

# CALLISTO

*Un portail sémantique pour le partage de données  
scientifiques*

CAPITOUL, 6 mai 2021



CALMIP (UAR 366)  
Espace Clément Ad  
[www.calmip.univ-toulouse](http://www.calmip.univ-toulouse)

## 1. Portail CALLISTO

- Contexte général : science ouverte
- Positionnement par rapport aux portails existants
- Fonctionnalités
- Aspects techniques : l'ontologie sous-jacente

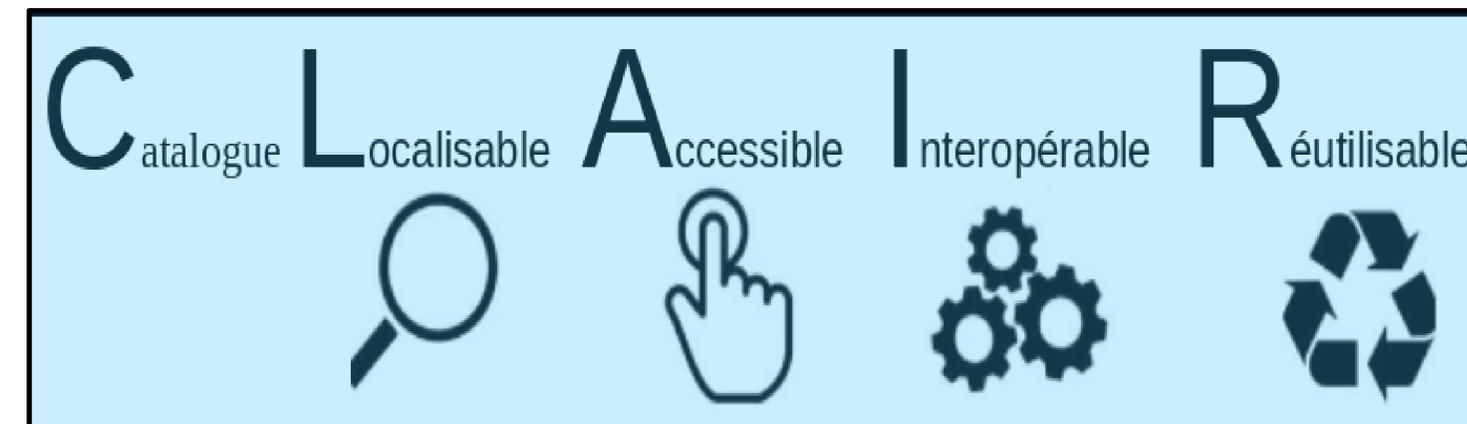
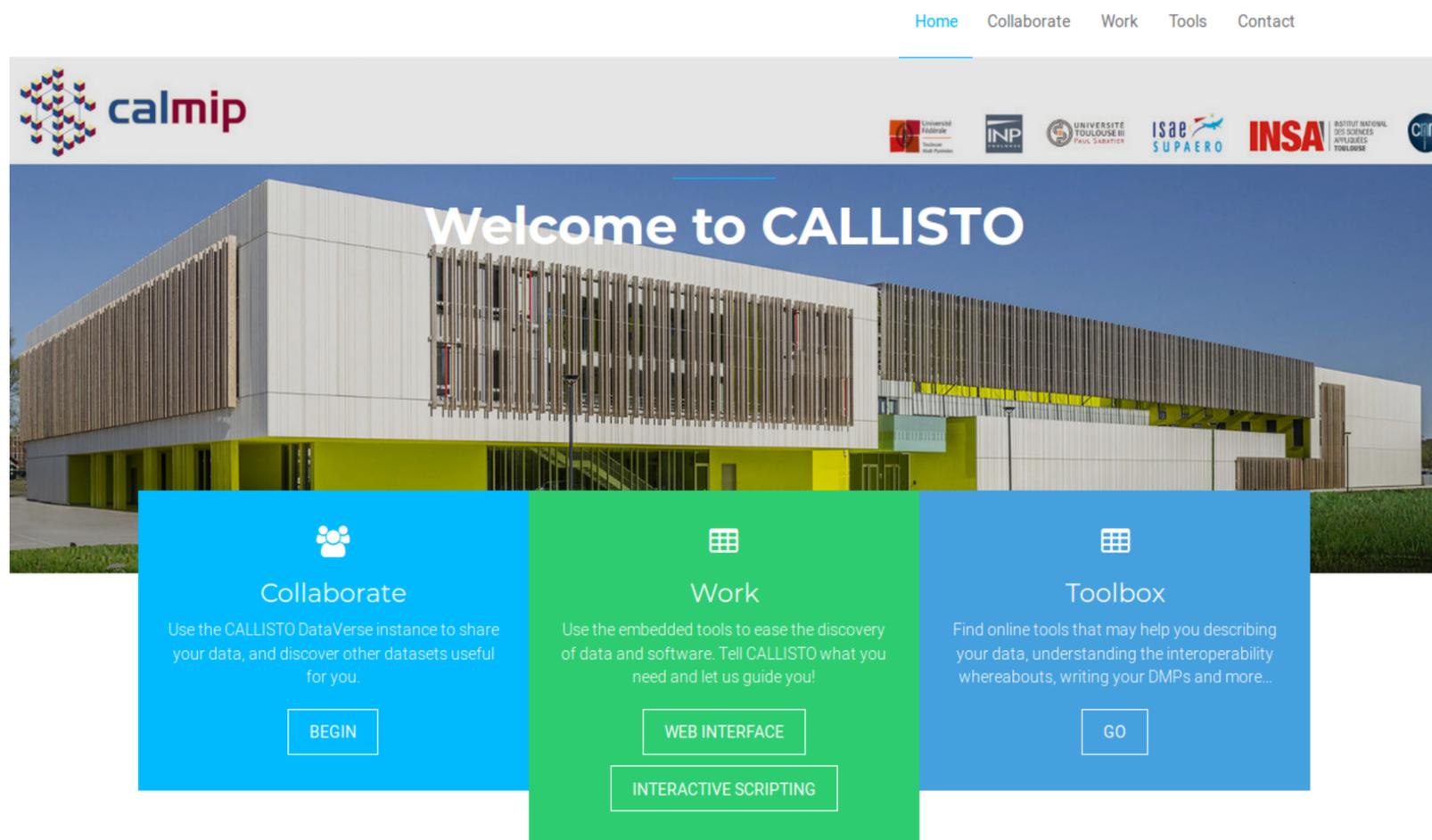
## 2. Architecture en conteneurs

- Rôle de chacun des containers
- Spécificités

Proxy, websockets, transport des variables d'identifiants Shibboleth...

## 3. Diffusion : Git et ansible

- Rôle, bénéfices et limites.



## Ce service de CALMIP s'inscrit dans un contexte:

- Adoption large des principes **FAIR/CLAIR**
  - Trouver/proposer des données bien documentées et respectant les standards applicables
- Mouvement vers une **science « ouverte »**
  - Articles, données affiliées disponibles comme un ensemble permettant la reproductibilité des résultats

## D'un point de vue pratique, ce contexte a des conséquences:

- Il est demandé aux scientifiques (**par l'ANR, les projets Européens...**) de produire un Plan de Gestion de Données (PGD) détaillant -entre autres- les modalités de partage (**dépôts**), de description (**métadonnées**) et de standardisation (**formats, données et métadonnées**) adoptées au sein de leur projet.

### **FAIR / CLAIR** : « Principes-guides » pour améliorer la ré-utilisabilité des données.

- Identifier les jeux de données (DOI par exemple)
- Donner accès à ces jeux de données à travers des interfaces « connues » (ouverture plus large que celle possible au travers de sites peu référencés ou des bases de données utilisées par quelques dizaines de personnes)
- Décrire ces données de façon standardisée (suivant des pratiques disciplinaires par exemple)
- Adopter des formats répandus

### Référence :

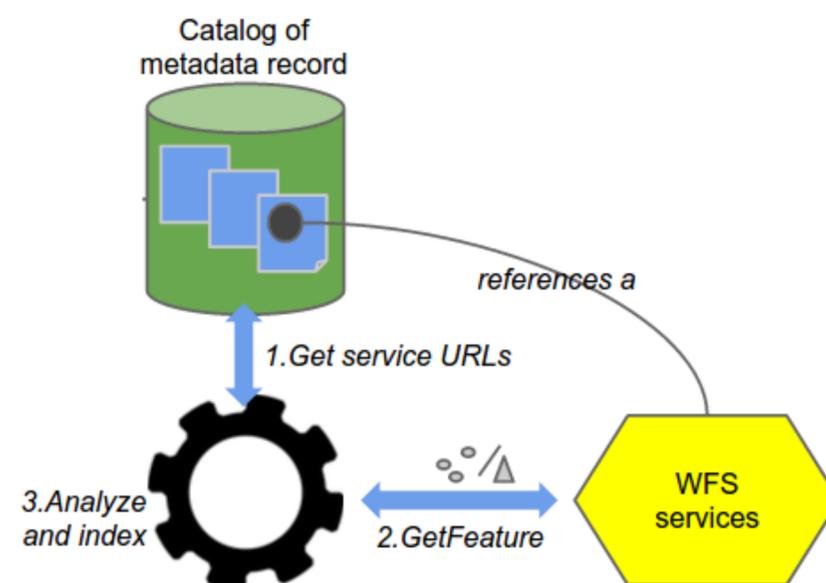
- <https://www.calmip.univ-toulouse.fr/spip.php?article718>
- <https://callisto.calmip.univ-toulouse.fr/callisto/tools.html>

**Métadonnées** : informations à propos des données.

- Elles peuvent se trouver en-dehors des données (dans une interface Web, un document externe...) comme dans les fichiers eux-mêmes (ex : hdf5).
- Il existe des ensembles de métadonnées très répandus (DublinCore, essentiellement bibliographique) et aussi des ensembles moins répandus mais très utilisés au sein de certaines disciplines (OGC, IVOA...).
- Exemple : Catalogue Web de métadonnées pour des données géographiques → Interrogation de métadonnées issues de plusieurs sources (<https://github.com/geonetwork>)
- Exemple d'implémentation technique

