

CALLISTO

*Un portail sémantique pour le partage de données
scientifiques*

CAPITOUL, 6 mai 2021



CALMIP (UAR 366)
Espace Clément Ad
www.calmip.univ-toulouse

Un portail sémantique pour le partage de données scientifiques

1. Portail CALLISTO

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

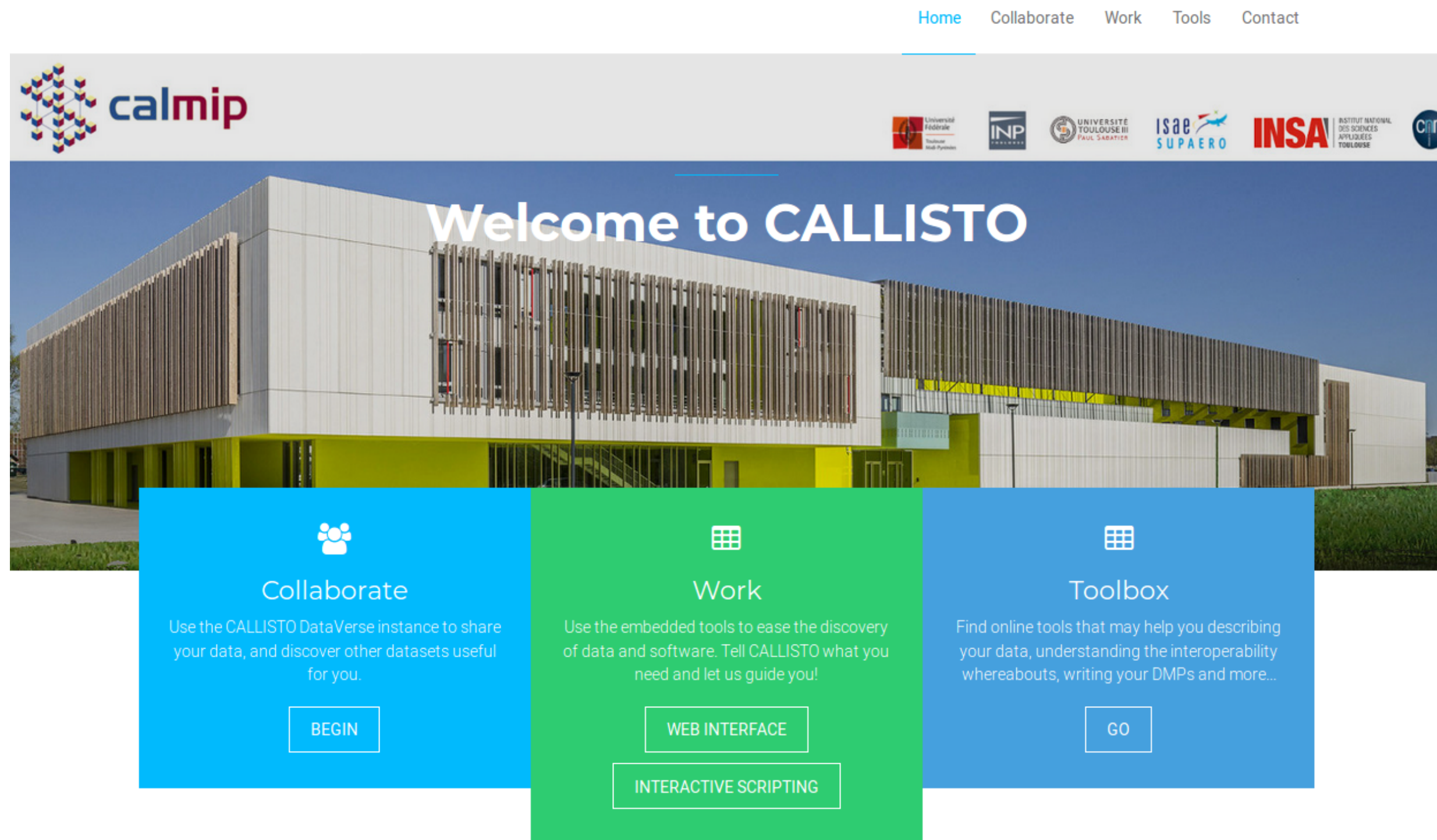
2. Architecture en conteneurs

- 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>

