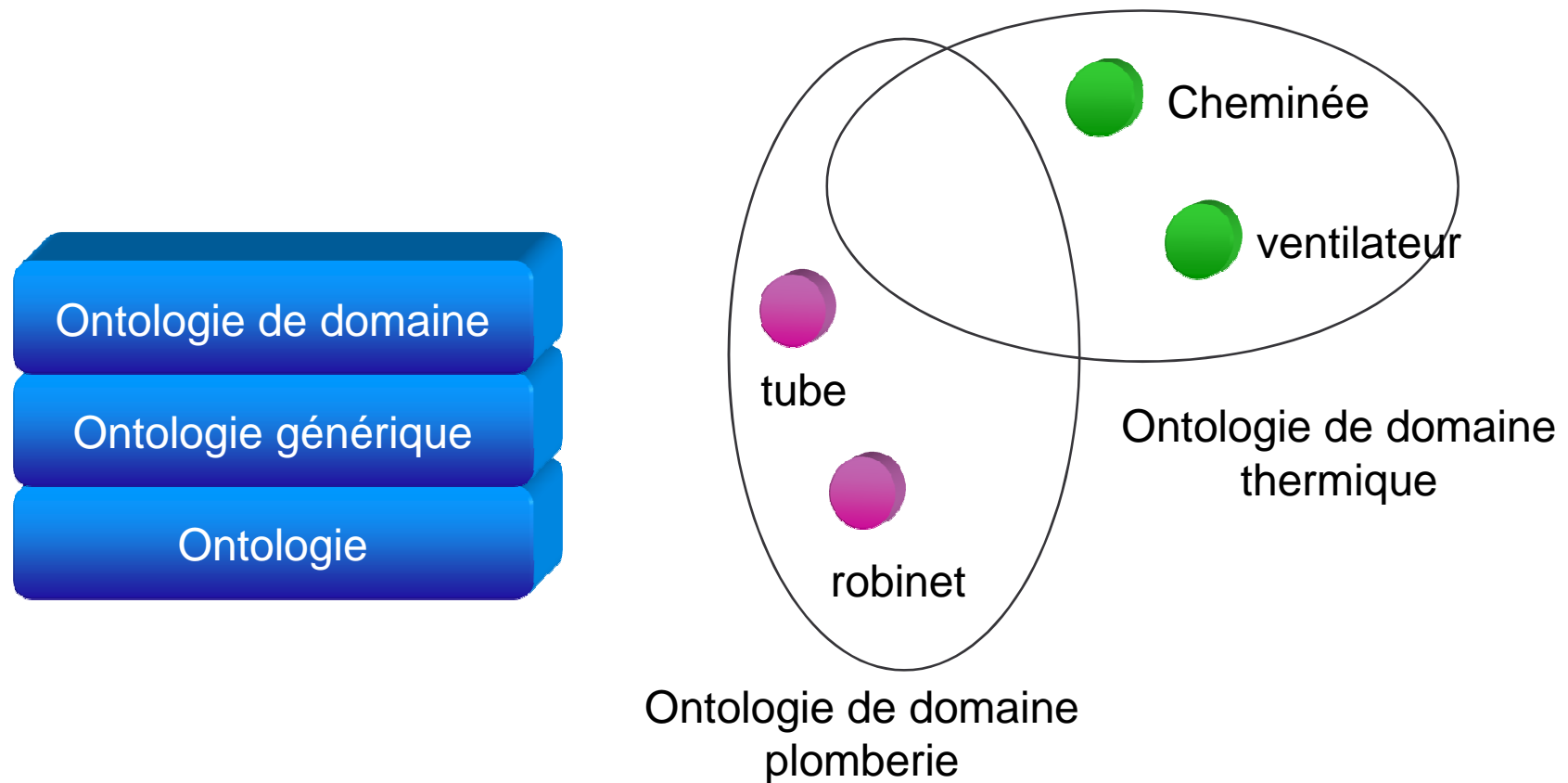
Ontologie générique

Ontologie

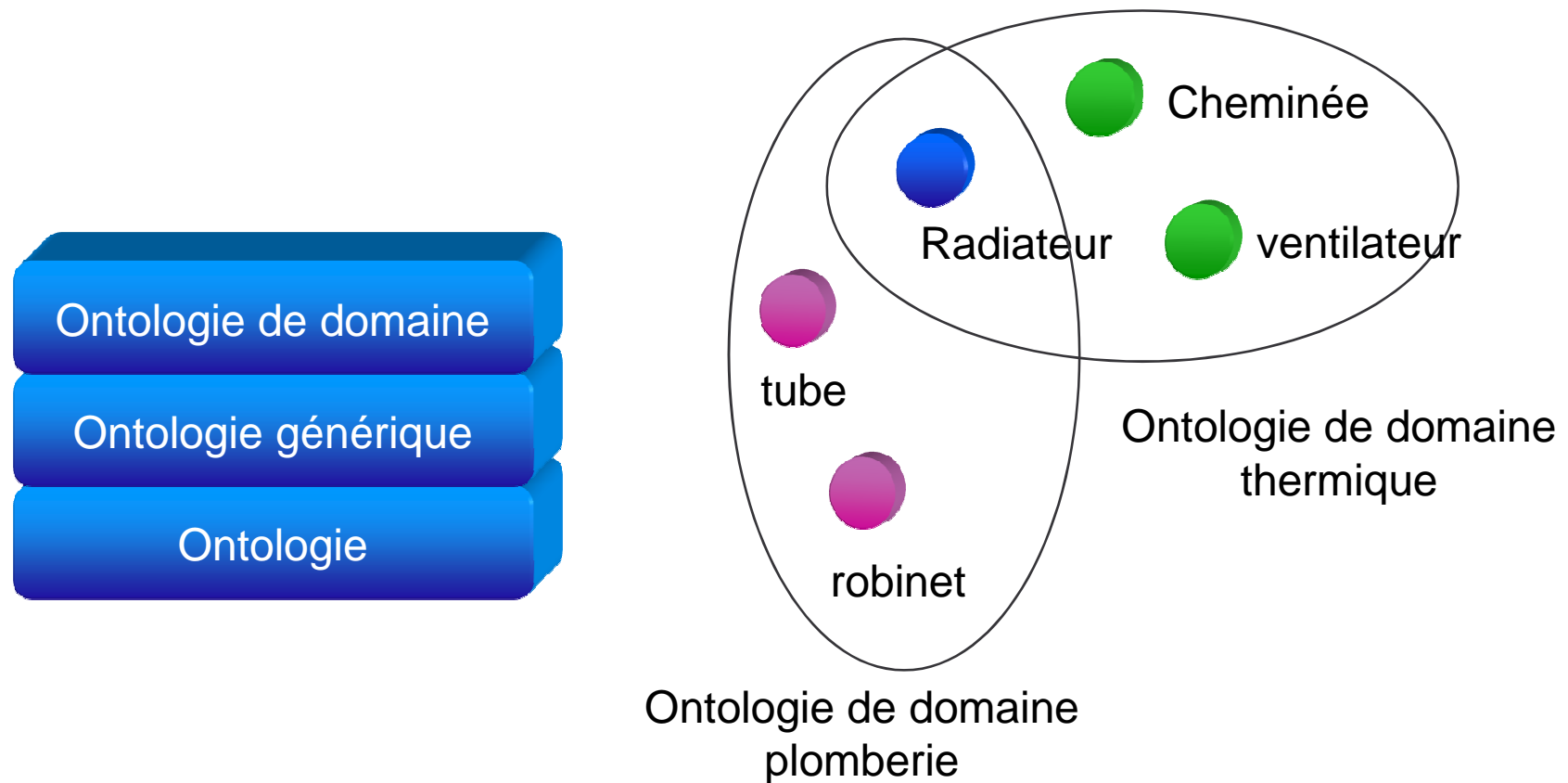
## Ontologie Générique



## Ontologie Générique



## Ontologie Générique





## Définition des “marques schématiques”

### schema XML

```
<xs:element name="Floor">
  <xs:complexType>
