Semantic artefact and ontology services for long-term data interpretation

PV 2023 Conference
CERN, Switzerland

Clement Jonquet – jonquet@lirmm.fr

May 3rd 2023

Mistea
Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie

INRAE
la science pour la vie, l'humain, la terre

LIRMM
Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier

Université Montpellier

What would happen if we would be able to preserve data for so long but we would have lost the ways to interpret the knowledge contained in these data?

The tape-unit reel-display system (RDS) shown mounted over tape units in the 6600 computing complex, in 1965.
Image credit: CERN.

Decouple the “model” from the data themselves

- Preserve the data and save this model for future interpretation.

- Separation data/model
  by annotating or describing
  data with semantic artefacts

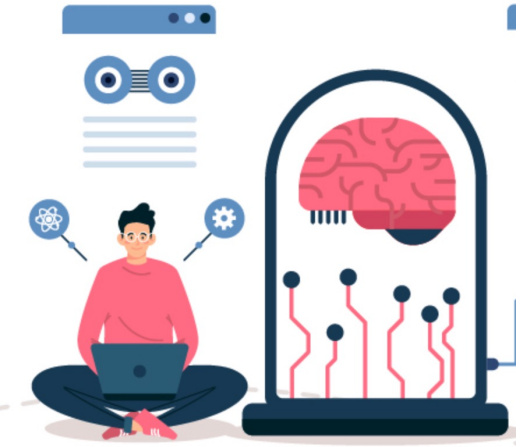
- Semantic artefacts: ontologies,
  terminologies, taxonomies,
  thesauri, vocabularies,
  metadata schemas and standards

The talks fits in 2 statements

#1

We need to develop services to adopt and facilitate semantic artefacts use when archiving data

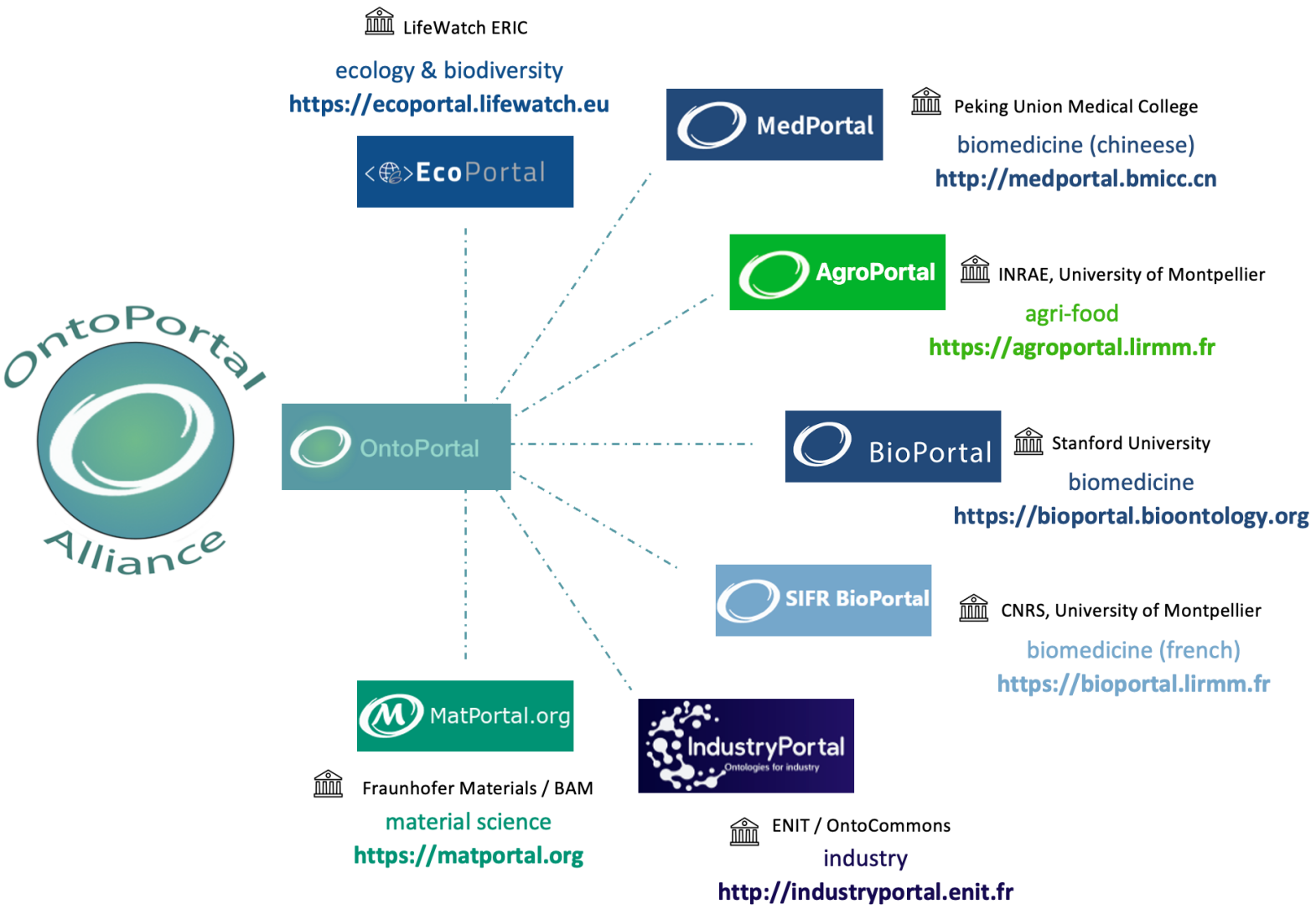
(e.g., semantic annotation, indexing, knowledge graphs, linked data)