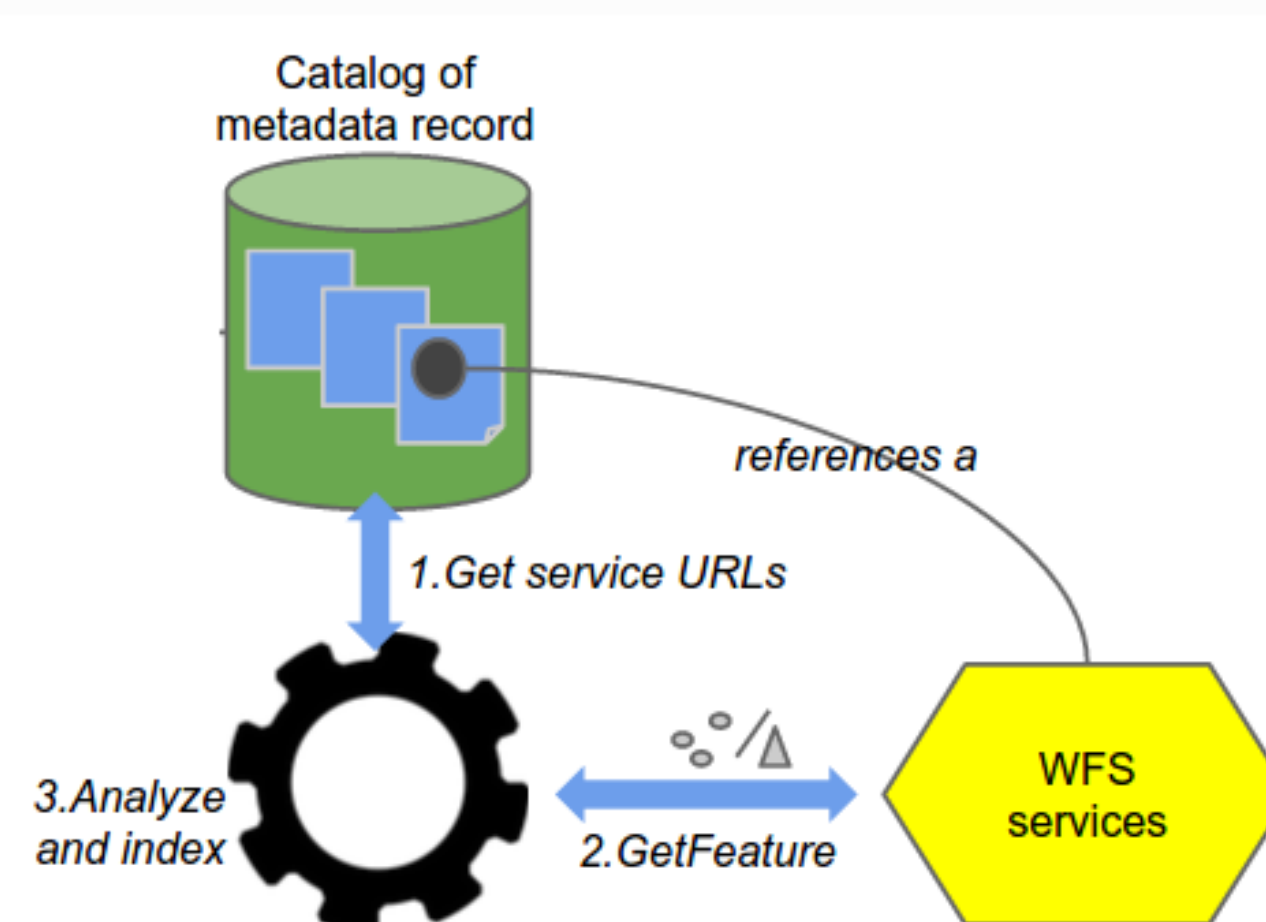
Contexte général : métadonnées

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

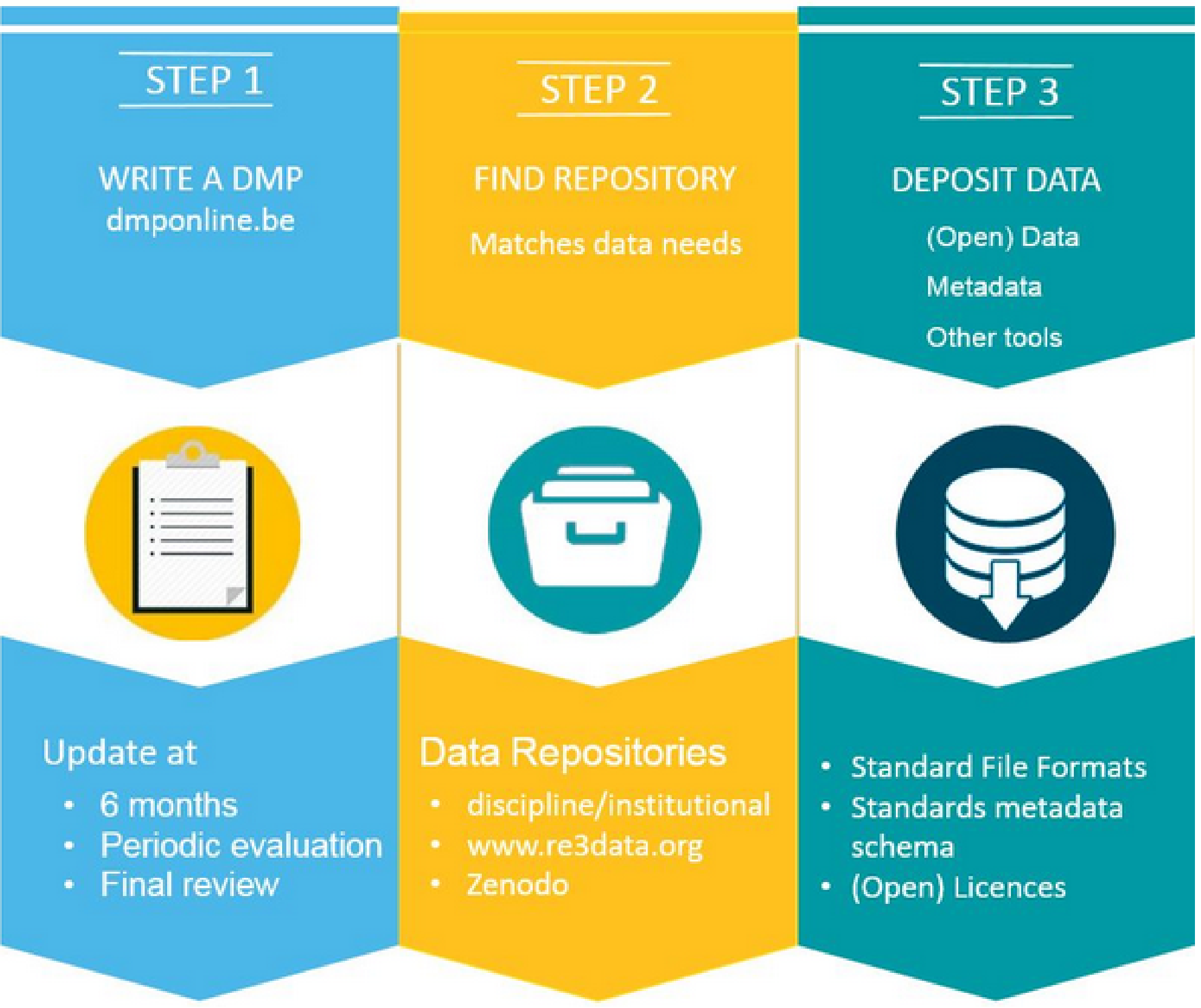
Data accessible using download services (WFS)

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.

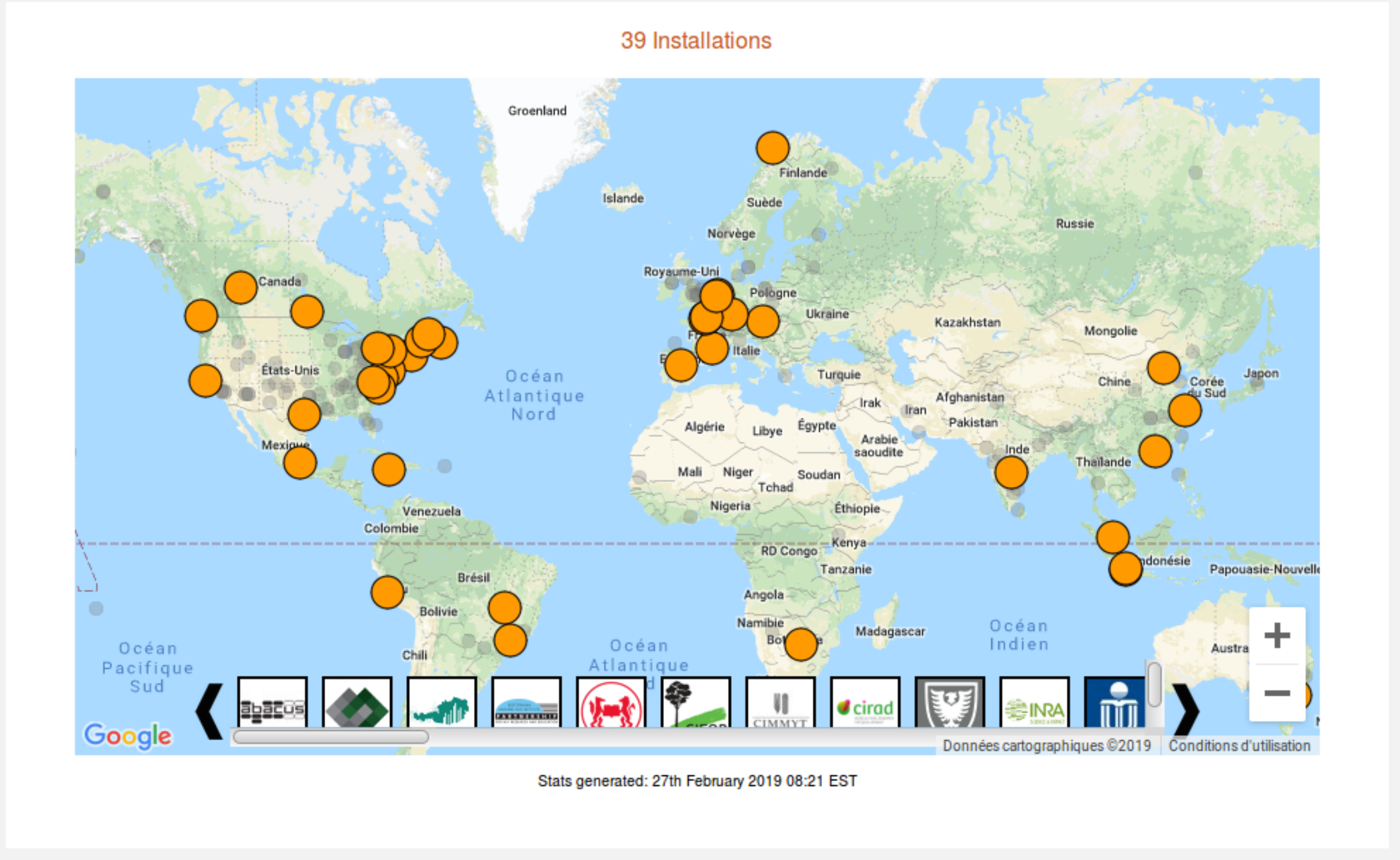


<https://geonetwork-opensource.org/manuals/trunk/fr/user-guide/analyzing/data.html>