    <xs:sequence maxOccurs="unbounded">
      <xs:element ref="Slab" minOccurs="0"/>
      <xs:element ref="Wall" minOccurs="0"/>
    </xs:sequence>
    <xs:attribute name="Label" type="xs:string" />
  </xs:complexType>
</xs:element>
```

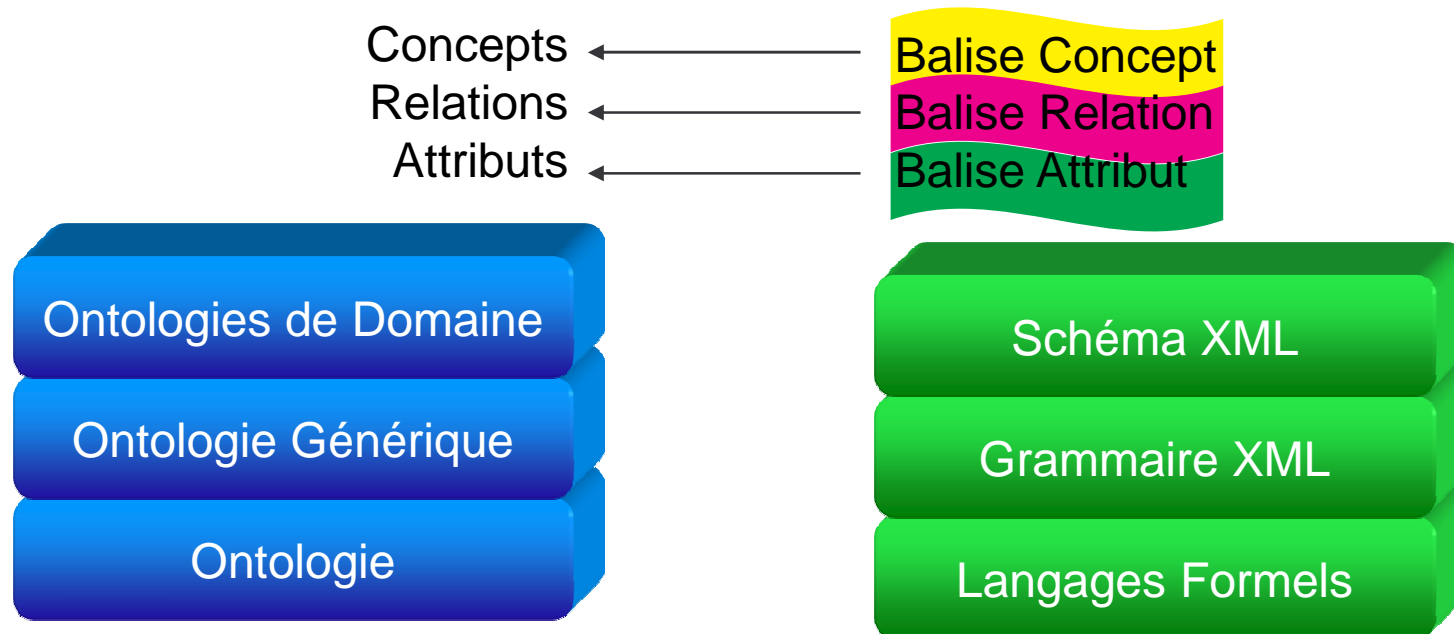
Balise Concept  
Balise Relation  
Balise Attribut

Schéma XML

Grammaire XML

Langages Formels

## Definition of Schematic marks



Définition de six règles pour  
l'intégration

## Trois étapes pour l'intégration:

- Définition de structures pour l'intégration des données
- Définition de l'intégration des schémas
- Définition de l'intégration des données

