

La sémantique dans VOTheory



Laboratoire d'Étude du Rayonnement et de la Matière en Astrophysique



Laboratoire Univers et Théories



Introduction

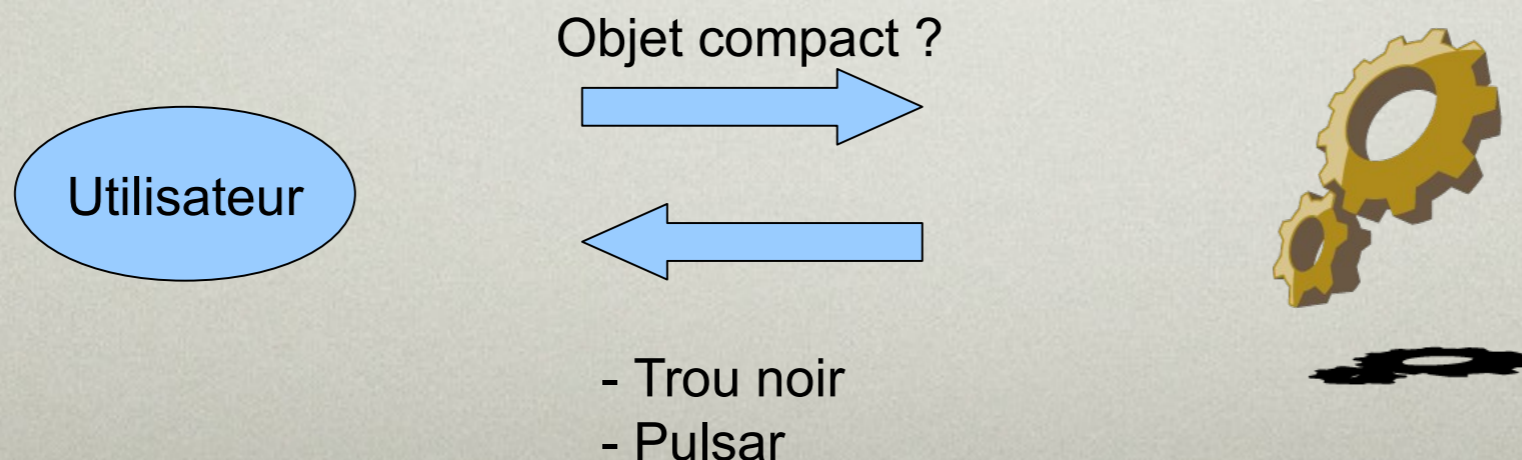
- SimDB permet de stocker des informations sur des simulations numériques et leurs résultats :
 - Code utilisé
 - Paramètres
 - Algorithmes
 - Processus physiques modélisés
 - ...

- Quels outils utiliser pour rechercher et rendre interopérables ces informations ?

- Objectifs :
 - Permettre à des logiciels/codes de comprendre les quantités manipulées (dans le cas d'un workflow par exemple)
 - Proposer une recherche efficace à un utilisateur humain en allant au delà de la simple recherche textuelle

Introduction

- SimDB permet de stocker des informations sur des simulations numériques et leurs résultats :
 - Code utilisé
 - Paramètres
 - Algorithmes
 - Processus physiques modélisés
 - ...
- Quels outils utiliser pour rechercher et rendre interopérables ces informations ?
- Objectifs :
 - Permettre à des logiciels/codes de comprendre les quantités manipulées (dans le cas d'un workflow par exemple)
 - Proposer une recherche efficace à un utilisateur humain en allant au delà de la simple recherche textuelle



Introduction

- Web sémantique : rendre des données publiées sur le web compréhensibles par des logiciels
- Les informations sont formalisées pour pouvoir être traitées automatiquement
- Standards W3C issus de ces travaux :
 - RDF (Resource Description Framework) : modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées.

```
<RDF>
  <Description about="http://fr.wikipedia.org/wiki/Galaxie">
    <modification>30/11/2011</modification>
    <homepage>http://fr.wikipedia.org</homepage>
  </Description>
</RDF>
```