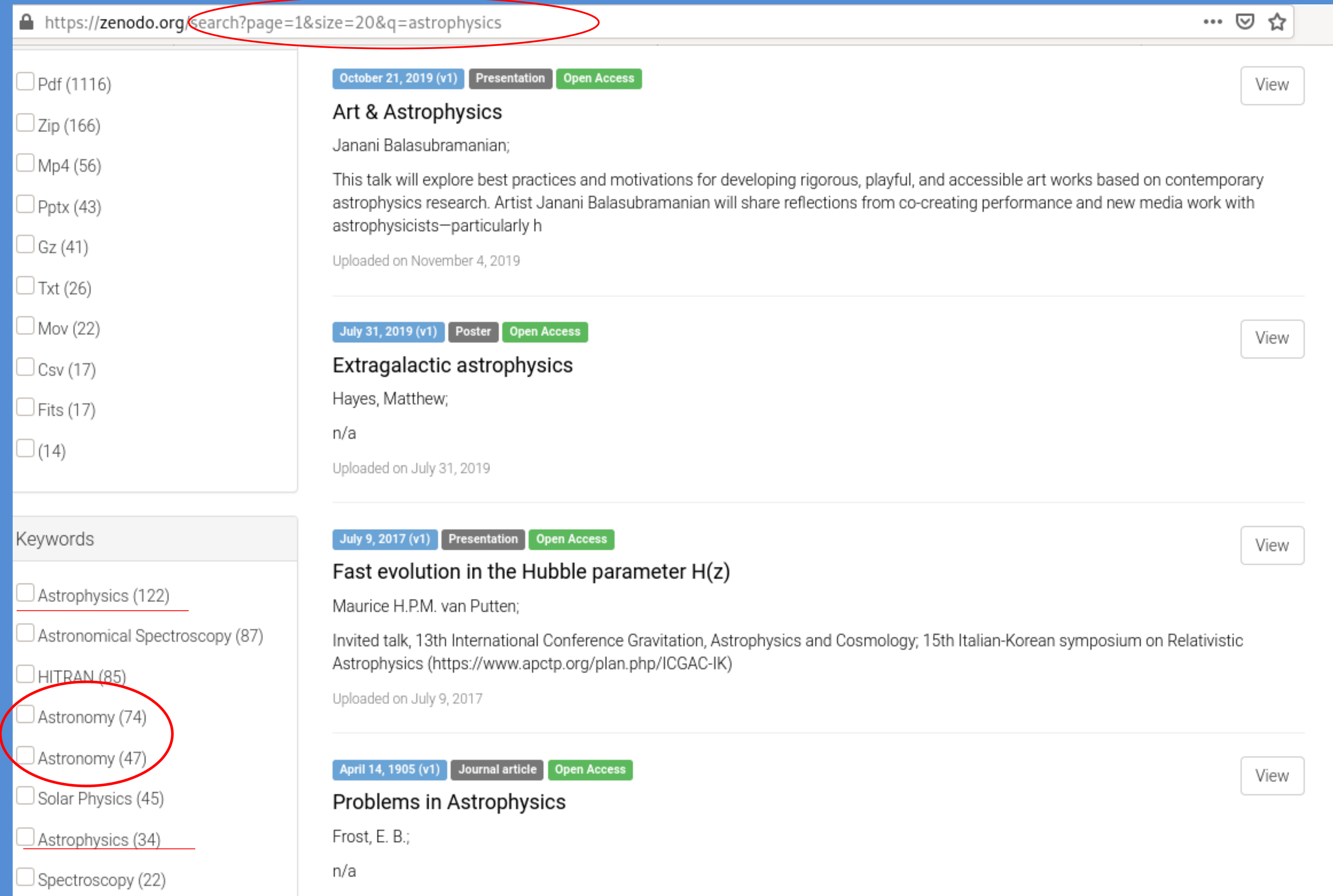
Source: OpenAIRE <https://www.openaire.eu>



DATaverse REPOSITORIES - A WORLD VIEW



Positionnement par rapport aux portails existants



Fonctionnalités de CALLISTO

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.

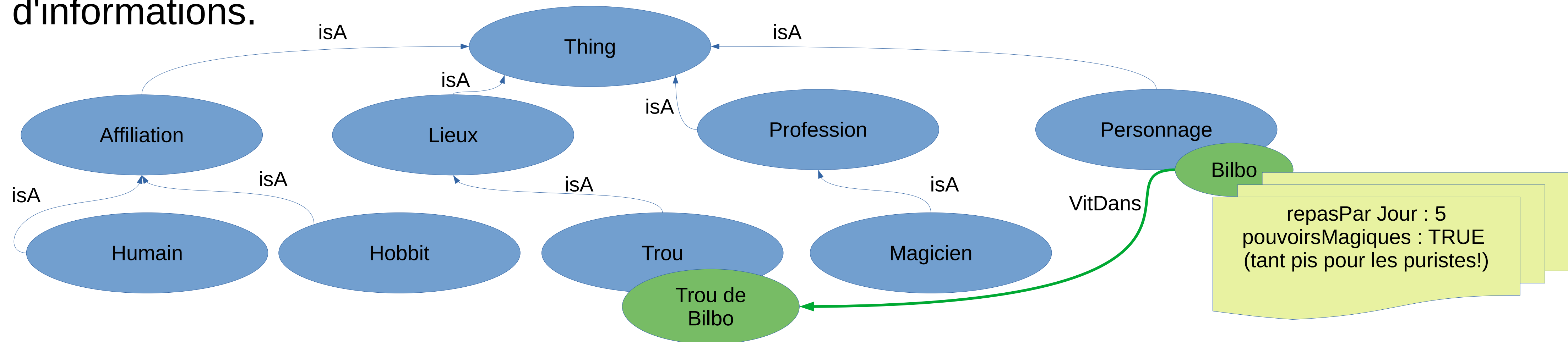
- Besoins exprimés : Espace de partage
- Travail effectué : ouverture d'un espace de dépôt soutenu par des métadonnées reliées à une ontologie spécifique construite par ISMS et CALMIP
- Travail en cours : amélioration de l'ontologie, conception d'outils de recherche de données avancés.

➤ 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.



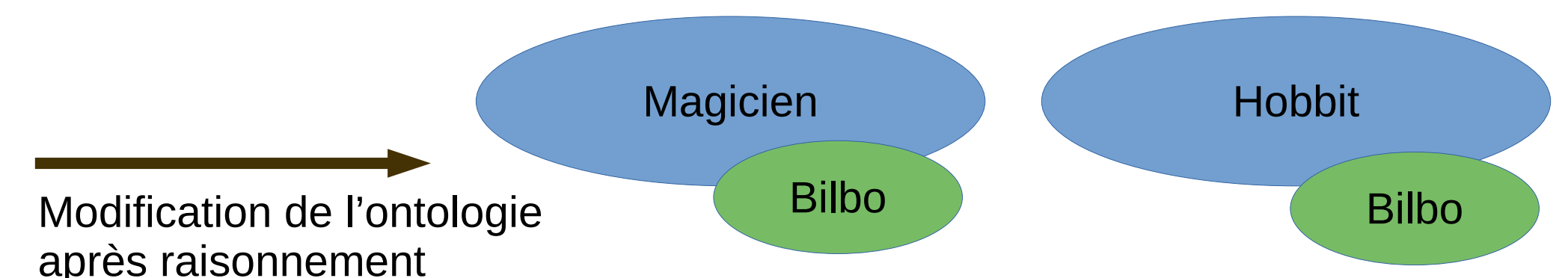
L'ontologie constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que des relations entre ces concepts. Elle est employée pour raisonner à propos des objets du domaine concerné.