### Data accessible using download services (WFS)

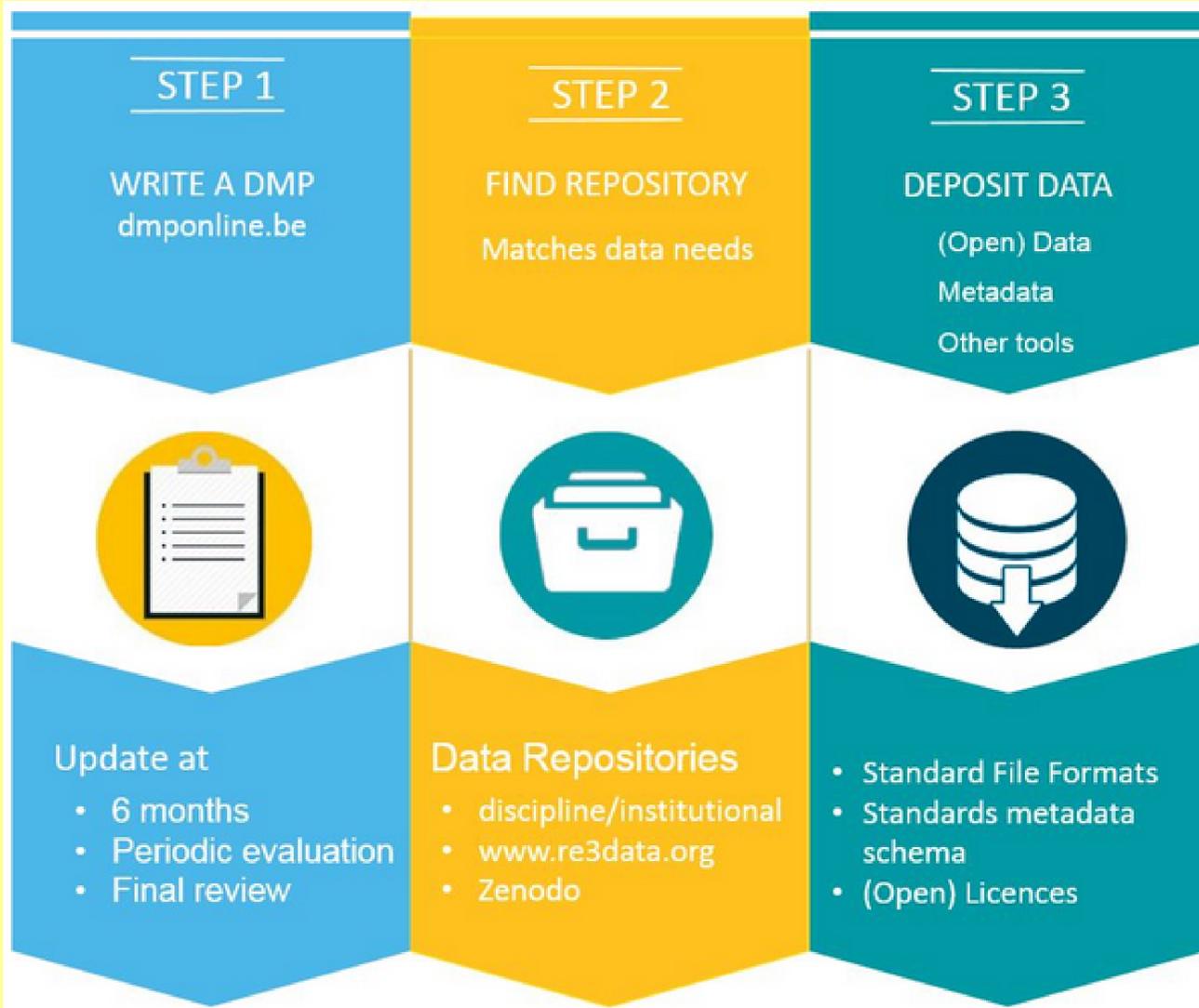
All datasets described in metadata records and accessible through a download services (ie WFS) can be analyzed to improve search and data visualization. The catalog collect WFS features and index them.



<https://geonetwork-opensource.org/manuals/trunk/fr/user-guide/analyzing/data.html>

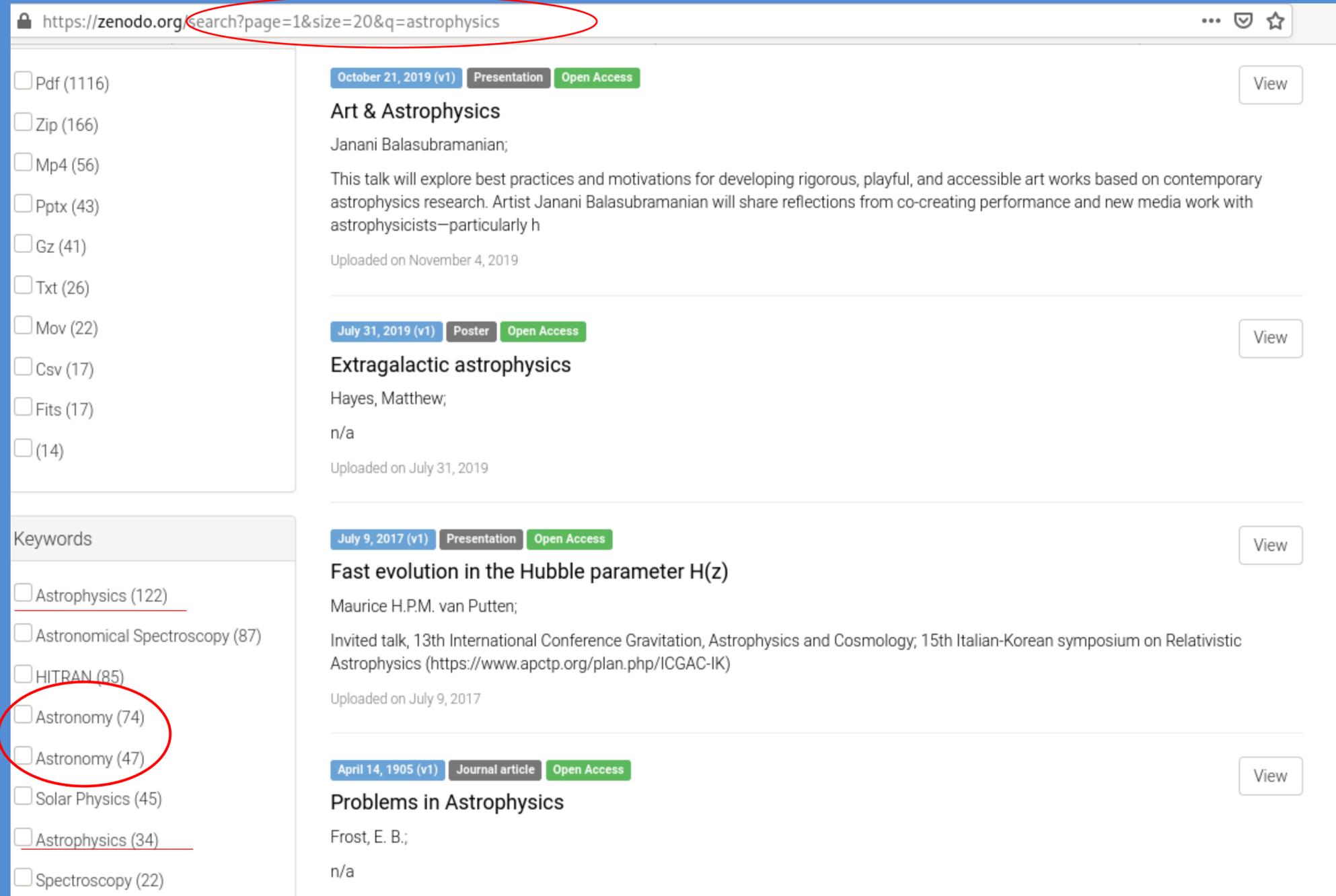
# Contexte général : panorama résumé.

Source: OpenAIRE <https://www.openaire.eu>



DATAVERSE REPOSITORIES - A WORLD VIEW





The screenshot shows a Zenodo search results page for the query 'astrophysics'. The URL in the browser's address bar is <https://zenodo.org/search?page=1&size=20&q=astrophysics>, which is circled in red. On the left side, there is a sidebar with file format filters (Pdf, Zip, Mp4, Pptx, Gz, Txt, Mov, Csv, Fits, and a general category) and a 'Keywords' section. In the 'Keywords' section, 'Astronomy (74)' is circled in red. The main content area displays three search results, each with a 'View' button:

- Art & Astrophysics**  
Janani Balasubramanian;  
This talk will explore best practices and motivations for developing rigorous, playful, and accessible art works based on contemporary astrophysics research. Artist Janani Balasubramanian will share reflections from co-creating performance and new media work with astrophysicists—particularly h  
Uploaded on November 4, 2019
- Extragalactic astrophysics**  
Hayes, Matthew;  
n/a  
Uploaded on July 31, 2019
- Fast evolution in the Hubble parameter  $H(z)$**   
Maurice H.P.M. van Putten;  
Invited talk, 13th International Conference Gravitation, Astrophysics and Cosmology, 15th Italian-Korean symposium on Relativistic Astrophysics (<https://www.apctp.org/plan.php/ICGAC-IK>)  
Uploaded on July 9, 2017
- Problems in Astrophysics**  
Frost, E. B.;  
n/a

L'interface de CALLISTO propose un espace de partage de données protégé par identifiant/mot de passe, enrichi d'une description précise des données hébergées afin d'améliorer leur ré-utilisabilité.

Le positionnement est celui d'un portail à un niveau de détails très fins (celui d'un projet scientifique) éventuellement accompagné d'une description ontologique du contexte scientifique.