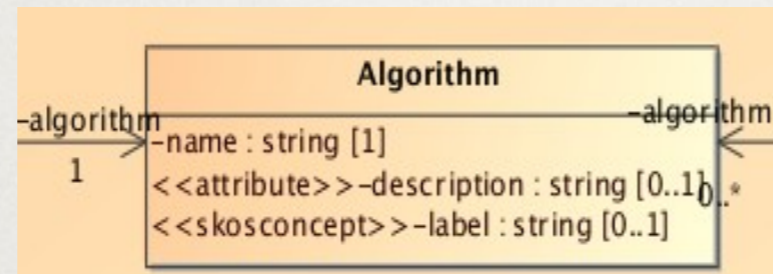
- RDF Schema : langage qui fournit des éléments de base pour la définition d'ontologies ou vocabulaires destinés à structurer des ressources RDF (notion de classes, d'héritage ...)
- OWL (Web Ontology Language) : langage permettant de créer des ontologies, vocabulaires plus complexes servant de support aux traitements logiques (inférences, classification automatique...)
- SPARQL (Sparql Protocol And Rdf Query Language) : langage de requêtes pour obtenir des informations à partir de graphes RDF.

Sémantique et simulation numérique

- Dans un premier temps, permettre à l'utilisateur de rechercher des simulations
- Choix d'utiliser un vocabulaire RDF : SKOS
- Permet d'écrire des systèmes d'organisation des connaissances comme des thesauri (liste de termes décrivant un domaine, reliés entre eux par des relations : synonymie, association, hiérarchie)
- Avantages :
 - Structure simple
 - Facile à générer depuis un vocabulaire déjà existant
 - L'utilisation d'ontologie aurait conduit à développer un modèle conceptuel supplémentaire
- Implémentation à travers le site votheory.obspm.fr
- Exemple de recherche dans Starformat

Utilisation de concepts sémantiques dans SimDB

- SimDM intègre un champ permettant d'affecter un concept skos aux instances de certaines classes



- Réalisation de 6 vocabulaires :

- Algorithms
- AstronomicalObjects
- DataObjectTypes (structure de données manipulée par le code : datacube, adaptative mesh cell)
- InputParameters
- JournalKeywords
- PhysicalProcesses
- PhysicalQuantities

- Exemples d'utilisation :

- Trouver un code en fonction des algorithmes utilisés
- Permettre une interopérabilité avec d'autres codes utilisant ces données :
 - Physical Quantity : Abundance, Column Density, Temperature ...

Utilisation de concepts SKOS dans SimDB

- Chaque vocabulaire est une liste de concepts éventuellement liés entre eux
- Chacun possède :
 - Un identifiant (une URI qui permet de l'identifier de manière automatisée)
 - Un prefLabel qui est son écriture principale
 - Zero ou plus altLabel (simple synonymes ou alias d'usage courant)
 - Une description
- Chacun de ces champs peut être écrit dans une ou plusieurs langues
- Les concepts sont reliés par un ensemble restreint de relations :
 - "broader"
 - "narrower"
 - "related"

Utilisation de concepts SKOS dans SimDB

- Chaque vocabulaire est une liste de concepts éventuellement liés entre eux
- Chacun possède :
 - Un identifiant (une URI qui permet de l'identifier de manière automatisée)
 - Un prefLabel qui est son écriture principale
 - Zero ou plus altLabel (simple synonymes ou alias d'usage courant)
 - Une description
- Chacun de ces champs peut être écrit dans une ou plusieurs langues
- Les concepts sont reliés par un ensemble restreint de relations :
 - "broader"
 - "narrower"
 - "related"

```
<skos:Concept rdf:about="http://purl.org/astronomy/vocab/IvoaAlgorithms/Stationary_Iterative_Method">  
  <skos:inScheme rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/scheme" />  
  <skos:prefLabel xml:lang="en">Stationary Iterative Method</skos:prefLabel>  
  <skos:narrower rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/Successive_Overrelaxation" />  
  <skos:narrower rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/Gauss-Seidel" />  
  <skos:narrower rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/Jacobi_Method" />  
  <skos:broader rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/Algorithm" />  
  <skos:related rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/krylov_Subspace_Method" />  
</skos:Concept>
```