« *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 - alpha=2.5 deg - f=350 Hz", <https://doi.org/10.5072/FK2/S4UL3Z>, Root, DRAFT VERSION 

 Cite Dataset ▾

 Learn about Data Citation Standards.

Description

Fichier au format ASCII. Forces aérodynamiques intégrées autour d'une aile d'A320 de corde $c = 0.15$ m, d'incidence $\alpha = 2.5^\circ$. 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 = $\pm 1^\circ$ 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 = C_x colonne 3 = C_z colonne 4 = C_m

Subject

Physics; Other

Keyword

T1 morphing, NSMB, C_x , C_z , C_m .plt

Files

Metadata

Terms

Versions

Dataverse natif

 Add + Edit Metadata

Métadonnées générales

Citation Metadata 

Dataset Persistent ID

doi:10.5072/FK2/S4UL3Z

Title

forces aérodynamiques morphing 2D - alpha=2.5 deg - 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 $\alpha = 2.5^\circ$. 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 = $\pm 1^\circ$ 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 = C_x colonne 3 = C_z colonne 4 = C_m

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/>
 C_x , C_z , C_m .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

None

ssor relaxation

None

KIS

None

unsteady

None

pointwise implicit

Select...

flapping distance

None

cmv

None

limiter variable flag

None

mesh deformation flag

None

time scheme

Select...

wall distance flag

Select...