#2

Semantic artefacts are themselves data so we need to develop long term repositories for archiving semantic artefacts

(e.g., gather them, save them, deals with format heterogeneity, versioning…)



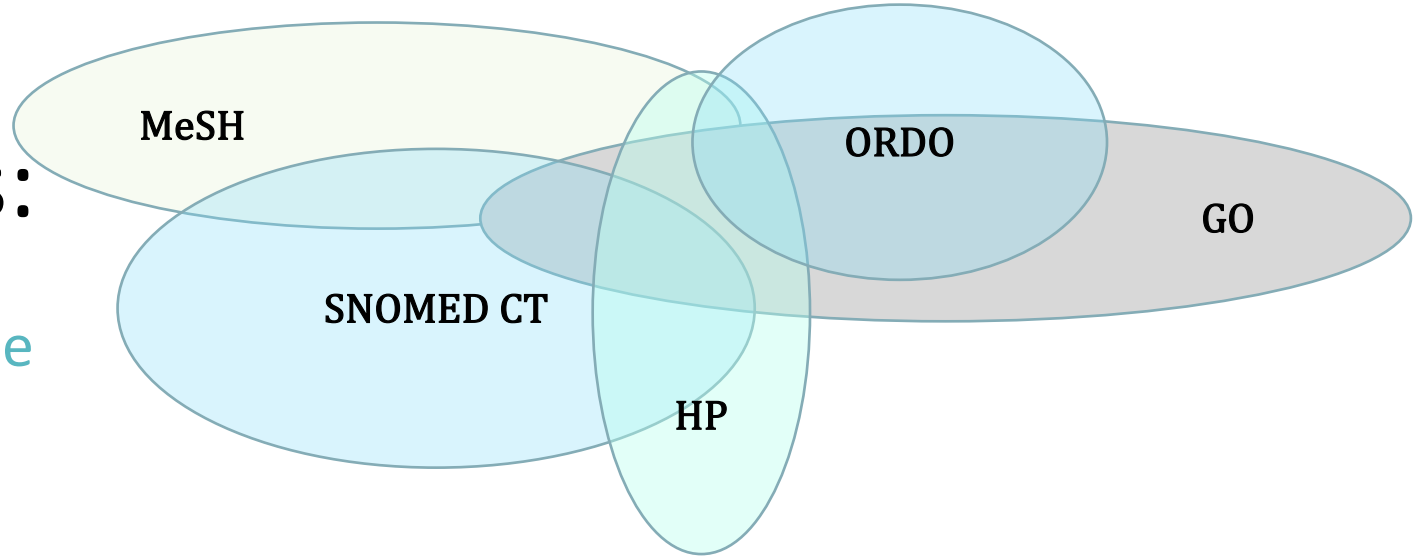
We develop an open-source semantic artefact and ontology repository technology called OntoPortal