Identification des concepts : PURL

- Chaque concept a un identifiant unique qui est utilisé pour marquer une entrée dans la base

```
<skos:Concept rdf:about="http://purl.org/astronomy/vocab/IvoaAlgorithms/Stationary_Iterative_Method">  
  <skos:inScheme rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/scheme" />  
  <skos:prefLabel xml:lang="en">Stationary Iterative Method</skos:prefLabel>  
  <skos:narrower rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/Successive_Overrelaxation" />  
  <skos:narrower rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/Gauss-Seidel" />  
  <skos:narrower rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/Jacobi_Method" />  
  <skos:broader rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/Algorithm" />  
  <skos:related rdf:resource="http://purl.org/astronomy/vocab/IvoaAlgorithms/krylov_Subspace_Method" />  
</skos:Concept>
```

<http://purl.org/astronomy/Algorithms/Concept>

Racine commune

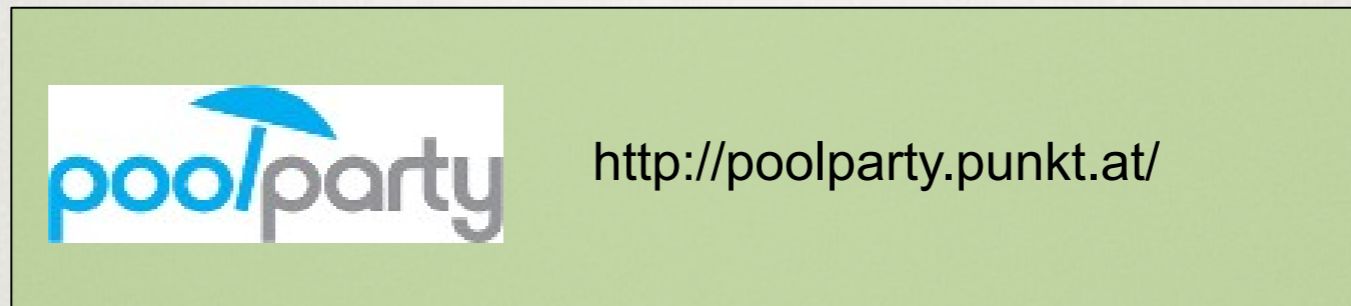
Vocabulaire

Concept

- Ces urls sont des identifiants permanents qui redirigent vers une page cible qui peut être modifiée
- Utilisation de PURL (Persistent Uniform Resource Locators)
- Service fourni par le Online Computer Library Center (consortium de bibliothèques)
- Permet d'assurer la stabilité de cet identifiant dans le temps
- Exemple : <http://purl.org/astronomy/vocab/Algorithms/AdaptiveMeshRefinement>

Création et mise à jour des vocabulaires

- Ils sont le plus souvent représentés en XML (assez verbeux)
- La mise à jour via un simple éditeur texte peut poser problème
- Pour simplifier ce processus, utilisation d'un logiciel spécialisé



- PoolParty est une application commerciale (15 k€/an) qui permet :
 - de créer et d'éditer des vocabulaires grâce à une interface web
 - d'accéder aux données via un wiki frontend
 - d'importer et d'exporter les données au format RDF
 - de valider la syntaxe de fichiers RDF
- Elle propose aussi des services web d'interrogation :
 - requête SPARQL
 - GetSemanticRelations
 - GetThesauri
 - GetConcepts
 - ...
- Licence académique