### Ø Outils logiciels mis à disposition :

- Entrepôt de données
- Formulaires de recherche de données
- Mise en place de fonctionnalités spécifiques (requiert la construction d'une ontologie, service en cours d'élaboration avec les utilisateurs)

### Ø Support technique associé :

- Assistance à la prise en main de l'interface
- Collaboration étroite pour la construction d'une ontologie associée aux données

CALLISTO est en développement continu, sa construction se fait avec les utilisateurs. Les réalisations menées jusqu'ici sont le résultat de la collaboration de CALMIP avec les membres des projets de calcul.

## Deux projets-types autour desquels CALLISTO évolue aujourd'hui :

### ➤ SMS Smart Morphing and Sensing

**Projet européen en mécanique des fluides, plusieurs intervenants internationalement.**

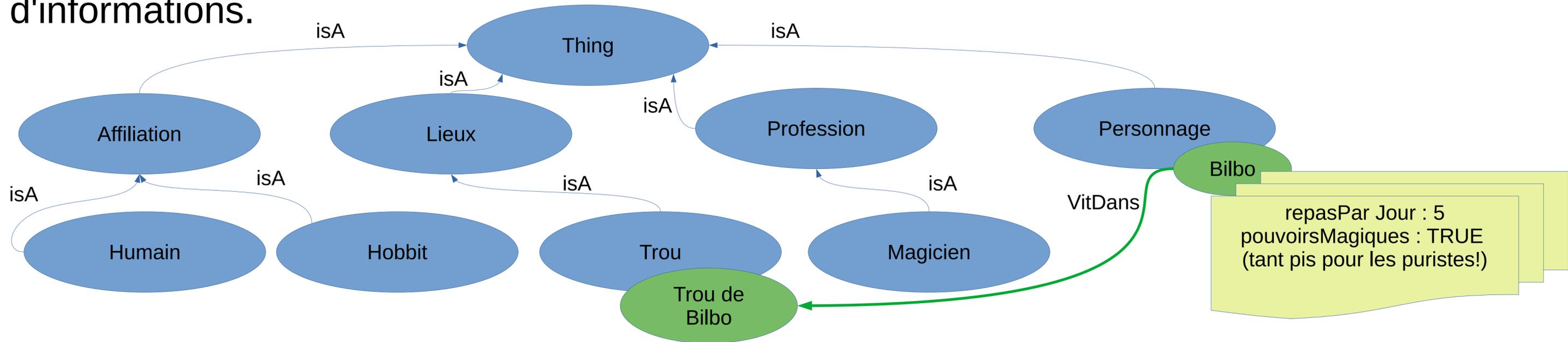
- Besoins exprimés : Espace de partage
- Travail effectué : ouverture d'un espace de dépôt soutenu par des métadonnées reliées à une ontologie spécifique construite par ISMS et CALMIP
- Travail en cours : amélioration de l'ontologie, conception d'outils de recherche de données avancés.

### ➤ HiperBorea

**Projet ANR, géophysique**

- Besoins exprimés : aide à la rédaction de PGD
- Travail effectué : proposition de plusieurs formats de métadonnées.  
Choix de l'un d'entre eux, construction d'une ontologie permettant la génération de champs de métadonnées pour DataVerse correspondant à ce format.
- Travail en cours : Choix des solutions de partage et de stockage, informations demandées dans le PGD

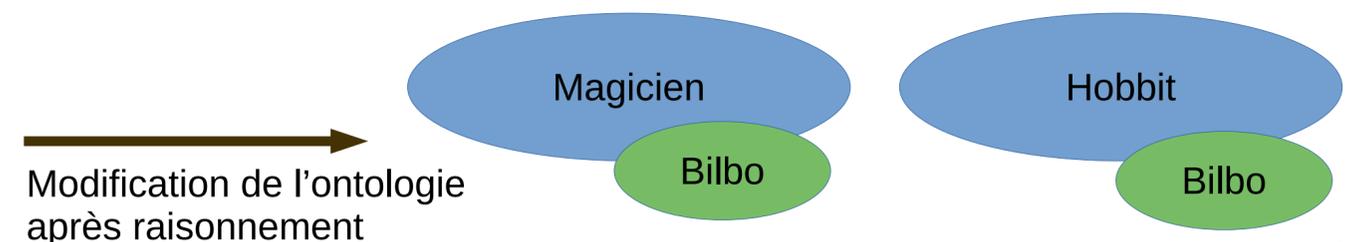
Une **ontologie** est l'ensemble structuré des termes et concepts représentant le sens d'un champ d'informations.



L'ontologie constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que des relations entre ces concepts. Elle est employée pour raisonner à propos des objets du domaine concerné.

« *Spécification d'une conceptualisation partagée* » (Gruber, 1993)

Exemple de conditions exprimées :  
 Personnages and ((VitDans some Trous) and (repasParJour min 5 xsd:int)) → hobbit  
 Personnages and (pouvoirsMagiques value true) → magicien



# L'ontologie sous-jacente

Titre du jeu de données

Dataverse natif

**forces aérodynamiques morphing 2D - alpha=2.5 deg - f=350 Hz** Draft Unpublished

TÔ, Jean-Baptiste, 2020, "forces aérodynamiques morphing 2D&nbsp;- alpha=2.5 deg&nbsp;- f=350 Hz", <https://doi.org/10.5072/FK2/S4UL3Z>, Root, DRAFT VERSION

**Description** Fichier au format ASCII. Forces aérodynamiques intégrées autour d'une aile d'A320 de corde c = 0.15 m, d'incidence alpha = 2.5°. Ce fichier a été généré à l'issue d'une simulation CFD de formulation volumes finis (VF) dans une configuration bidimensionnelle (2D). Paramètres : Type de morphing = vibration sinusoïdale du bord de fuite Longueur de la déformation = 5c/100 Amplitude de la déformation = +/-1° autour du profil neutre Fréquence de vibration = 350 Hz Nombre de Mach à l'infini = 0.78 Nombre de Reynolds = 2.06e6 Température à l'infini = 261.22 K Pression à l'infini = 67574 Pa Viscosité = suit le modèle de Sutherland Descriptif des données : colonne 1 = t (en secondes) colonne 2 = Cx colonne 3 = Cz colonne 4 = Cm

**Subject** Physics; Other

**Keyword** T1 morphing, NSMB, Cx\_Cz\_Cm.plt

Files Metadata Terms Versions

Dataverse natif

**Métadonnées générales**

Citation Metadata

**Dataset Persistent ID** doi:10.5072/FK2/S4UL3Z

**Title** forces aérodynamiques morphing 2D&nbsp;- alpha=2.5 deg&nbsp;- f=350 Hz

**Author** TÔ, Jean-Baptiste

**Contact** Use email button above to contact.  
TÔ, Jean-Baptiste

**Description** Fichier au format ASCII. Forces aérodynamiques intégrées autour d'une aile d'A320 de corde c = 0.15 m, d'incidence alpha = 2.5°. Ce fichier a été généré à l'issue d'une simulation CFD de formulation volumes finis (VF) dans une configuration bidimensionnelle (2D). Paramètres : Type de morphing = vibration sinusoïdale du bord de fuite Longueur de la déformation = 5c/100 Amplitude de la déformation = +/-1° autour du profil neutre Fréquence de vibration = 350 Hz Nombre de Mach à l'infini = 0.78 Nombre de Reynolds = 2.06e6 Température à l'infini = 261.22 K Pression à l'infini = 67574 Pa Viscosité = suit le modèle de Sutherland Descriptif des données : colonne 1 = t (en secondes) colonne 2 = Cx colonne 3 = Cz colonne 4 = Cm

**Subject** Physics; Other

**Keyword** T1 morphing (SMS Ontology) <https://callisto.calmip.univ-toulouse.fr/OntoWiki/>  
NSMB (SMS Ontology) <https://callisto.calmip.univ-toulouse.fr/OntoWiki/>  
Cx\_Cz\_Cm.plt (SMS Ontology) <https://callisto.calmip.univ-toulouse.fr/OntoWiki/>

**Métadonnées générales**

Related Publication

Citation

ID Type Select... ID Number

URL Enter full URL, starting with http://

Notes

Depositor Admin, Dataverse

Deposit Date 2020-05-15

Spécifiques à un projet

Métadonnées spécifiques générées à partir de l'ontologie

SMS\_04Mars2020\_v3

None

ssor relaxation None

unsteady None

flapping distance None

limiter variable flag None

time scheme Select...

alpha None

updblock implicit None

ris None

pointwise implicit Select...