alpha

None

chordofairfoil

None

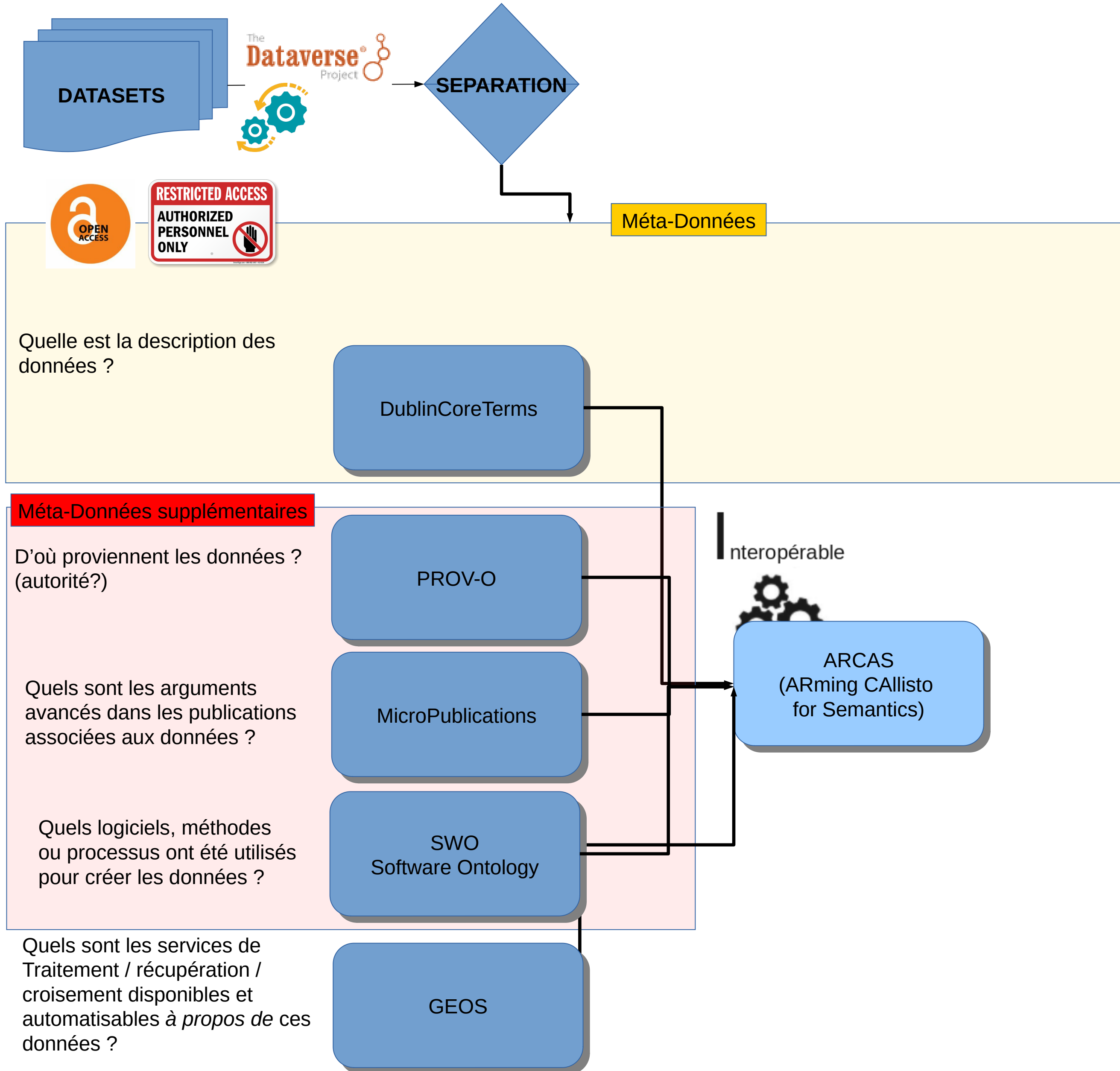
updblock implicit

None

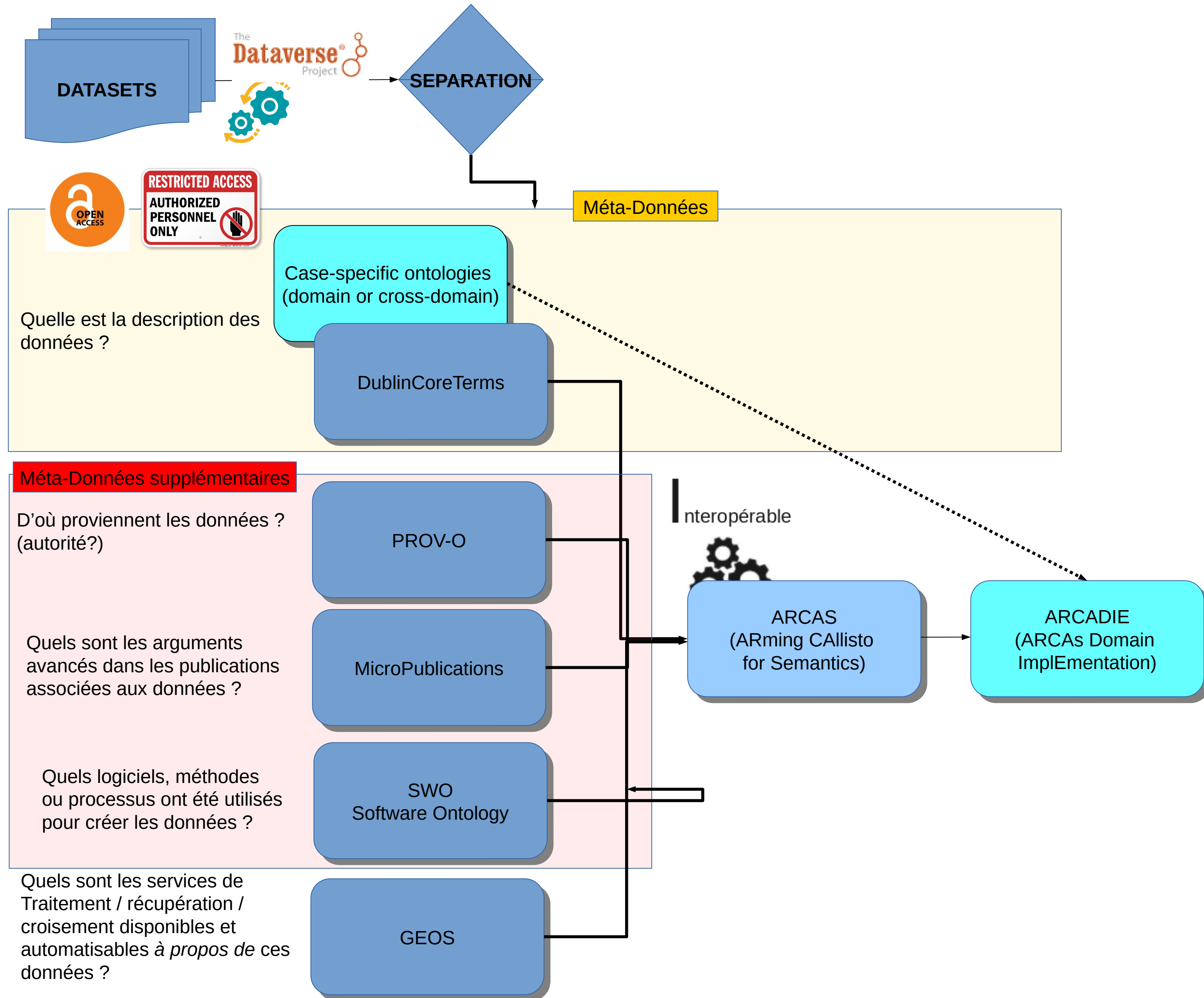
pressure infinity

None

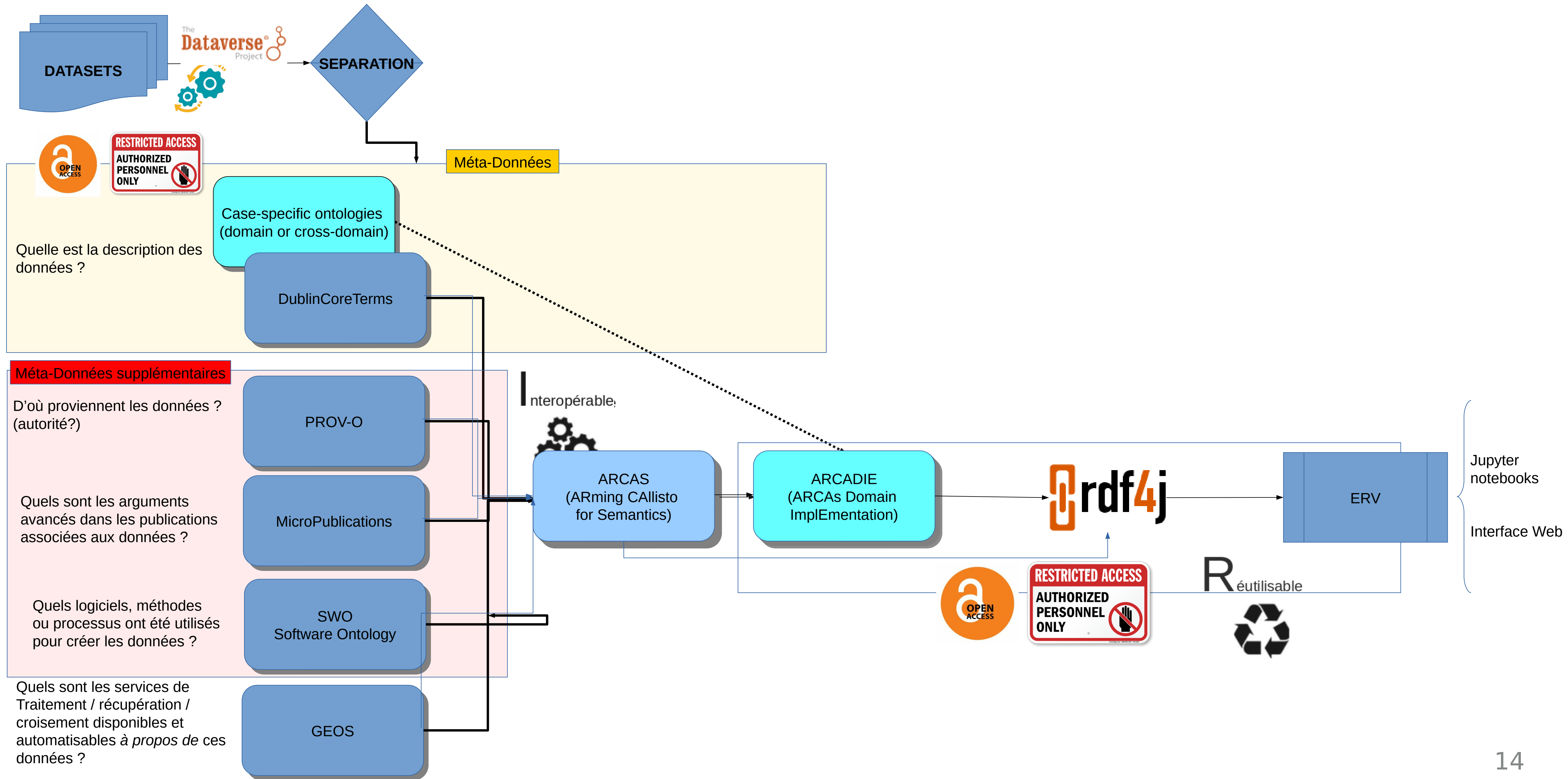