Création et mise à jour des vocabulaires

PROJECT DOCUMENTS TOOLS OPTIONS HELP ABOUT POOLPARTY

IvoaAlgorithms

- Astronomical Algorithms Vocabulary
 - Algorithm (71)
 - 3+1 Formalism (0)
 - Accelerated Lambda Iteration (0)
 - Adaptive Mesh Refinement (0)
 - Advection Upstream Splitting Method (0)
 - Alternating Direction Implicit (0)
 - Bulirsch-Stoer (0)
 - Coupled Escaped Probability (0)
 - Crank-Nicolson (0)
 - Escape Probability (1)
 - Euler (0)
 - Exact Radiative Transfer Method (0)
 - Exact Riemann Solver (0)
 - Fast-Multipole Method (0)
 - Finite Difference (9)**
 - Finite Element (0)
 - Finite Volume (3)
 - Fokker-Planck Solver (0)
 - Forward-Time Central-Space (0)
 - Fourier Technique (1)
 - Friends-Of-Friends (0)
 - Gauss-Seidel (0)
 - Gear Method (0)
 - Godunov (0)
 - Harten-Lax-van Leer (2)
 - Harten-Lax-van Leer-Contact (0)
 - Harten-Lax-van Leer-Einfeldt (0)
 - Hartree-Fock (0)
 - Henye (0)
 - Isochrones Synthesis (0)
 - Iterative Method (2)

Selected Concept

Finite Difference

http://purl.org/astronomy/vocab/ivoaAlgorithms/Finite_Difference

SKOS Metadata Linked Data Triples Visualization Geo

Broader Concepts

- [Algorithm](#)

Narrower Concepts

- [Crank-Nicolson](#)
- [Euler](#)
- [Forward-Time Central-Space](#)
- [Gear Method](#)
- [Lax-Friedrichs](#)
- [Lax-Wendroff](#)
- [Leap Frog](#)
- [MacCormack](#)
- [Runge-Kutta](#)

Related Concepts

Exact Matching Concepts

Close Matching Concepts

Preferred Label ([translate](#))

- Finite Difference

Alternative Labels

Hidden Labels

Notation

Scope Notes

Definitions

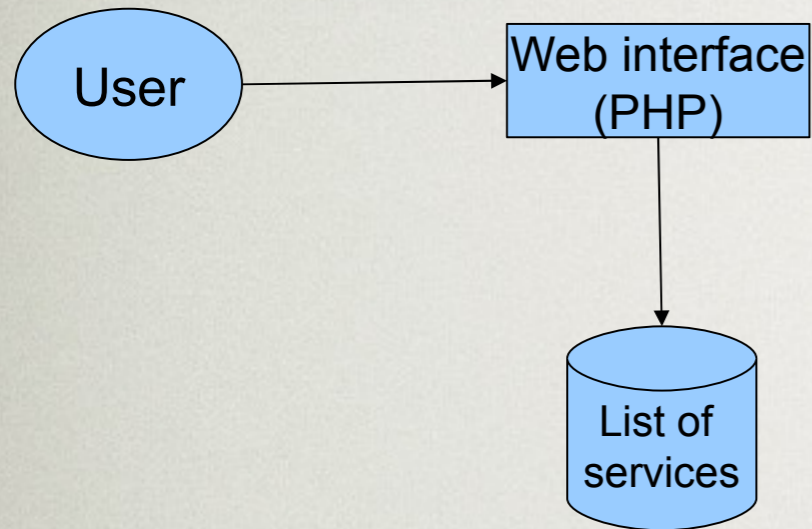
Création et mise à jour des vocabulaires

- Nous disposons d'un compte sur le serveur hébergé par l'éditeur
- Pas d'accès au serveur ou au code de l'application
- L'objectif est uniquement de simplifier la mise à jour et de centraliser les vocabulaires
- Aucun couplage avec cette application
- Les vocabulaires sont ensuite exportés vers notre propre serveur pour être utilisés
- Vocabulaires consultables : <http://votheory.obspm.fr/terms>