cmv None

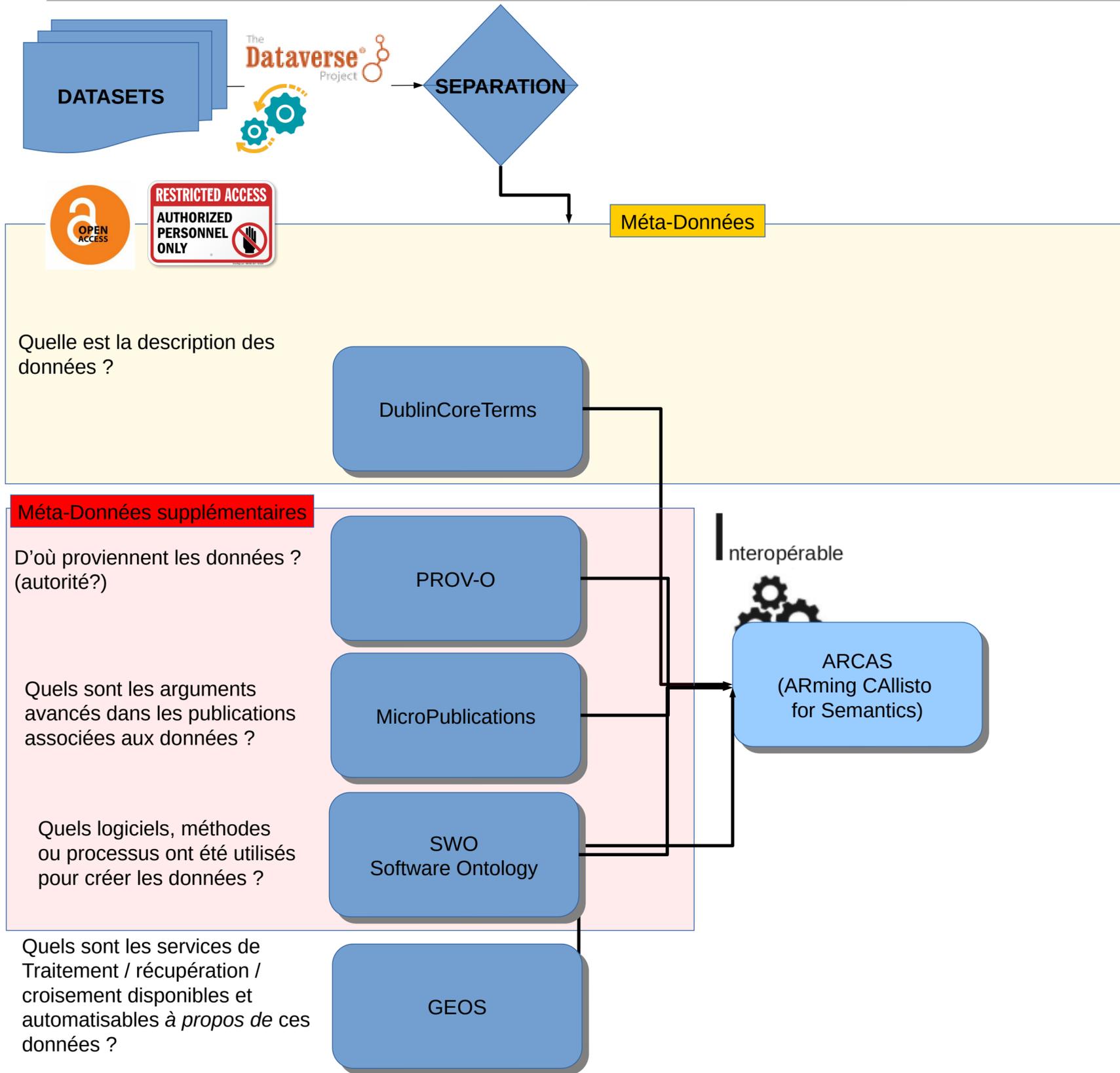
mesh deformation flag None

wall distance flag Select...