L'ontologie sous-jacente



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

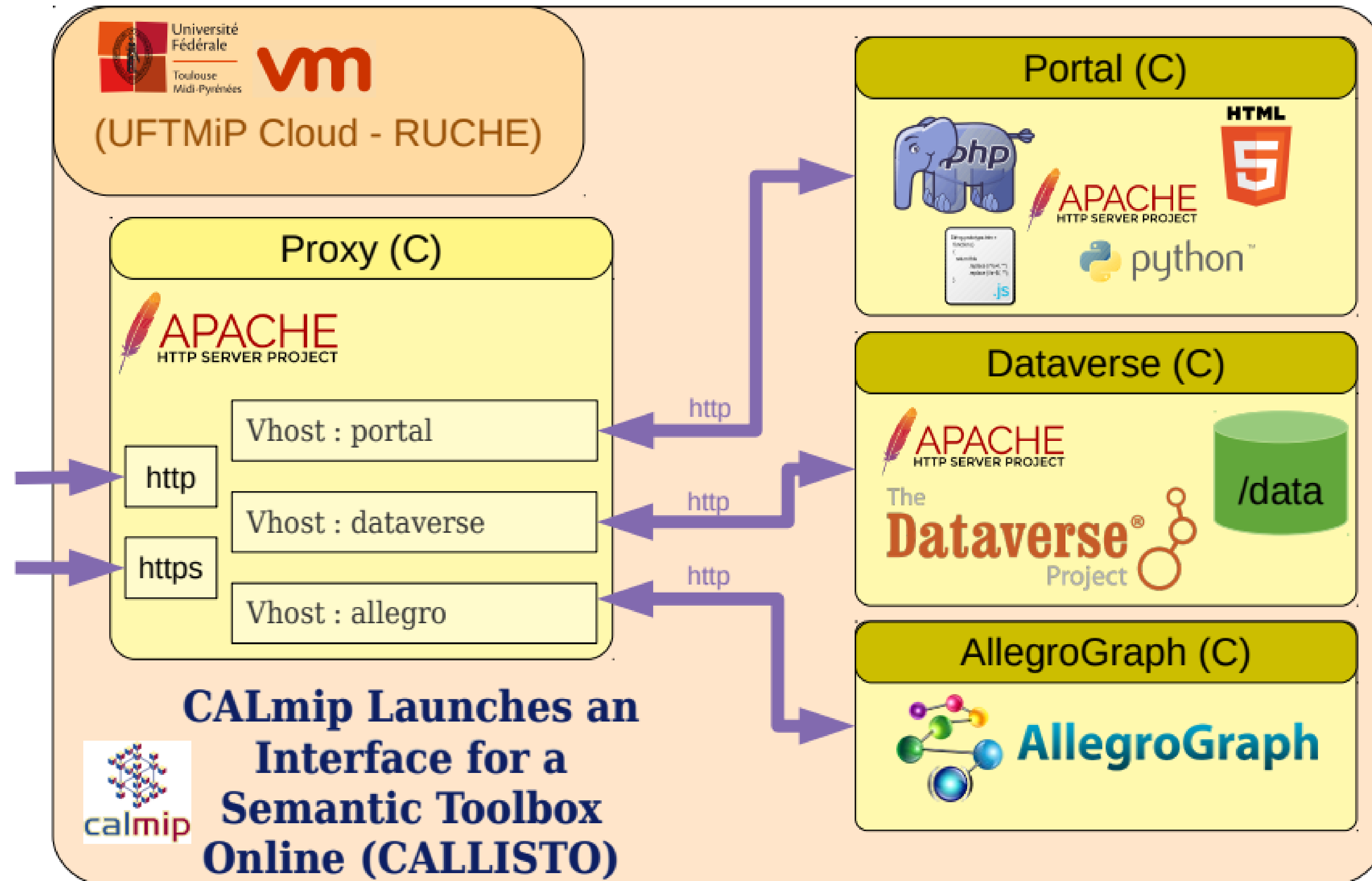
- **Shibboleth**

- **Diffusion avec git et ansible**

- Rôle, bénéfices et limites.