Implémentation : le site Votheory

- Utilisation de SKOS concepts sur les bases PDR et Starformat
- Classes concernées :
 - Physics
 - Algorithms
 - TargetObjectTypes
- L'objectif est de trouver des protocoles et des projets
- Les relations entre concepts ne sont pas prises en compte

Architecture



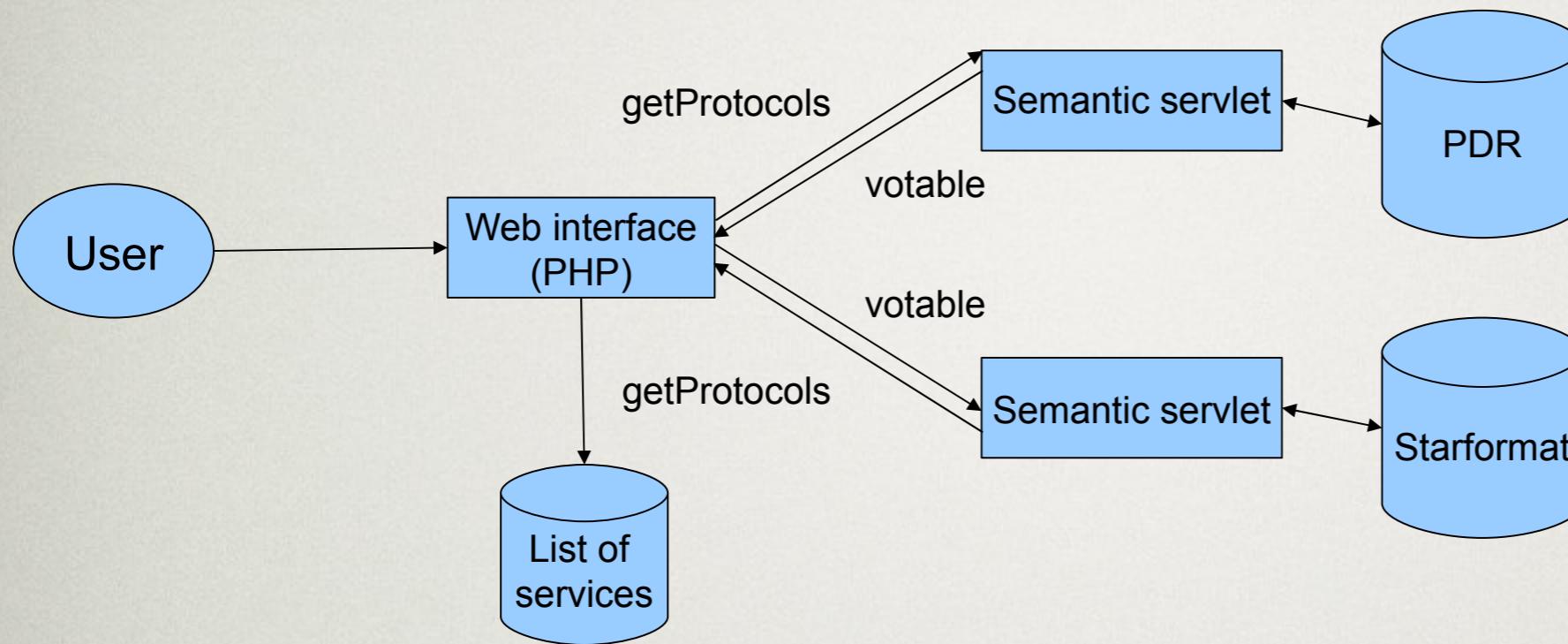
1. L'utilisateur entre sa requête

2. Une requête pour trouver les protocoles est envoyée à tous les services de la liste

3. Une requête pour trouver les projets est envoyée à tous les services de la liste

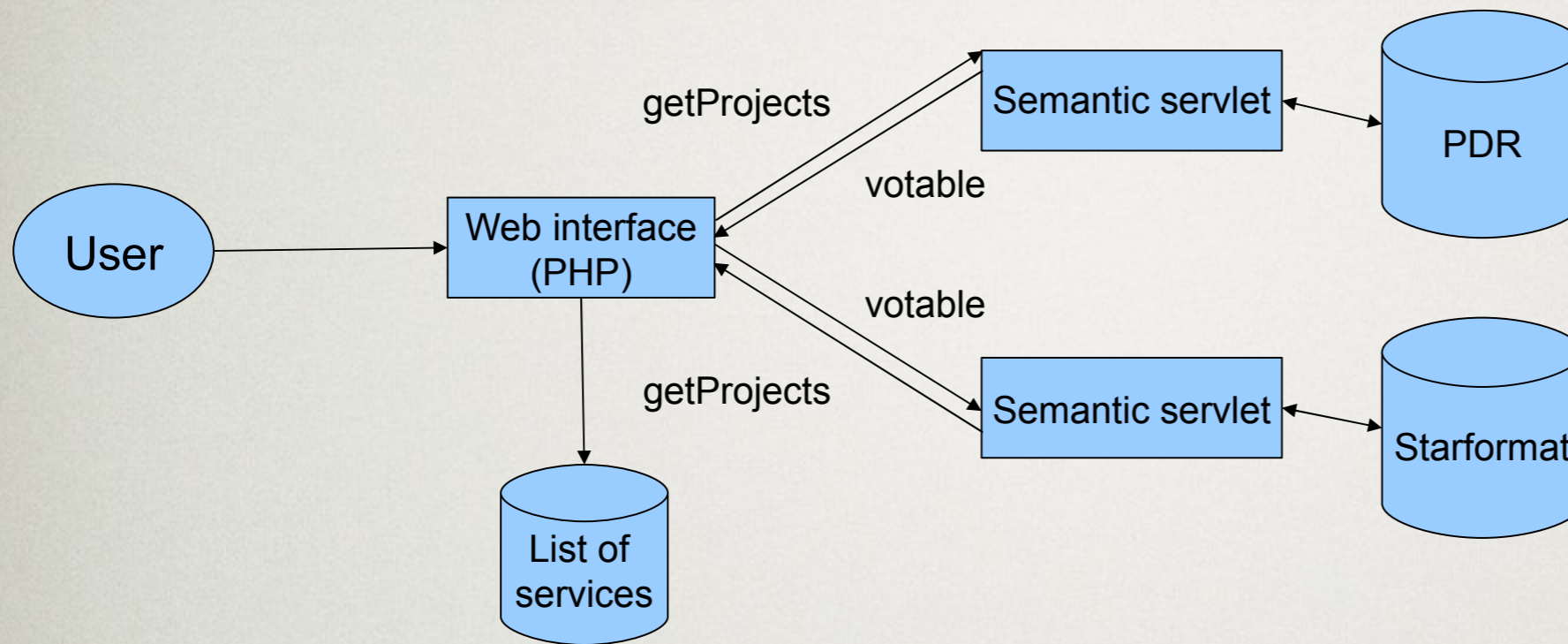
4. Les VOTables sont affichées