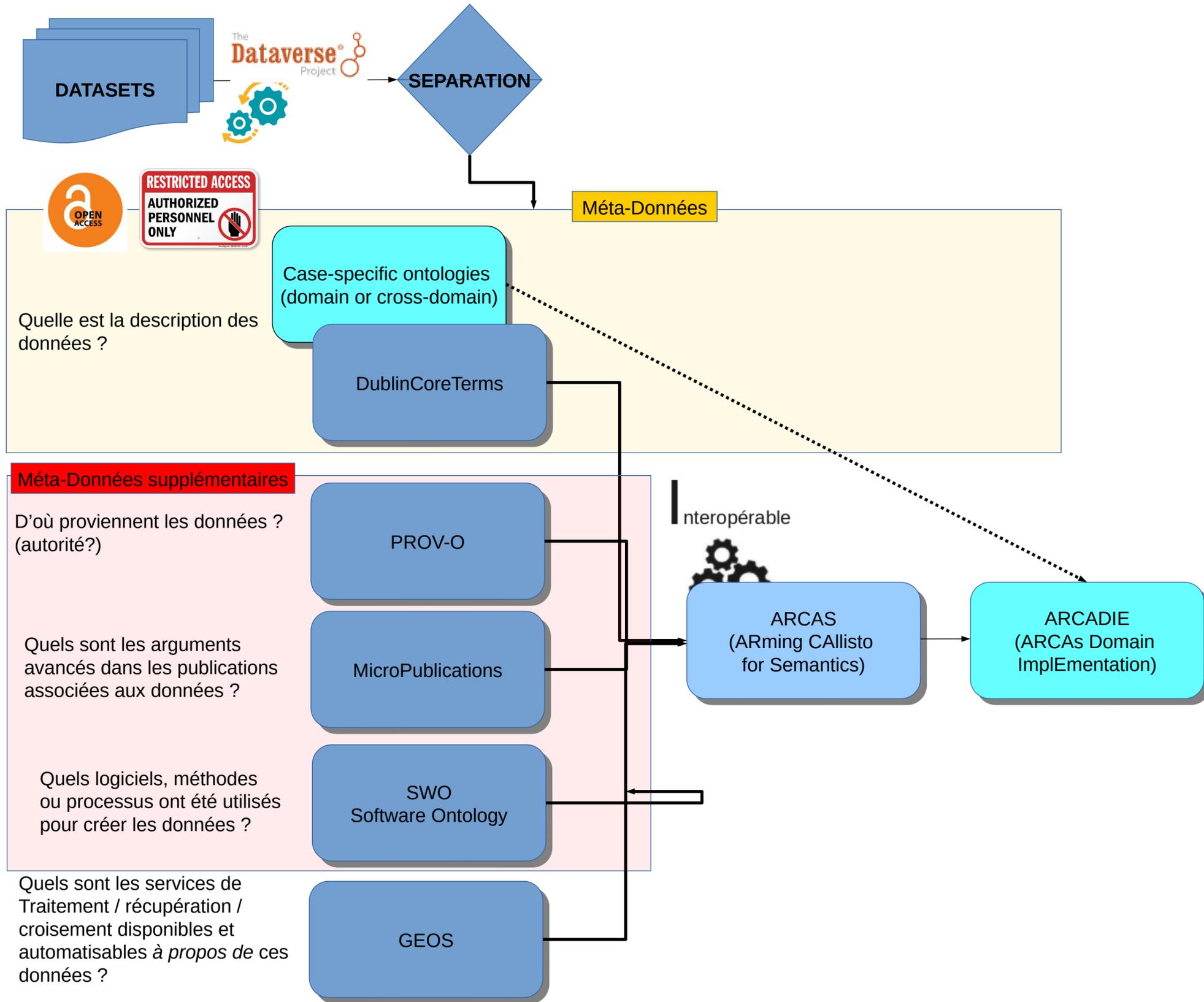
chordofairfoil None

pressure infinity None

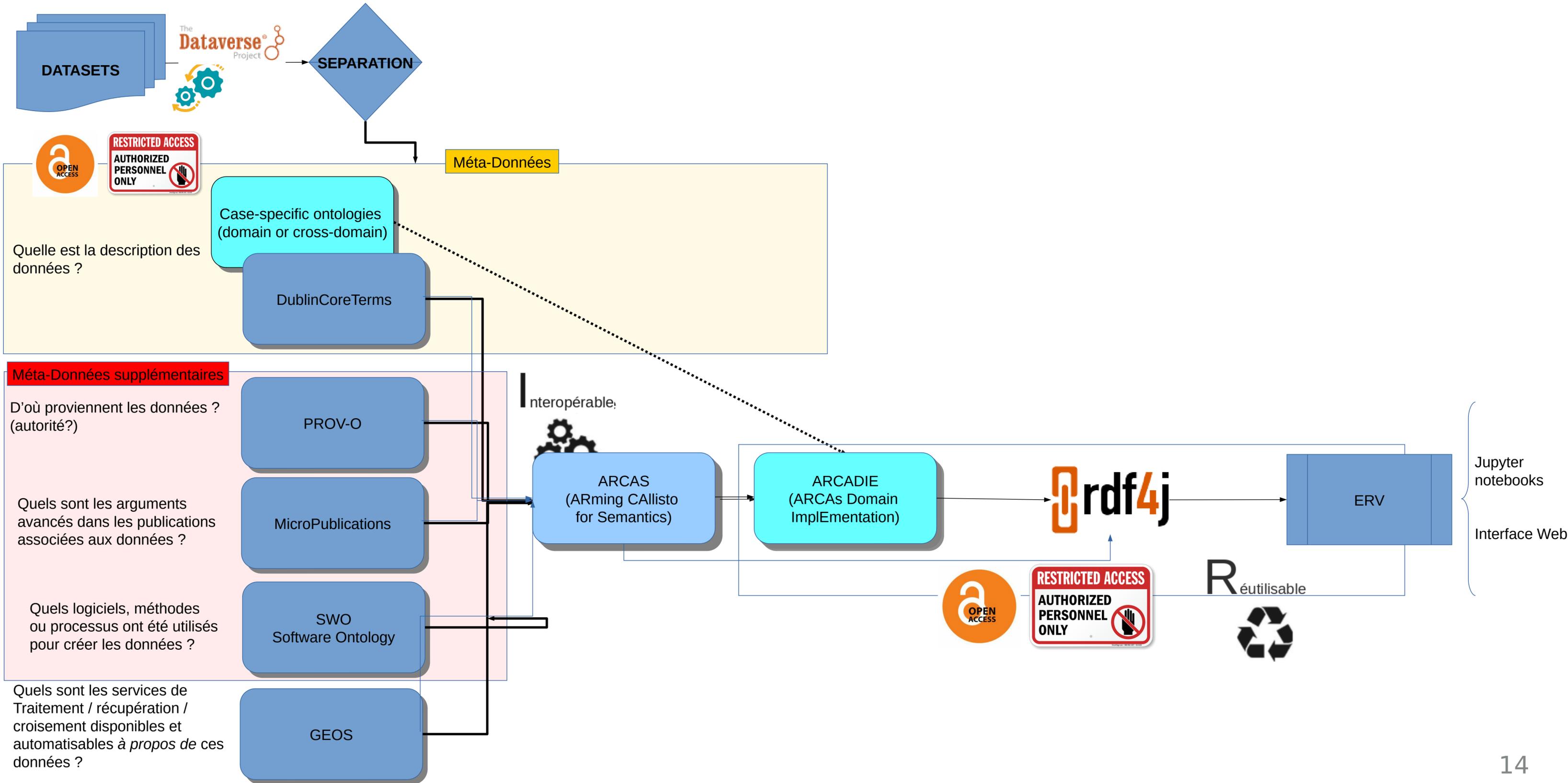
# L'ontologie sous-jacente



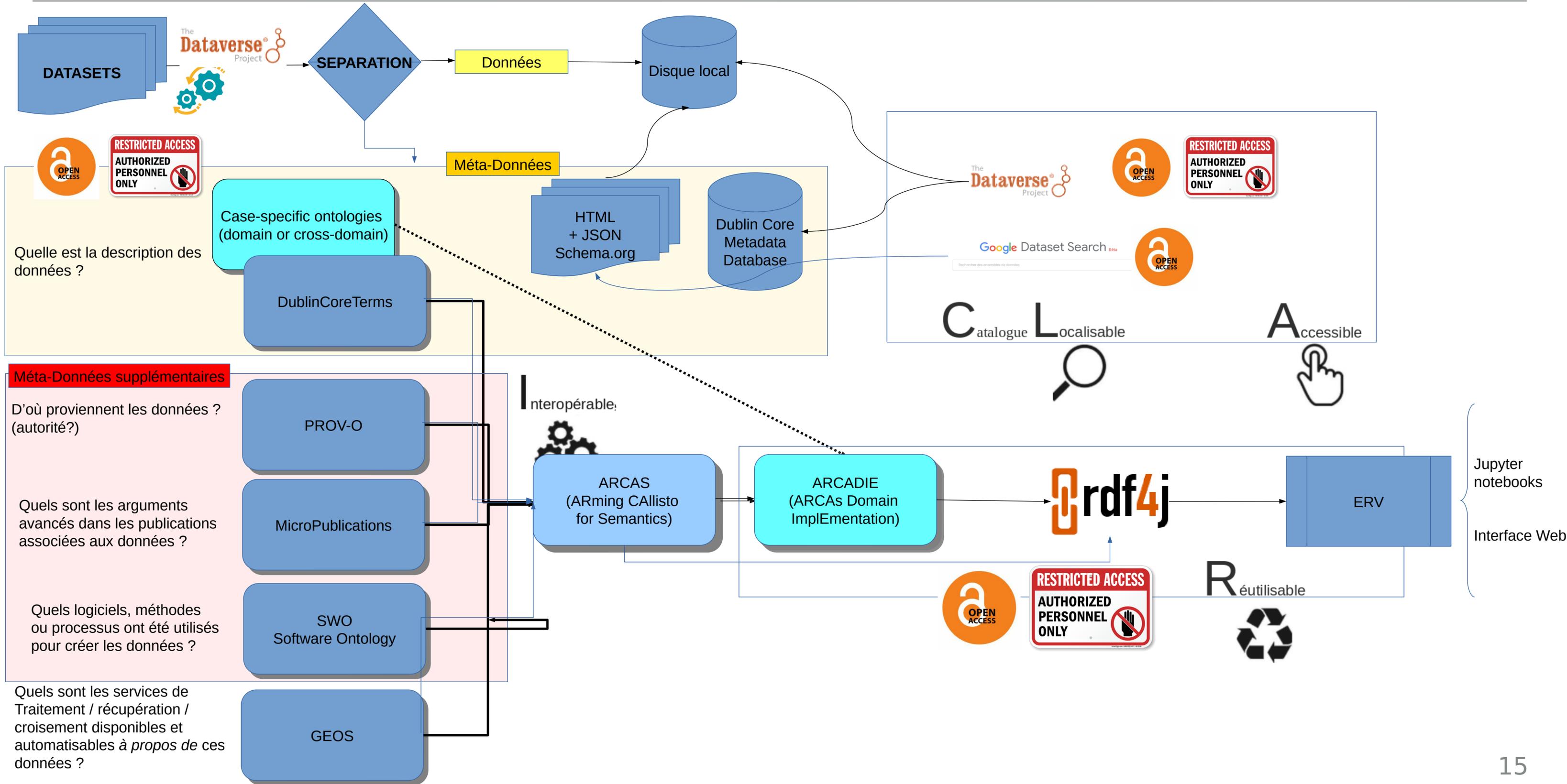
# L'ontologie sous-jacente



# L'ontologie sous-jacente



# L'ontologie sous-jacente



---

**CALLISTO – CALmip** Launches an **Interface for Semantic Tollbox Online** est un portail qui permet :

- De partager des données de manière sécurisée
- D'enrichir le partage avec des métadonnées pertinentes
- De mettre automatiquement en relation données, articles, logiciels et traitements

Et ceci dans un contexte et objectif (FAIR) de ré-utilisation de ces données.

L'ensemble des tutoriels : <https://www.calmip.univ-toulouse.fr/spip.php?article751>

## Portail Callisto

- **Architecture en conteneurs**

- Motivations
- Pour aller plus loin
- Quelques points techniques

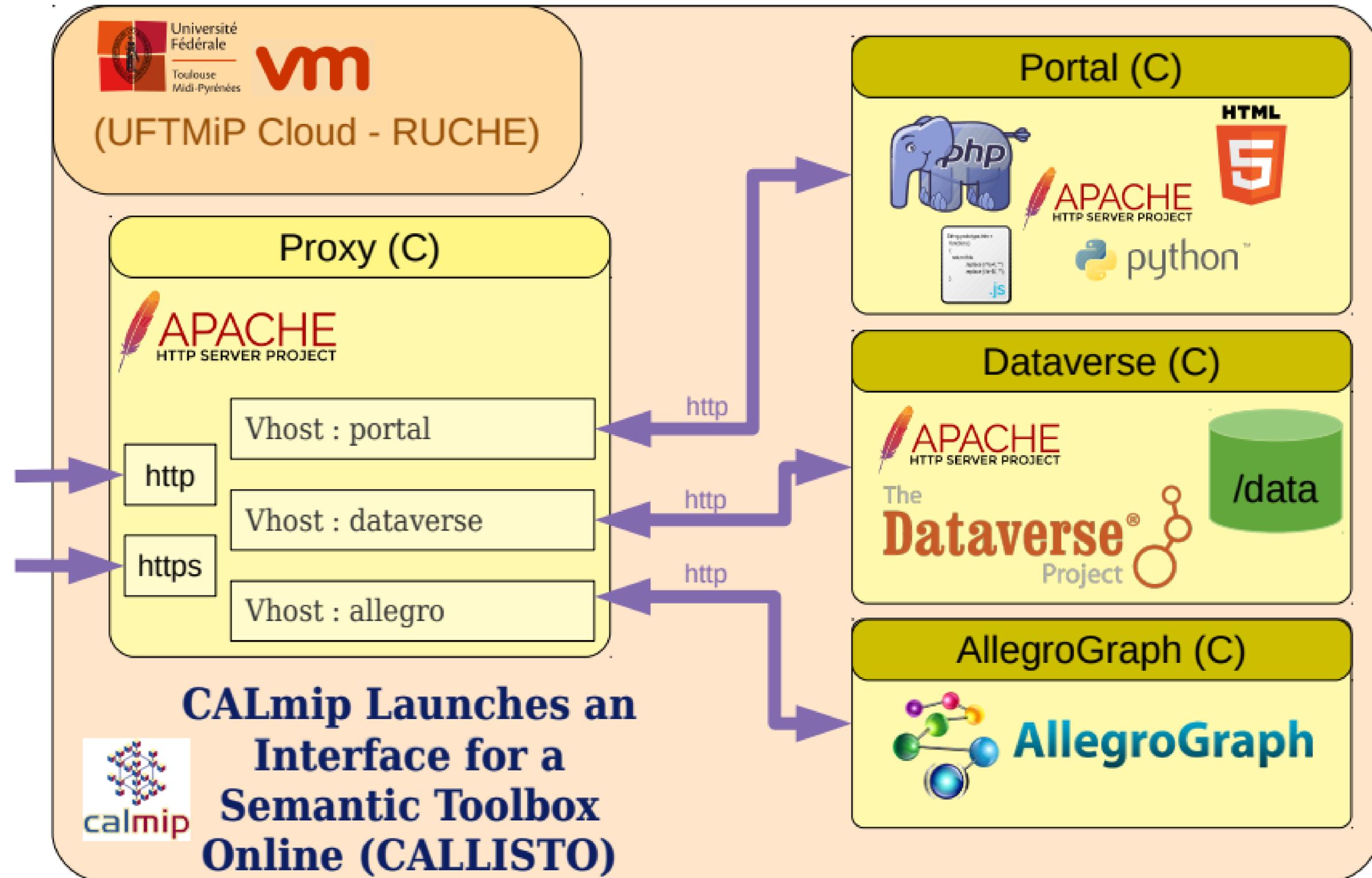
- **Shibboleth**

- **Diffusion avec git et ansible**

- Rôle, bénéfices et limites.