## Connaissances

Schémas XML

S.X. 1 S.X. 2 ... S.X. n

Balise Concept  
Balise Relation  
Balise attribut

O.D. 1 O.D. 2 ... O.D. n

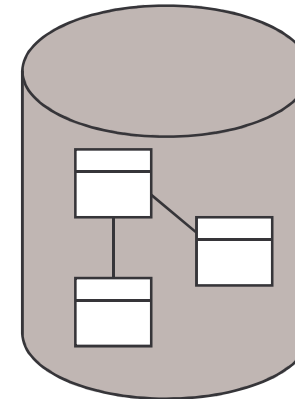
Ontologie générique

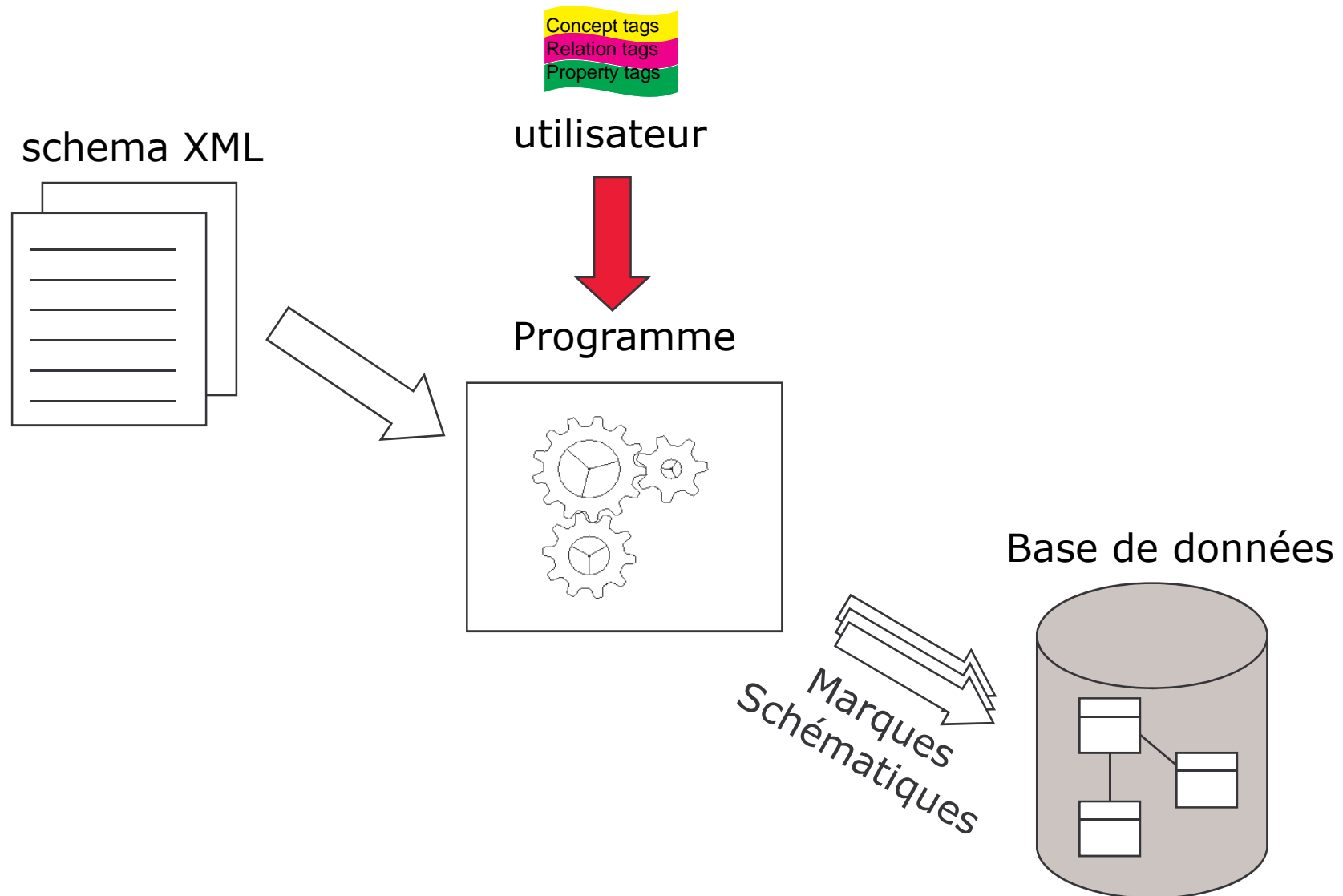
## Données

Documents XML

D.X. 1 D.X. 2 ... D.X. m

## Base de données





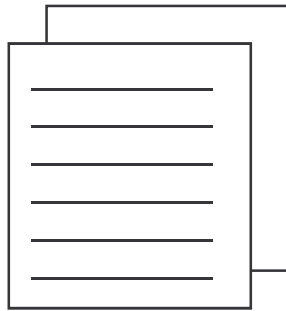
## Définition des règles

The screenshot displays the 'INTEGRATION DE SCHEMAS XML' application. On the left, a tree view shows the schema structure: 'schema' contains 'purchaseOrder', which contains 'shipTo' (with elements 'name', 'street', 'city', 'state', 'zip') and 'billTo' (with elements 'name', 'street').

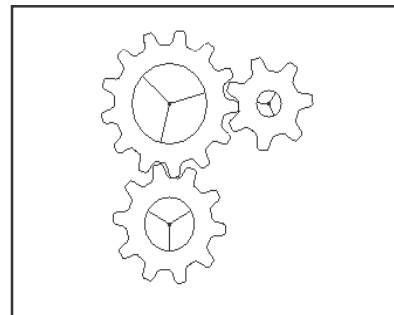
On the right, a table defines the mapping rules for the integration. The table has columns: Element, Rien, Concept, Relation, Attribut, and Nom.

Element	Rien	Concept	Relation	Attribut	Nom
purchaseOrder	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	chaseOrder ▼
shipTo	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	shipTo
name	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	name
street	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	street
city	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
state	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
zip	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
billTo	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	billTo
comment	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
items	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
item	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	objet
productItems	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

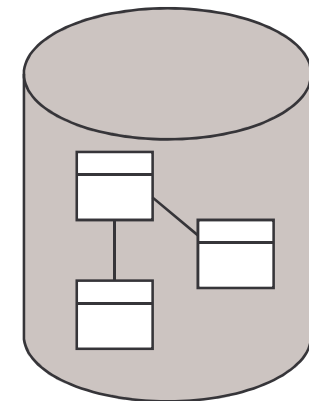
Documents XML



Programme



Base de données



Données

Sémantiques

## Identifiant global

INTÉGRATION DE SCHEMAS XML

Menu

- schema
  - purchaseOrder
    - shipTo
      - name
      - street
      - city
      - state
      - zip
    - billTo
      - name
      - street
      - city
      - state
      - zip
    - comment
    - items
      - item
        - productName
        - quantity
        - USPrice
        - comment
        - shipDate