Architecture



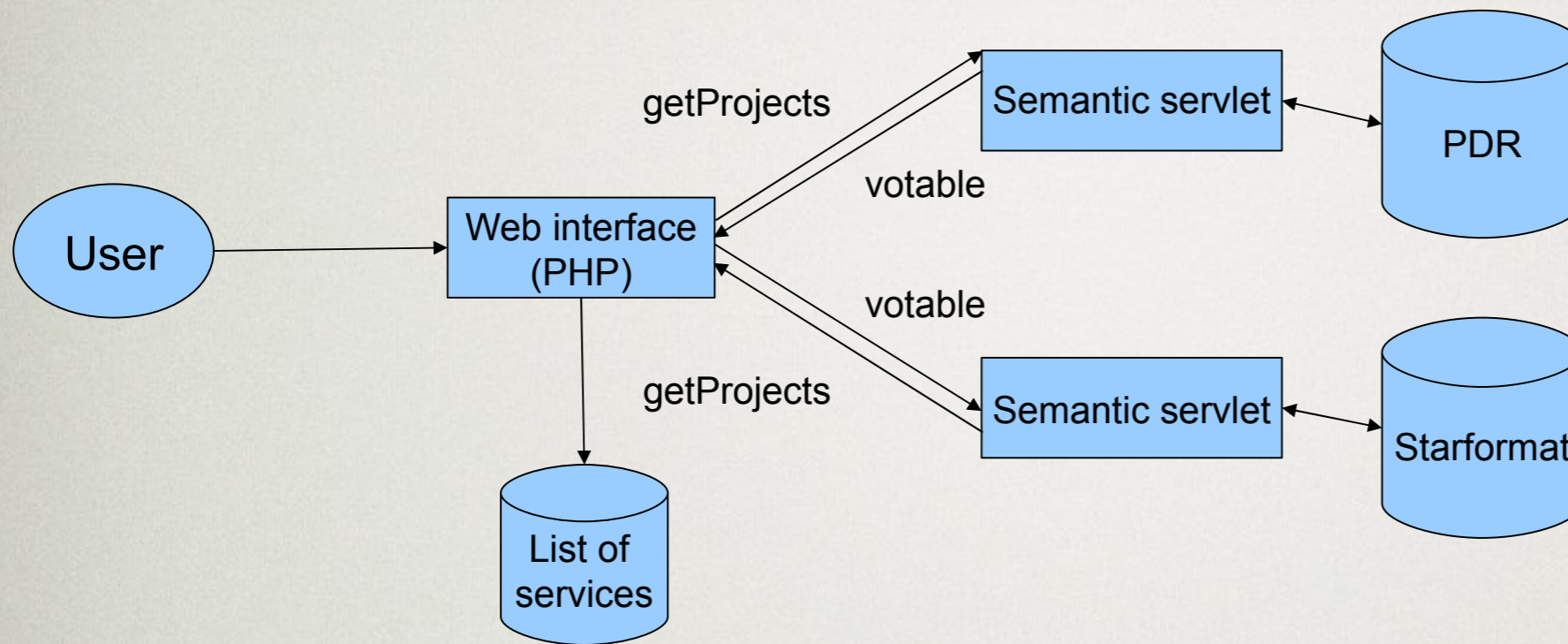
1. L'utilisateur entre sa requête
2. Une requête pour trouver les protocoles est envoyée à tous les services de la liste
3. Une requête pour trouver les projets est envoyée à tous les services de la liste
4. Les VOTables sont affichées