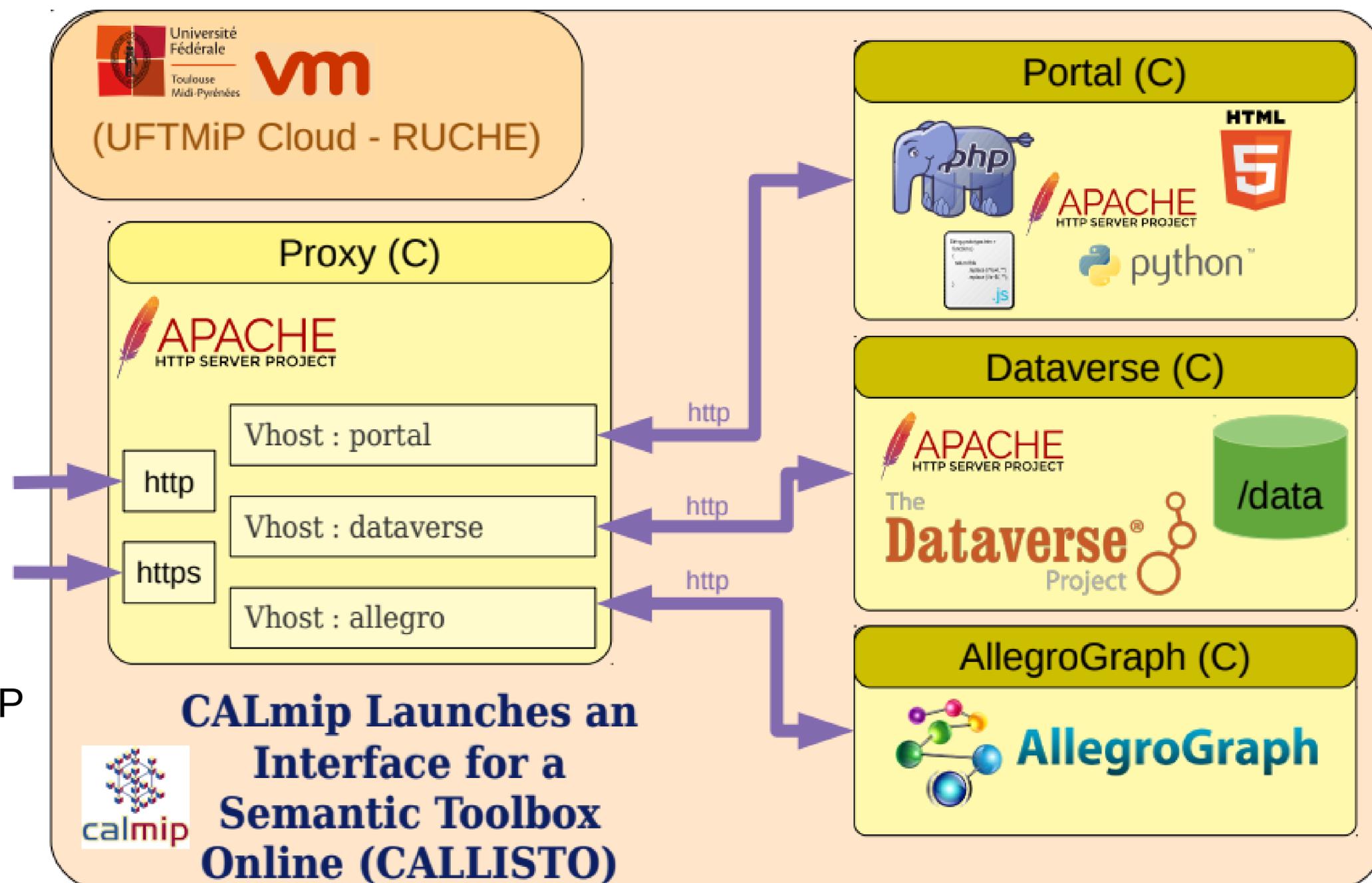
## • Architecture

- Une seule Machine Virtuelle
- Un conteneur proxy
- Un conteneur par brique logicielle



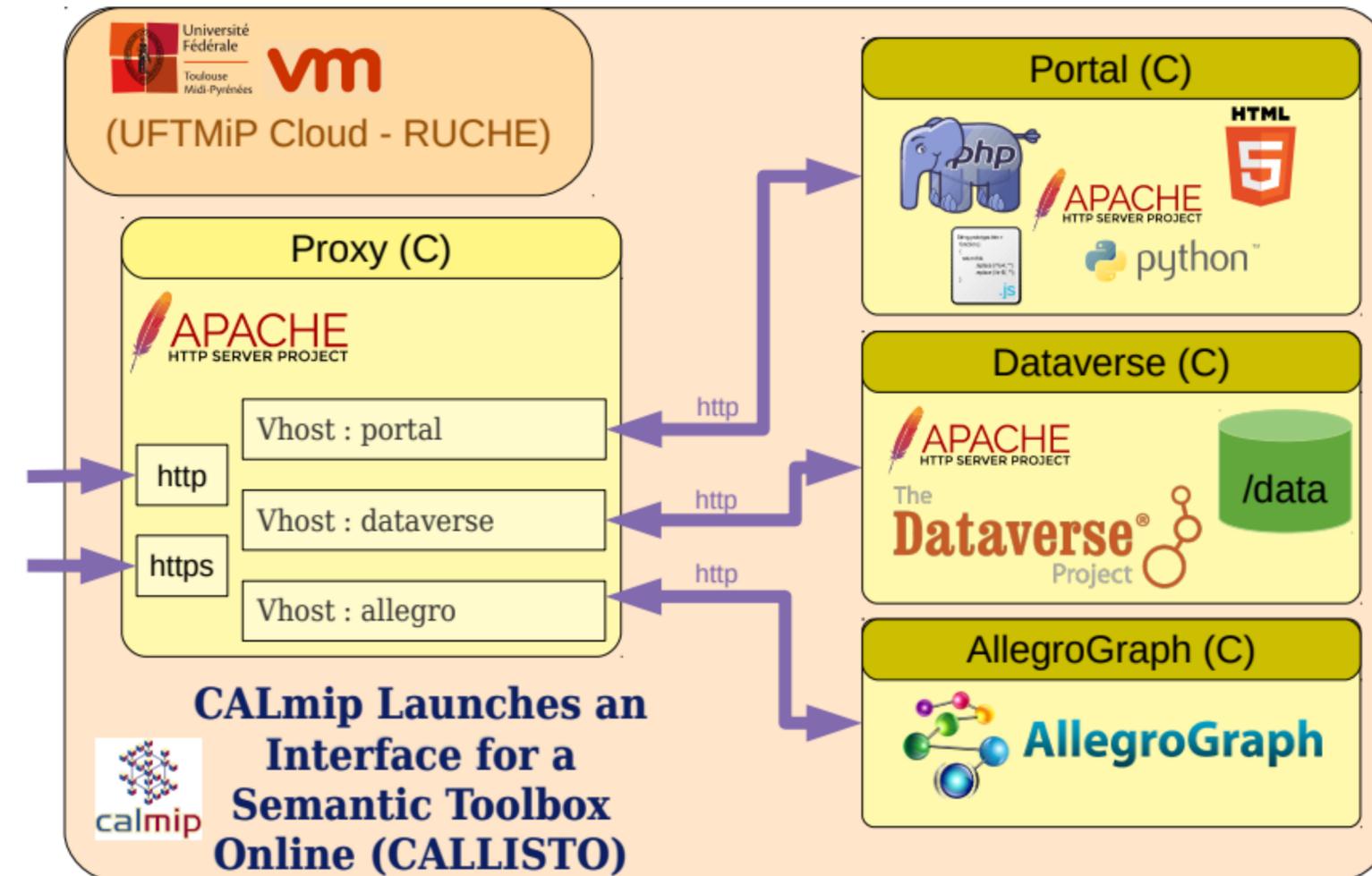
## Motivations

- **Modulaire :**
  - Un conteneur = une application
  - Briques remplaçables
- **Adaptation aux besoins :**
  - Développements maison
  - Applications externes open source
  - Applications externes propriétaire
- **Autonome :**
  - Pas besoin des admins sys
  - Pas besoin des collègues de l'UFTMiP
- **Simple et rapide :**
  - Un nouveau conteneur = ~1 minutes



## • Et pour aller plus loin ?

- Stockage permettant d'économiser l'espace disque (btrfs, zfs, ...)
- Déploiement automatisé (ansible, ...)



### • **Quelques points techniques**

- Un tunnel IPsec entre le cloud UFTMiP et l'Espace Clément Ader
- Un wildcard DNS
- Proxy websocket
- Authentification et proxy shibboleth