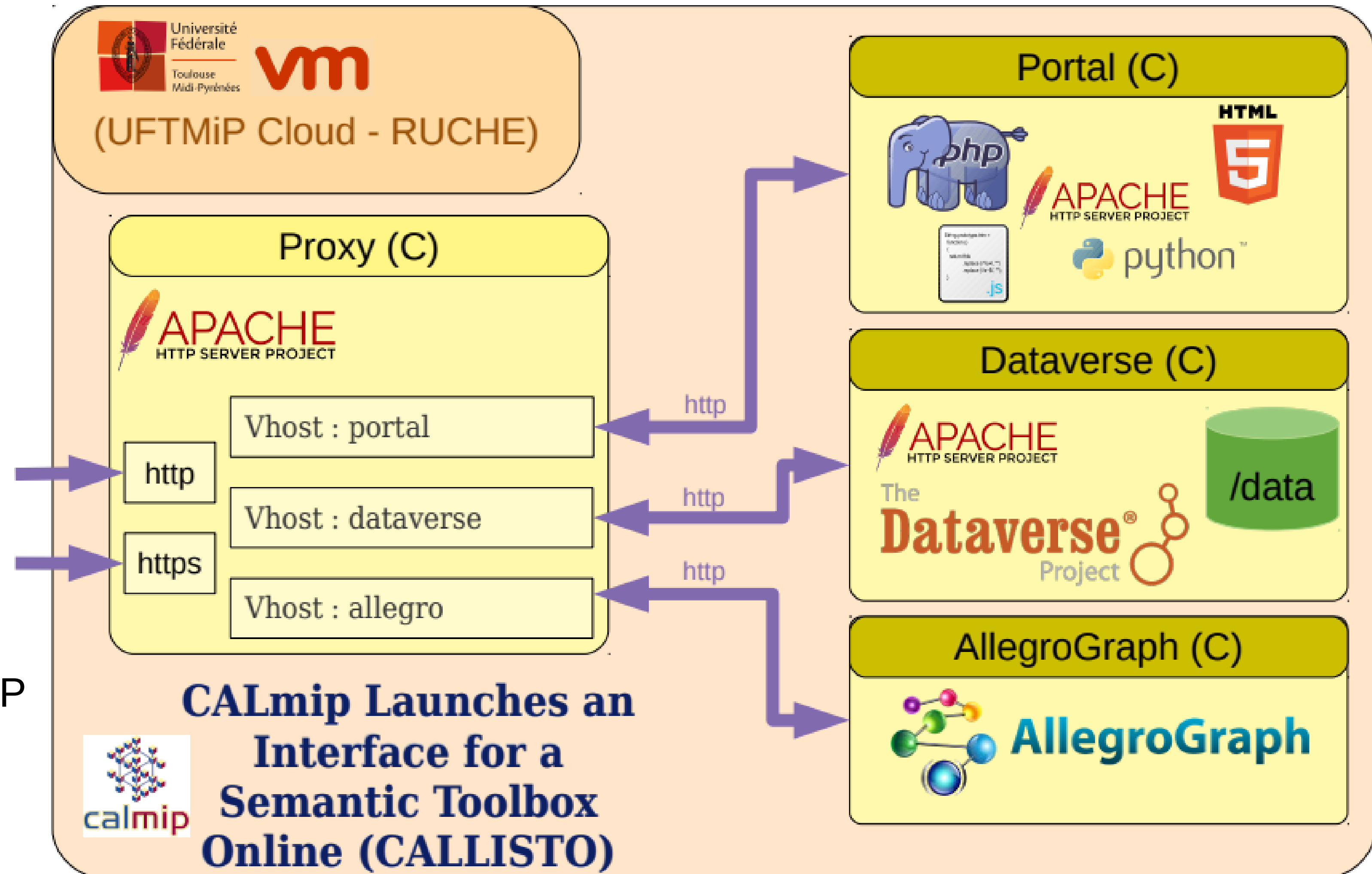
•Architecture

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



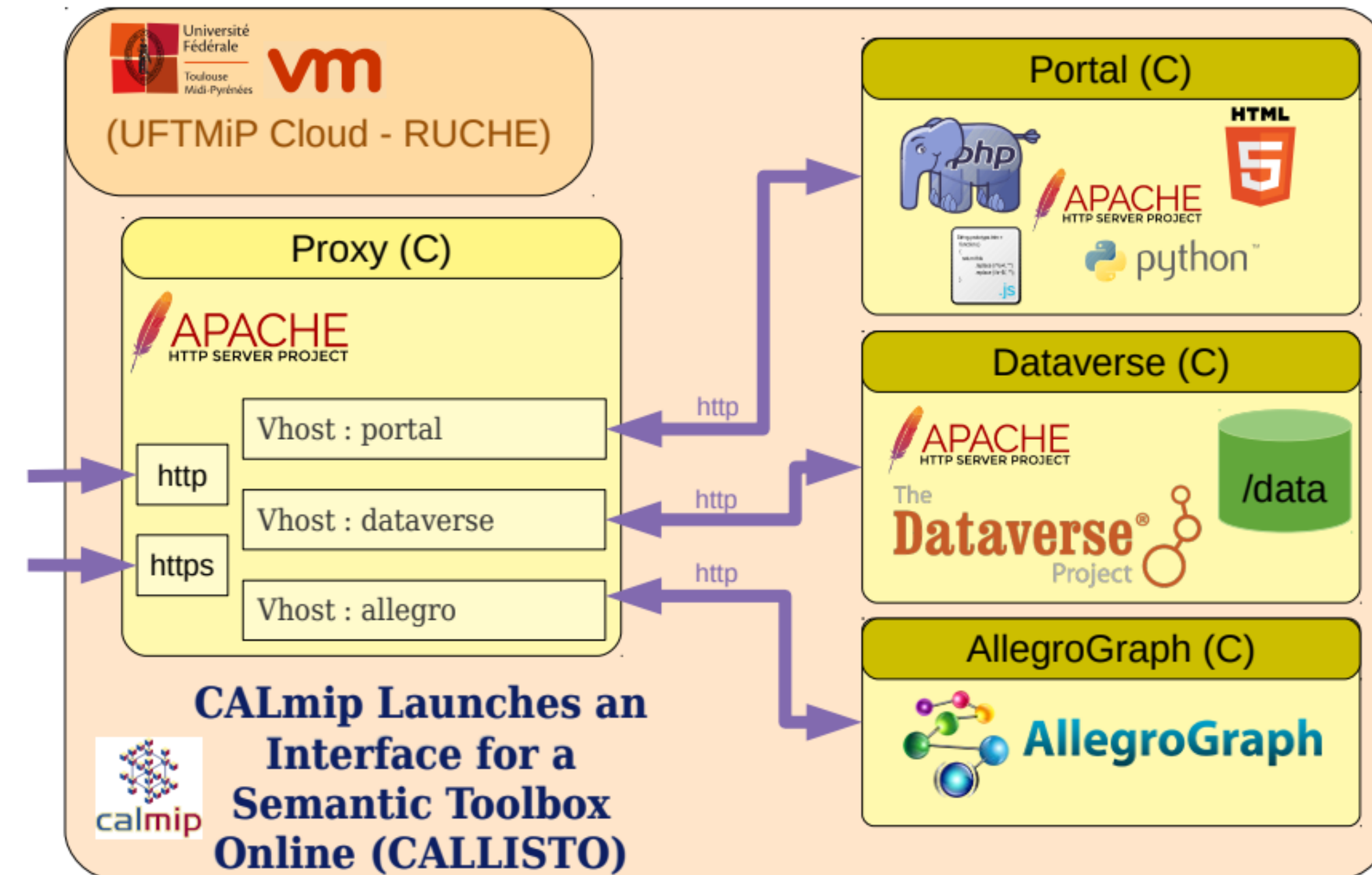
Motivations

- **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

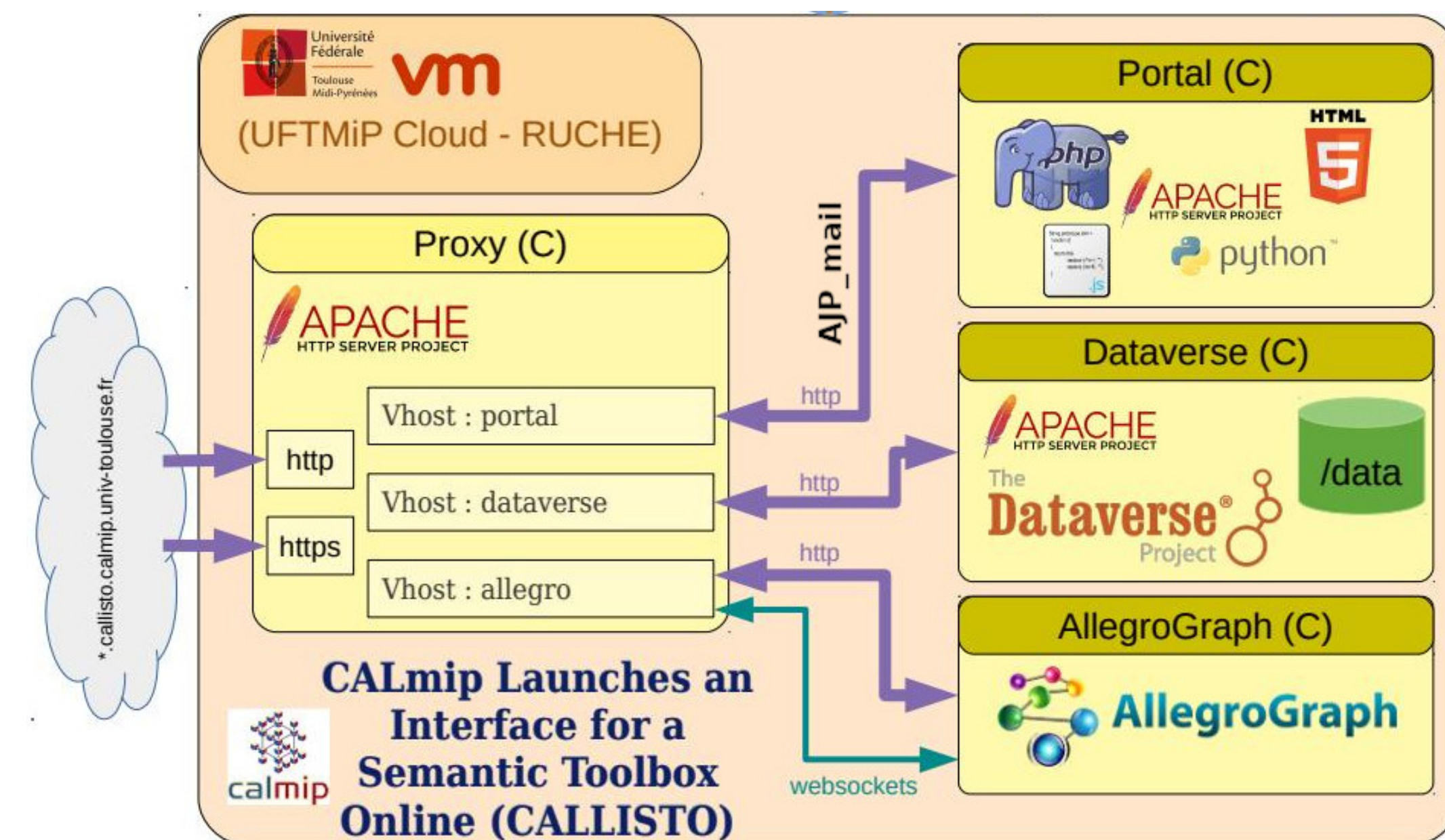


Authentification par la Fédération d'Identités Renater

- 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

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

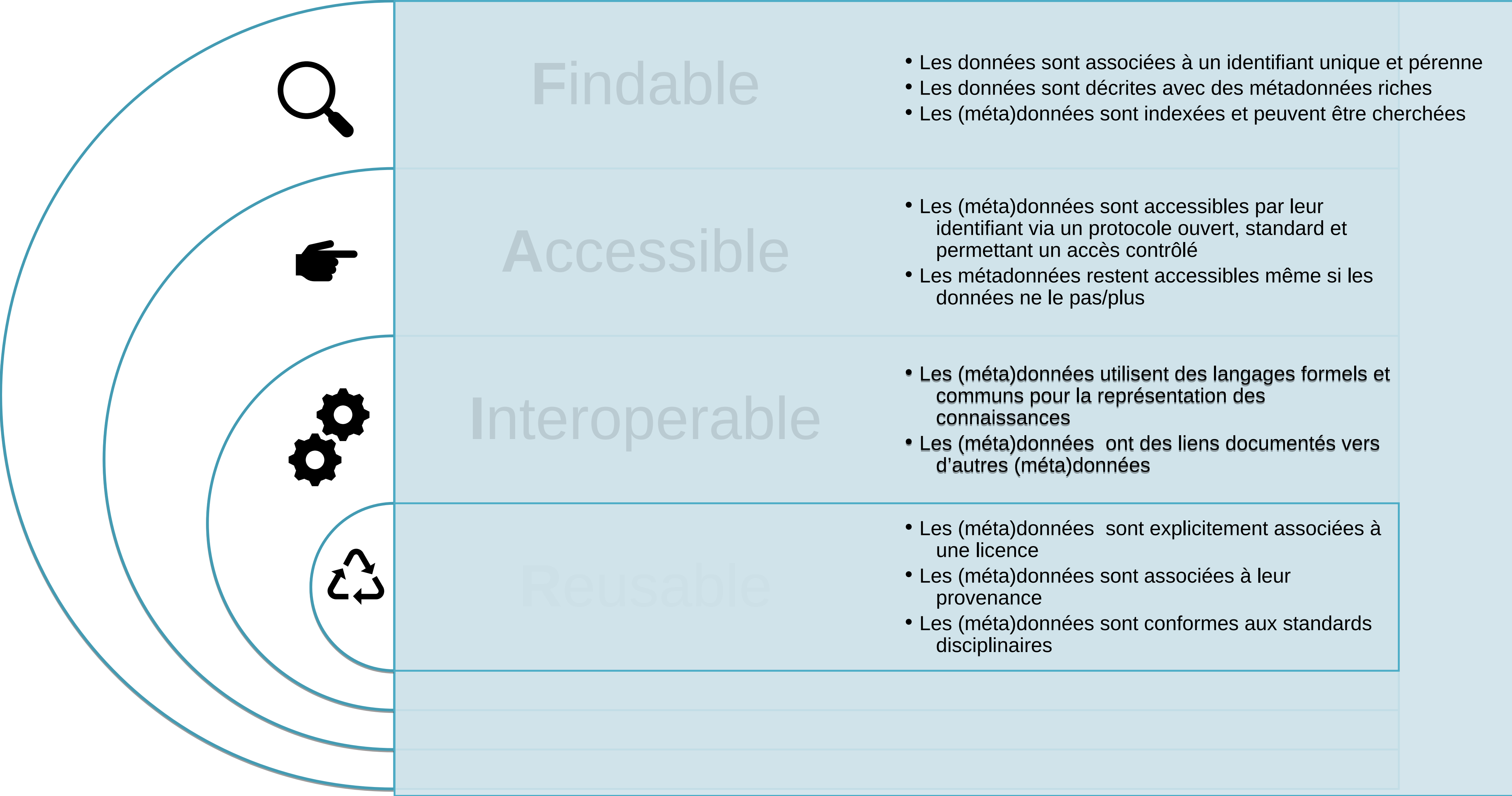
En écoutant leurs besoins

Que Callisto avance

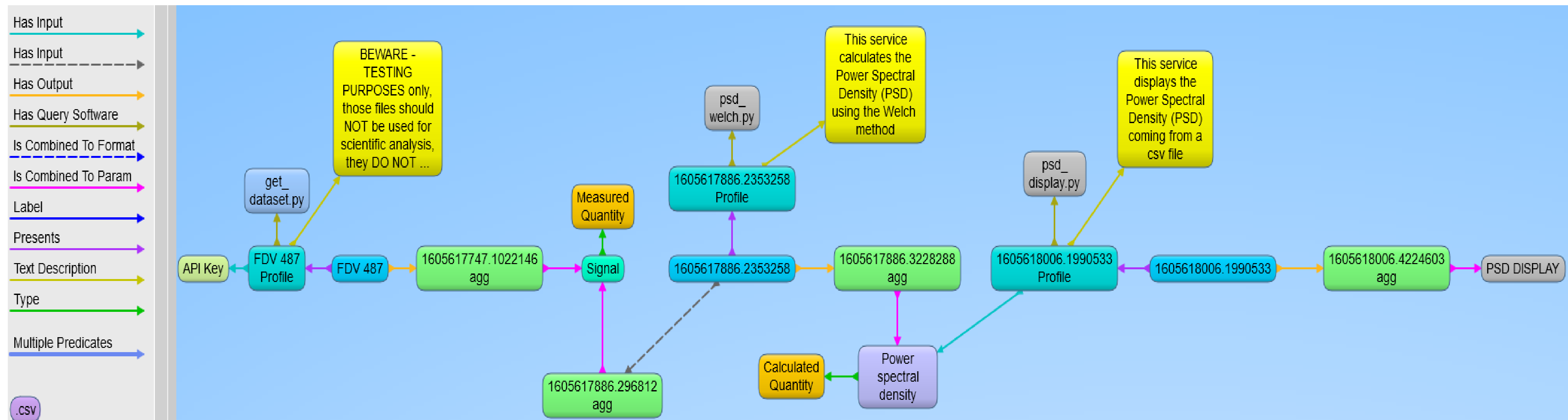
CALLISTO

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.

Activités

java-lang-Thread

16 févr. 14:12

untitled-ontology-12 (http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12) : [/home/thierry/ontologies/BilbLeBo.owl]

File Edit View Reasoner Tools Refactor Window Help

< > 🔍

untitled-ontology-12 (http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12)

Search...

Active Ontology

Classes Object

Class hierarchy:

owl:Thing

Affiliation

Hobbit

Humain

Orque

Lieux

Start reasoner

Synchronize reasoner

Stop reasoner

Explain

Configure...

Starts a new reasoner and initializes a cache of reasoning results including the inferred class hierarchy and the inferred types of individuals.

Hermit 1.3.8.413

Pellet

Pellet (Incremental)

None

Class Annotations

Class Usage

Annotations: Hobbit

Annotations

Description: Hobbit

Equivalent To

Personnages and ((VitDans some Trous) and (repasParJour value 5))

SubClass Of

Affiliations

General class axioms

SubClass Of (Anonymous Ancestor)

Instances

Target for Key

Disjoint With

Disjoint Union Of

untitled-ontology-12 (http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12) : [/home/thierry/ontologies/BilbLeBo.owl]

File Edit View Reasoner Tools Refactor Window Help

< > 🔍

untitled-ontology-12 (http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12)

Search...

Active Ontology x Entities x Individuals by class x DL Query x

Classes Object properties Data properties Annotation properties Datatypes Individuals

Individuals: Bilbo

Bilbo

trouDeBilbo

Individual Annotations

Individual Usage

Annotations: Bilbo

Annotations

Description: Bilbo

Types

Personnages

Same Individual As

Different Individuals

Property assertions: Bilbo

Object property assertions

VitDans trouDeBilbo

Data property assertions

repasParJour 5

