

# INU3011 Documents structurés

## Cours 13

### Publication Web de documents XML

# Plan

- Retour sur Mini-quiz 3
- Publication Web de documents XML
- Bases de données XML
- Travail de Publication Web (TP 6)
  - Protocole
  - Démonstration en classe
- Travail supervisé

# Prochain (et dernier) cours

- Lundi 24 avril
- Brève conclusion du trimestre
- Travail supervisé en laboratoire

# Retour sur le Mini-quiz 3

- Moyenne = 71,4 %
- Question XSLT (39%)
- Identifier les chemins XPath (50%)

`//pivoine`

`/vin/type[2][@*]`

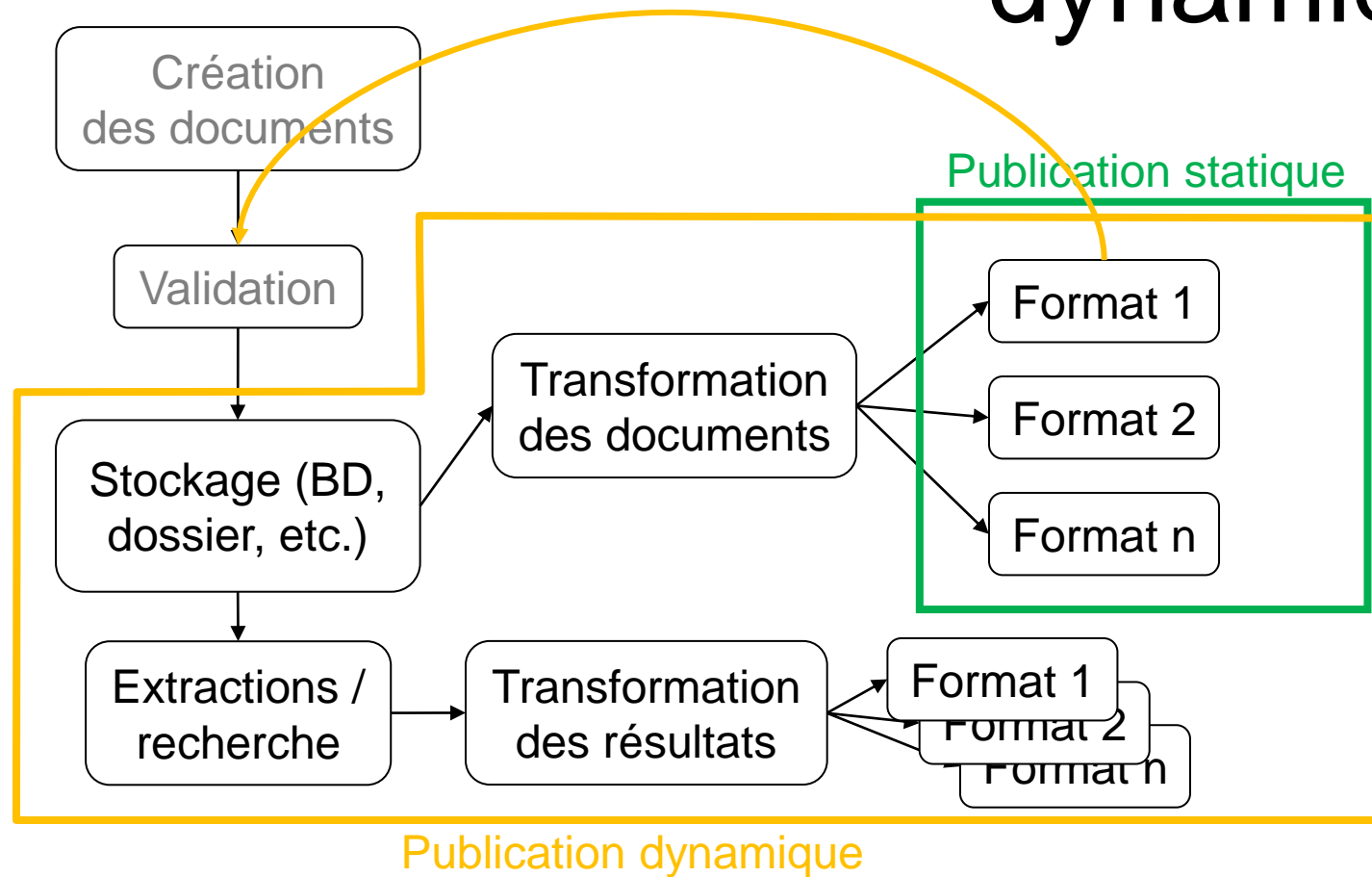
`pin`

`position()`

`count(../pers[@e-mail])`

`@type = ../@genre`

# Publication statique vs dynamique



# Langages typiquement utilisés

- Transformations : XSLT
- Extractions / recherches : XSLT / XQuery

# XQuery (1/2)

- XML Query (XQuery)
- Statut normatif
  - [XQuery 3.1: An XML Query Language](#)
    - W3C Recommendation 2017-03-21
  - [XQuery and XPath Full Text 3.0](#)
    - W3C Recommendation 2015-11-24

# XQuery (2/2)

- Langage normalisé d'interrogation pour bases de données XML
- XPath est au cœur de XQuery :
  - XQuery 1.0 : XPath 2.0 (dont XPath 1.0 est un sous-ensemble)
  - XQuery 3.1 : XPath 3.1

N.B. : Il n'y a jamais eu de XQuery 2.0



# Publication statique

- "Exportation" nécessaire vers le Web :
  - XML directement (XSLT en navigateur)
  - HTML (XSLT préalable, en lot)
- L'exportation doit être refaite après chaque modification des documents
- Accès :
  - par navigation via "index" ou "page d'accueil"
  - par requête (via post-indexation)