Architecture



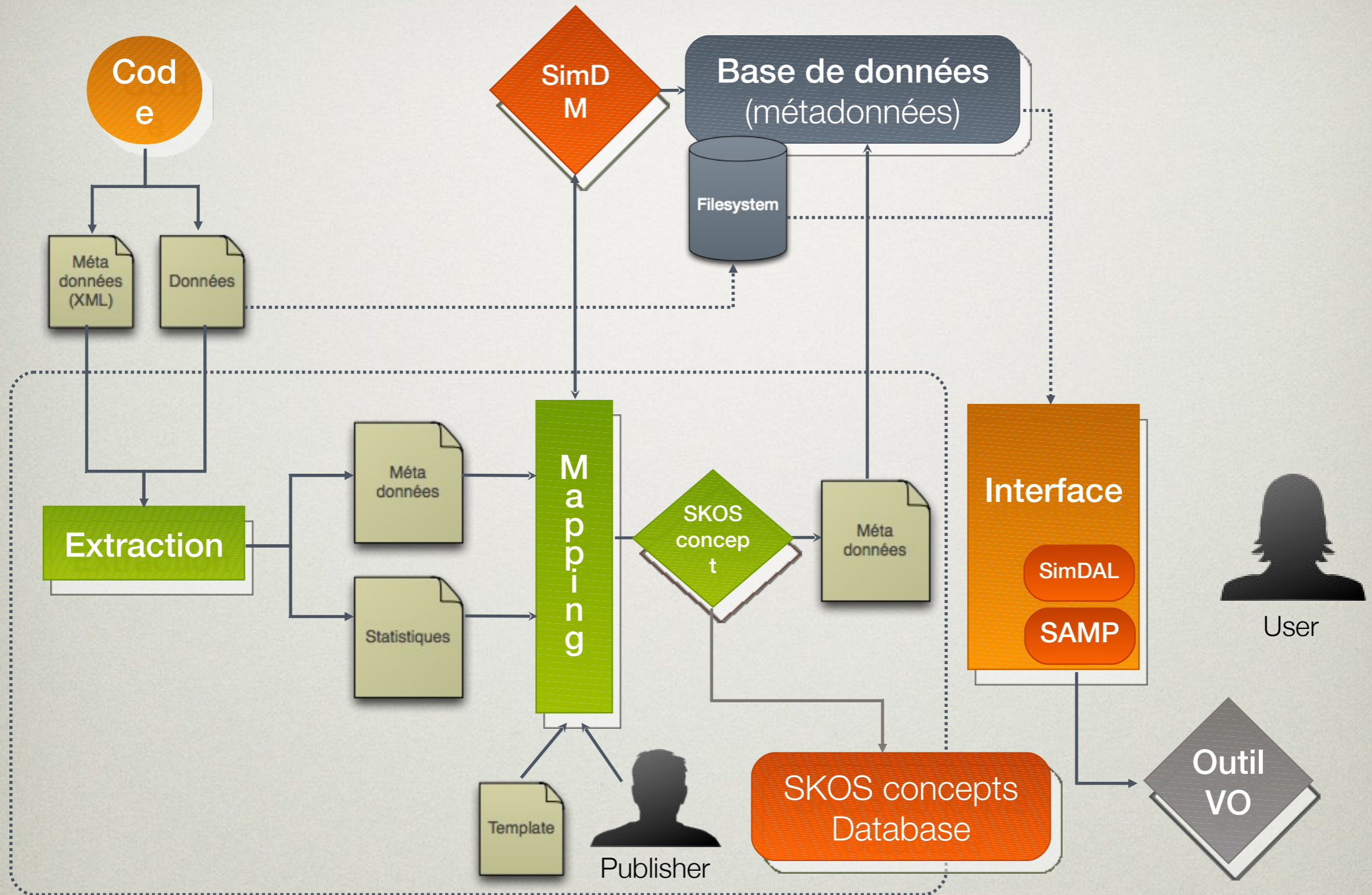
1. L'utilisateur entre sa requête
2. Une requête pour trouver les protocoles est envoyée à tous les services de la liste
3. Une requête pour trouver les projets est envoyée à tous les services de la liste
4. Les VOTables sont affichées

Architecture



1. L'utilisateur entre sa requête
2. Une requête pour trouver les protocoles est envoyée à tous les services de la liste
3. Une requête pour trouver les projets est envoyée à tous les services de la liste
4. Les VOTables sont affichées

Ingestion des concepts dans SimDB



- Les concepts doivent être ajoutés aux métadonnées dans le pipeline d'ingestion
- Leur validité peut être testée directement via leur URI
- Retourne soit le xml du concept soit une table d'erreur

Services disponibles

- Lister les vocabulaires
- Retourner vocabulaire complet en RDF
- Recherche de concepts par URI
- Recherche de concepts par prefLabel/altLabel
- Exemple d'application : Dictionnaire de termes astronomiques