pouvoirsMagiques true

Negative object property assertions

Negative data property assertions

http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12#trouDeBilbo

To use the reasoner click Reasoner > Start reasoner

Show Inferences

L'ontologie sous-jacente

Activités

java-lang-Thread

16 févr. 14:12

untitled-ontology-12 (http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12) : [/home/thierry/ontologies/BilbLeBo.owl]

File Edit View Reasoner Tools Refactor Window Help

untitled-ontology-12 (http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12)

Active Ontology x Entities x Individuals by class x DL Query x

Classes Object properties Data properties Annotation properties Datatypes Individuals

Class hierarchy: Magicien

owl:Thing

- Affiliations
 - Hobbit
 - Humain
 - Orque
- Lieux
 - Trous
- Personnages
 - Hobbit
 - Humain
 - Magicien
- Professions
 - Magicien

Magicien — http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12#Magicien

Class Annotations Class Usage

Annotations: Magicien

Description: Magicien

Equivalent To

- Personnages and (pouvoirsMagiques value true)

SubClass Of

- Professions
- Personnages

General class axioms

SubClass Of (Anonymous Ancestor)

Instances

- Bilbo

Target for Key

Disjoint With

Disjoint Union Of

Reasoner active ☒ Show Inferences

Activités

java-lang-Thread

16 févr. 14:12

untitled-ontology-12 (http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12) : [/home/thierry/ontologies/BilbLeBo.owl]

File Edit View Reasoner Tools Refactor Window Help

untitled-ontology-12 (http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12)

Active Ontology x Entities x Individuals by class x DL Query x

Classes Object properties Data properties Annotation properties Datatypes Individuals

Class hierarchy: Hobbit

owl:Thing

- Affiliations
 - Hobbit
 - Humain
 - Orque
- Lieux
 - Trous
- Personnages
 - Hobbit
 - Humain
 - Magicien
- Professions
 - Magicien

Hobbit — http://www.semanticweb.org/thierry/ontologies/2019/10/untitled-ontology-12#Hobbit

Class Annotations Class Usage

Annotations: Hobbit

Description: Hobbit

Equivalent To

- Personnages and ((VitDans some Trous) and (repasParJour value 5))

SubClass Of

- Affiliations
- Personnages

General class axioms

SubClass Of (Anonymous Ancestor)

Instances

- Bilbo

Target for Key

Disjoint With

Disjoint Union Of

Reasoner active ☒ Show Inferences