# Post-indexation du statique

- Indexation du *texte seulement* des documents, via outil comme [FreeFind](#) ou [Google Programmable Search Engine](#)
- Recherche par requêtes sur le *texte* des documents, mais pas sur la structure XML
- Fondamentalement statique
  - Réexportation + réindexation requises si changements

# Publication dynamique

- Accès à la BD via le Web
  - Interrogation / recherche
    - Accès par navigation *et* par requêtes
  - Administration de la BD (accès, backups, etc.)
  - Saisie de documents
- Pas d'exportation à refaire à chaque modification
  - La BD interrogée est toujours à jour

# Bases de données XML

- Exemples de SGBD XML :
  - BaseX <<https://baseX.org/>>
  - eXist <<https://exist.sourceforge.net/>>
  - [MarkLogic Server](#)

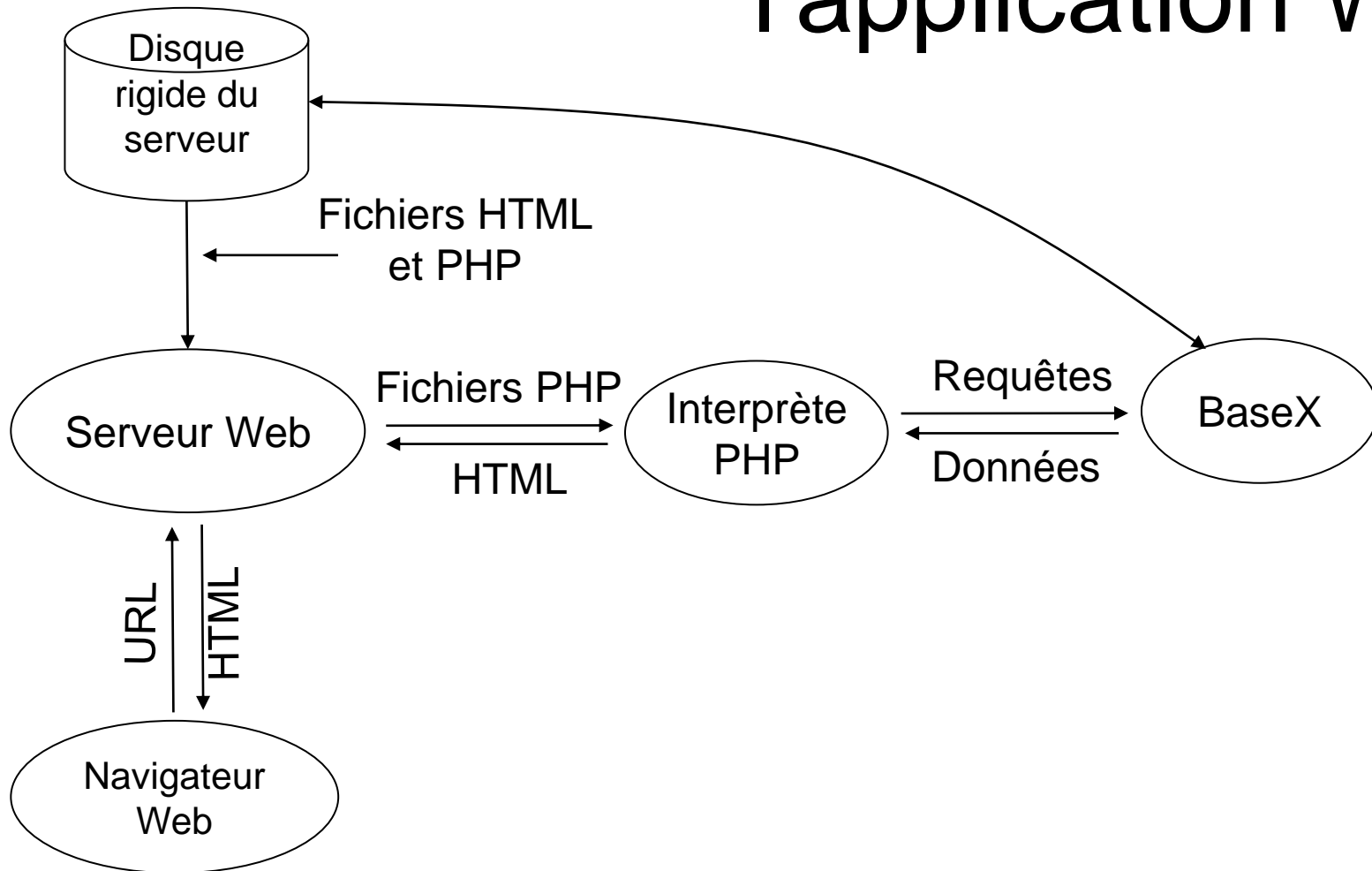
# BD XML : structure générale

- Très analogue à un système de fichiers
- Les différentes "fiches" (unités principales d'information) peuvent être dans un seul document ou dans des documents séparés
- "Collections" et "sous-collections"
  - Analogue à des dossiers et sous-dossiers
  - Permettent de stocker dans la BD des fichiers auxiliaires, comme images, etc.

# BaseX

- Implantation en Java
- Facile d'installation
- Prise en charge de XQuery
- Peut être "connecté" à oXygen, qui devient alors un *module de saisie / modification* pour les documents des BD
- Extensions pour la recherche en texte intégral

# Architecture de l'application Web



# Travail de publication Web

- [Protocole](#)
- Démonstration