ATTRIBUTS

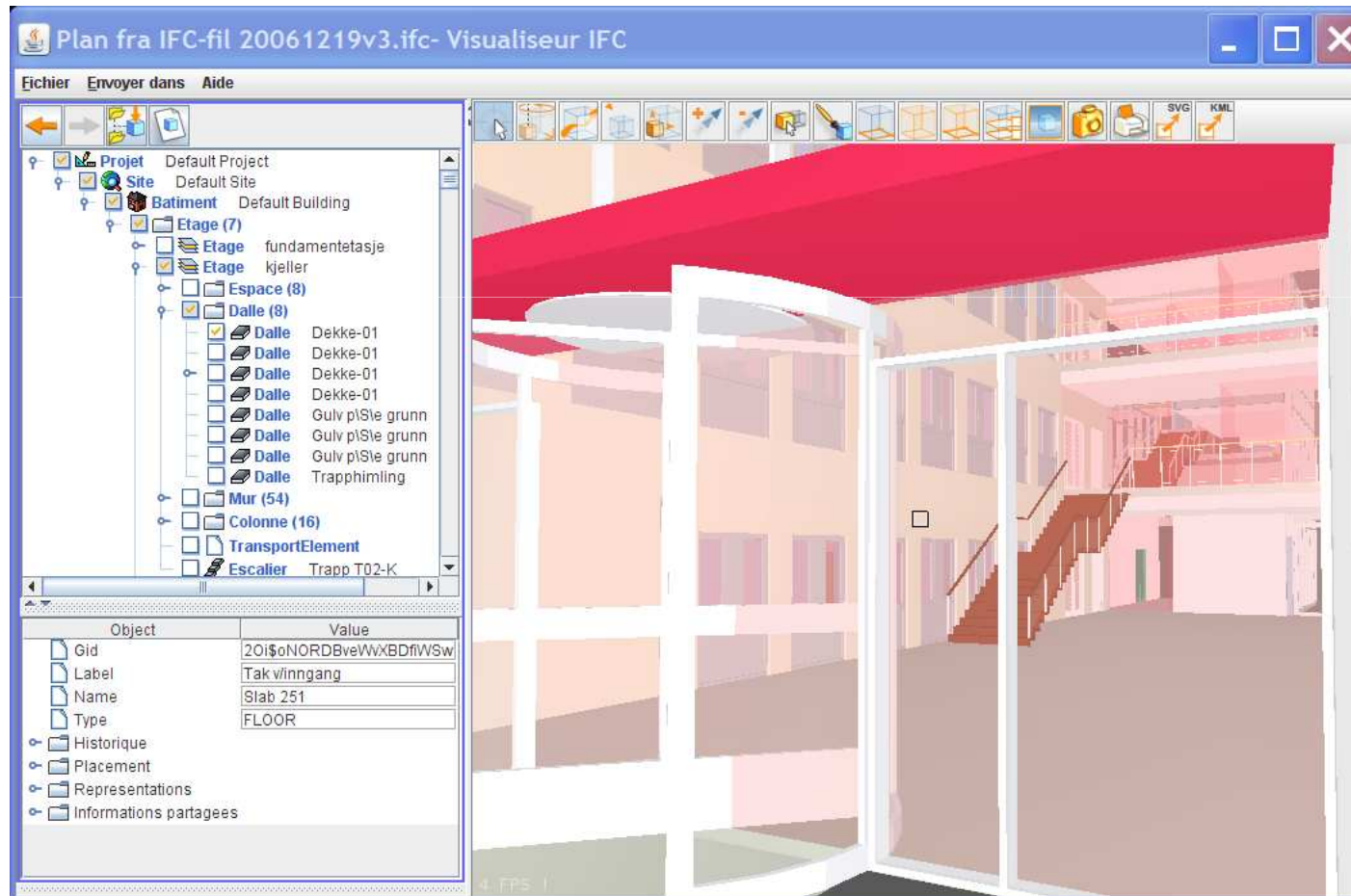
Attribut	Requis	id
partNum	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Element	Rien	Concept	Relation	Attribut	Nom
purchaseOrder	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	purchaseOrder
shipTo	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	shipTo
name	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	name
street	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	street
city	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
state	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
zip	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
billTo	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	billTo
comment	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
items	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
item	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	objet
productName	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
quantity	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
USPrice	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
shipDate	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	



## Application

### Informations architecturales et géométriques



## Travaux futurs...

- Prototypage du processus d'extraction en fonction du contexte
- Web Services et schémas XML