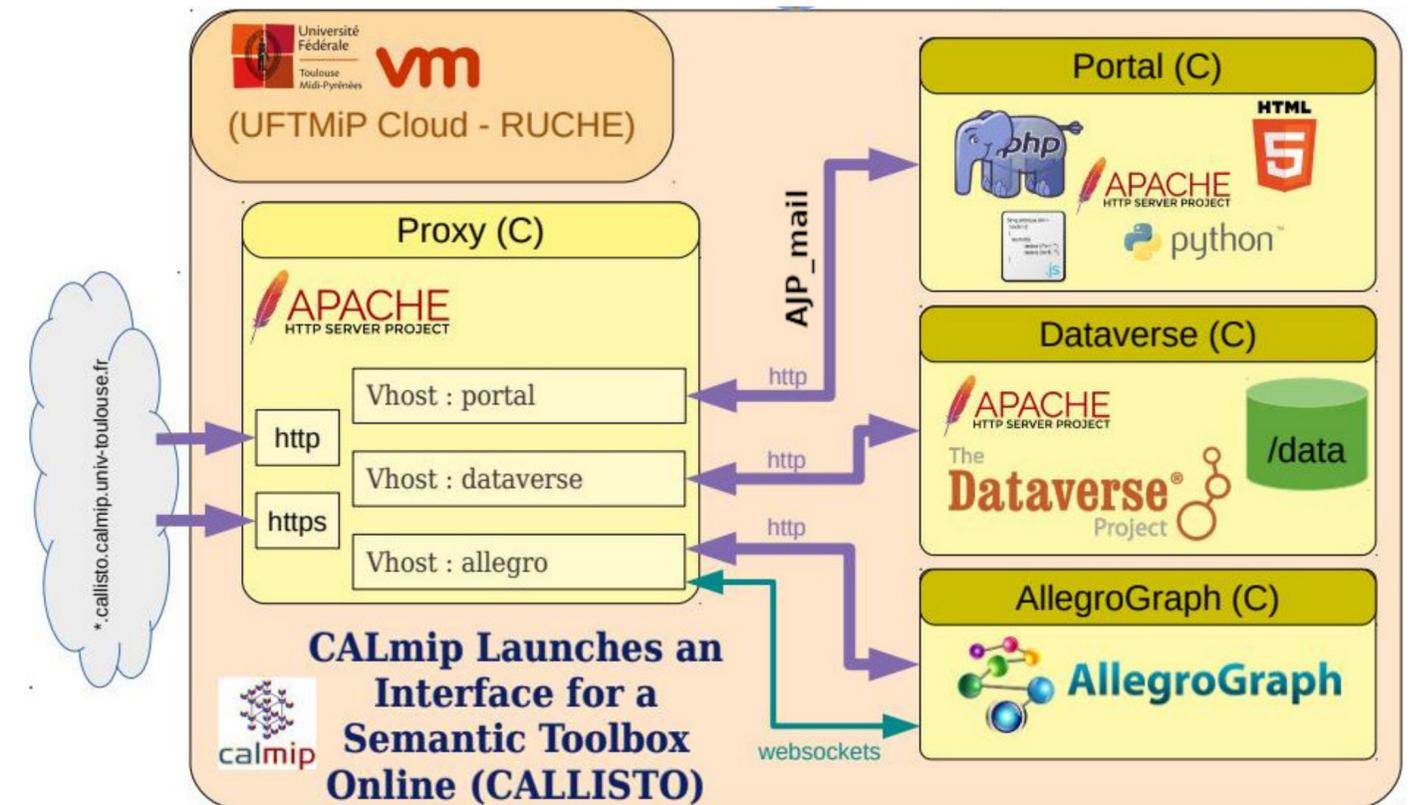


# Authentification par la Fédération d'Identités Renater

- L'authentification se fait au niveau du **reverse proxy** sur lequel est installé **shibboleth**
- Les *attributs* sont récupérés **en tant que en-tête http**
  - Paramètre **ShibUseHeaders On** dans la conf de **mod\_shibd**
- Dans le conteneur Dataverse ils sont « transformés » en variable d'environnement :
  - **SetEnvIf AJP\_mail "(.\*)"** **AJP\_mail=\$1**

On utilise le module **rewrite** pour *ne pas* authentifier certaines URL :

- L'API dataverse (qui utilise une clé)
- Les documents publics (ceux qui ont un doi)



Merci à Geoffroy Arnoud de Renater et à la liste [federation-utilisateurs@listes.renater.fr](mailto:federation-utilisateurs@listes.renater.fr)

### Rôle de Ansible

- Création des conteneurs
  - Utilisation d'une image CentOS7
  - Configuration des variables de fonctionnement (noms de domaine, d'hôte...)
  - Le rôle Ansible implémente les « bonnes pratiques » d'administration système (logwatch, unattended-upgrade etc.)
- Installation de dataverse (un rôle ansible est fourni)
- Copie de jeux de données de tests
  - Facilite la création de dépôts Dataverse de tests
  - Permet la création d'un Triple Store de tests (Allegro)

### Bénéfices

- Standardisation des développements et meilleure qualité logicielle
- Dataverse étant distribué sous forme de rôle ansible, le déploiement de Callisto et celui de Dataverse sont cohérents

### Limites

- Pour l'instant, Callisto n'est pas distribué sous forme d'« image », l'installation est un peu longue.
- Après l'installation, Callisto n'est toujours pas « Plug and Play ». Il faut des compétences en ontologies, en métadonnées pour pouvoir comprendre et utiliser l'application.
- Le passage vers Ansible demande un certain effort pour les développeurs : il faut rejouer le déploiement à chaque modification de fichier.

**C'est par le dialogue avec les utilisateurs**

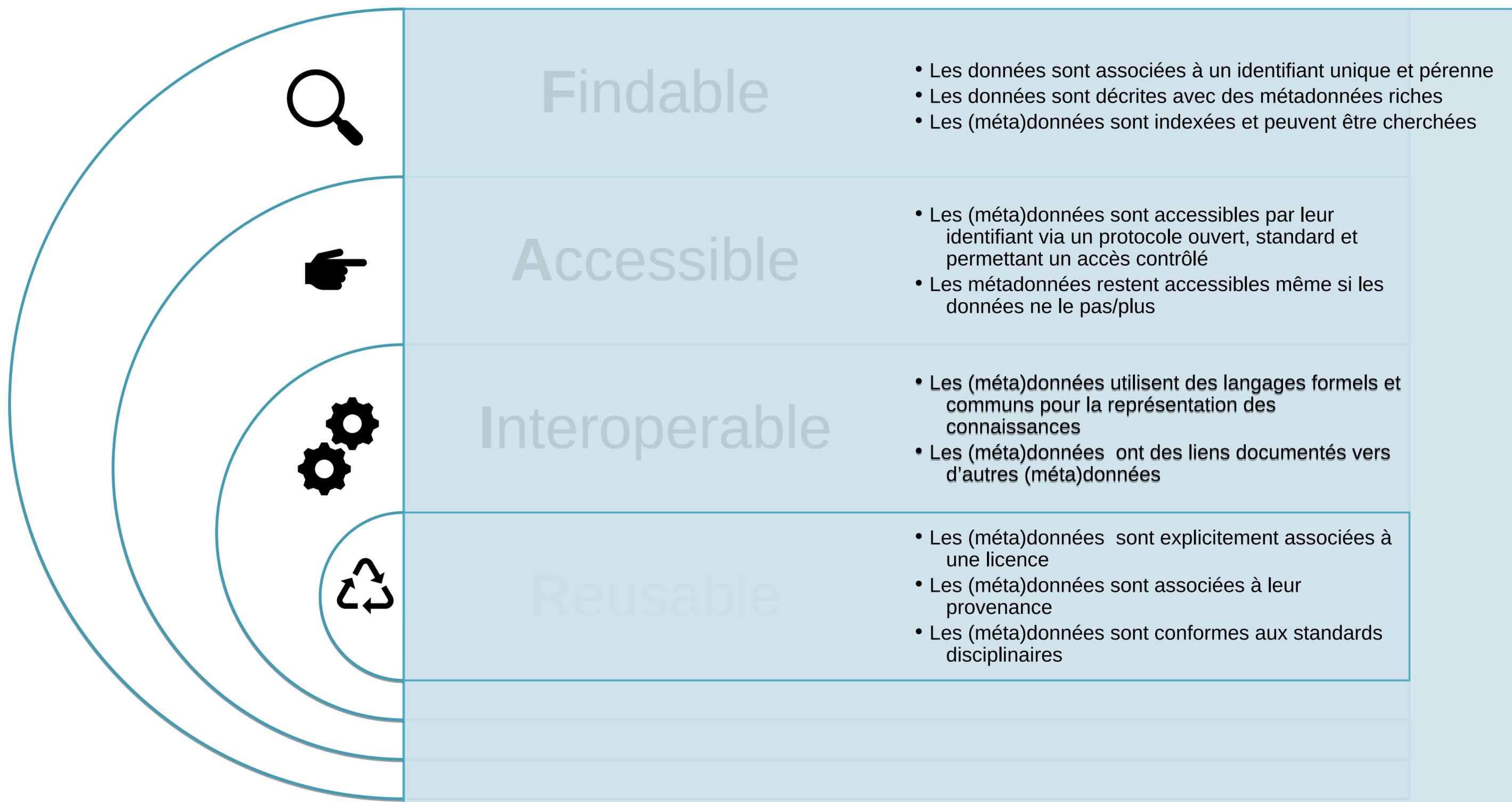
**En écoutant leurs besoins**

**Que Callisto avance**

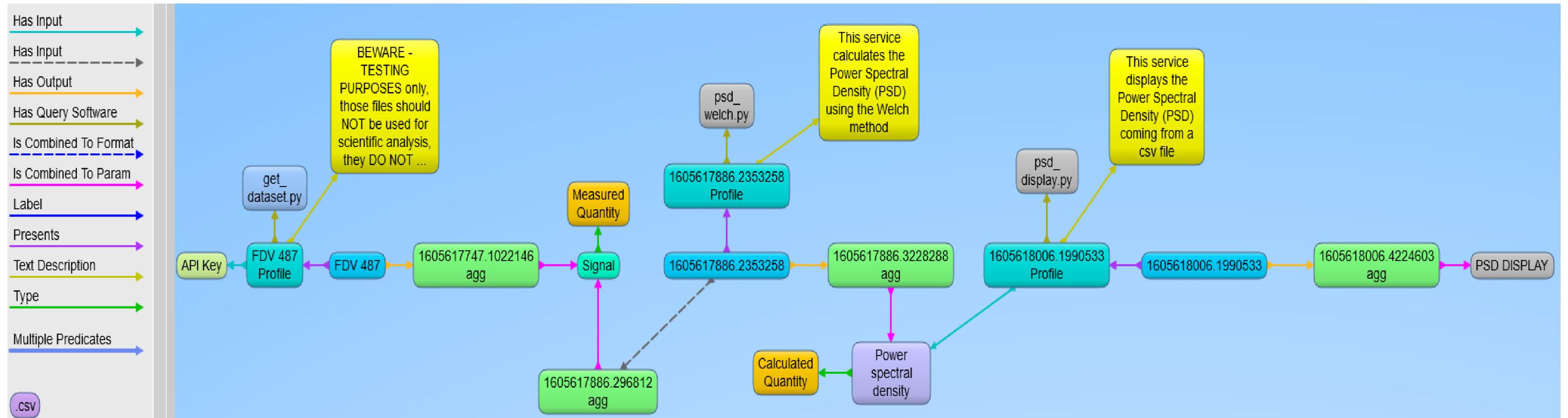
**CALLISTO**

**Merci à vous tous !**

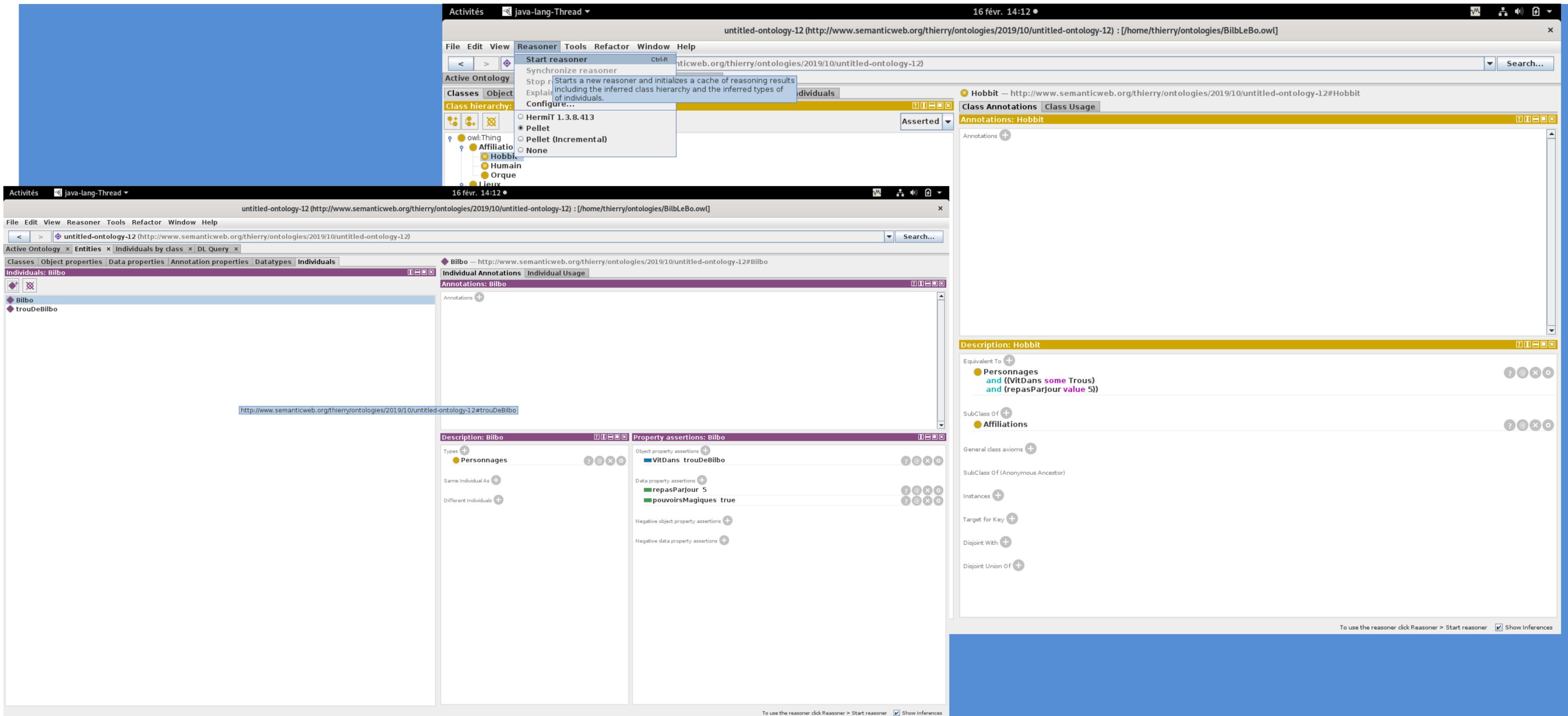




# L'ontologie dans les fonctionnalités de CALLISTO



Cette chaîne de traitements construite à la demande (et pas prédéfinie) est une automatisation rendue possible par la mise en relation d'éléments de métadonnées, de logiciels et de contexte scientifique. La plupart des éléments pourraient être rencontrés dans un PGD. C'est en quelque sort un « concentré de PGD », mais également un document complémentaire, formalisé par définition.



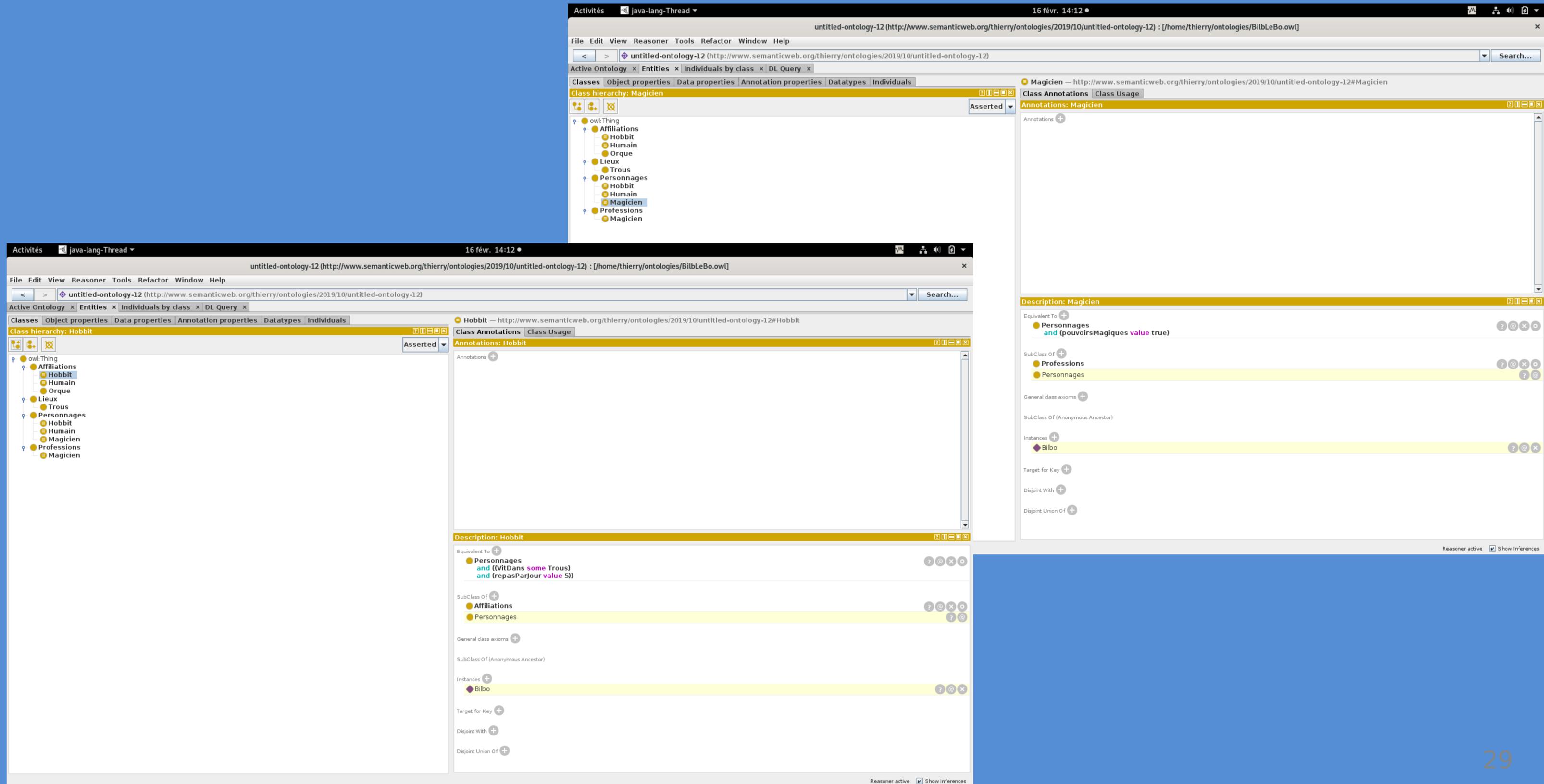
The screenshot displays the Protege ontology editor interface. The main window shows the 'untitled-ontology-12' with a menu bar including 'File', 'Edit', 'View', 'Reasoner', 'Tools', 'Refactor', 'Window', and 'Help'. The 'Reasoner' menu is open, showing options like 'Start reasoner', 'Synchronize reasoner', 'Stop reasoner', 'Explain', and 'Configure...'. The 'Start reasoner' option is highlighted, with a tooltip explaining: 'Starts a new reasoner and initializes a cache of reasoning results including the inferred class hierarchy and the inferred types of individuals.'

The interface is divided into several panes:

- Class hierarchy:** Shows a tree structure starting with 'owl:Thing', including 'Affiliation', 'Hobbit', 'Humain', 'Orque', and 'Lieux'.
- Active Ontology:** Displays the current ontology's structure.
- Individuals:** Shows a list of individuals, including 'Bilbo' and 'trouDeBilbo'.
- Description: Bilbo:** Shows the inferred class for 'Bilbo' as 'Personnages'.
- Property assertions: Bilbo:** Lists assertions for 'Bilbo', including 'VitDans trouDeBilbo', 'repasParJour 5', and 'pouvoirsMagiques true'.
- Description: Hobbit:** Shows the inferred class for 'Hobbit' as 'Personnages and ((VitDans some Trous) and (repasParJour value 5))'.
- SubClass Of:** Lists subclasses for 'Hobbit', including 'Affiliations'.

The bottom status bar indicates: 'To use the reasoner click Reasoner > Start reasoner' and 'Show Inferences' (checked).

# L'ontologie sous-jacente



The image displays two screenshots of the Protégé ontology editor interface, showing the class hierarchy and the description of two classes: Hobbit and Magicien.

**Top Screenshot (Magicien):**

- Class hierarchy:** owl:Thing -> Affiliations -> Hobbit, Humain, Orque; Lieux -> Trou; Personnages -> Hobbit, Humain, Magicien; Professions -> Magicien.
- Description: Magicien:**
  - Equivalent To: Personnages and (pouvoirsMagiques value true)
  - SubClass Of: Professions, Personnages
  - Instances: Bilbo

**Bottom Screenshot (Hobbit):**

- Class hierarchy:** owl:Thing -> Affiliations -> Hobbit, Humain, Orque; Lieux -> Trou; Personnages -> Hobbit, Humain, Magicien; Professions -> Magicien.
- Description: Hobbit:**
  - Equivalent To: Personnages and ((VitDans some Trou) and (repasParJour value 5))
  - SubClass Of: Affiliations, Personnages
  - Instances: Bilbo