https://ontoportal.org

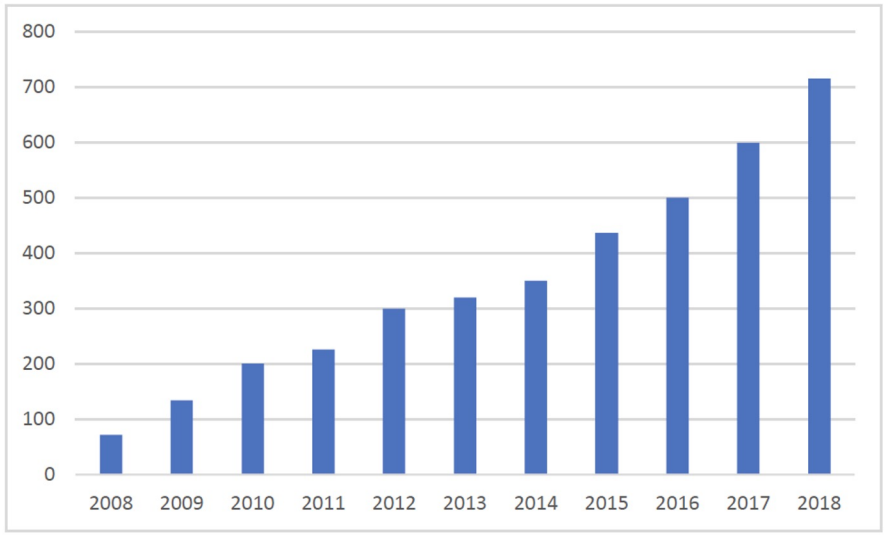
A few elements on ontology repositories

Issues with ontologies:

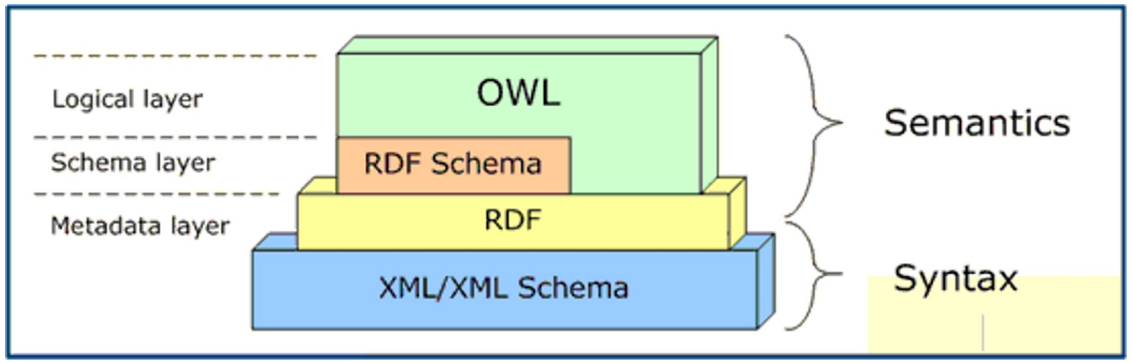
- spread out,
- in different formats, of different size
- with different structures
- increasing number
- overlapping



Overlapping ontologies



Number of ontologies in the NCBO BioPortal



Variety of representation languages

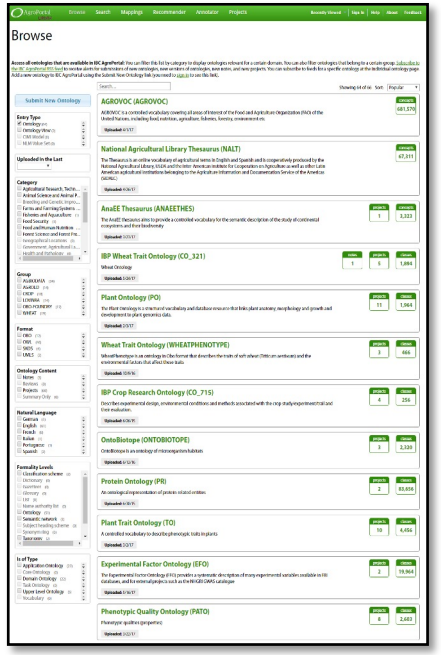
Why ontology repositories are important?

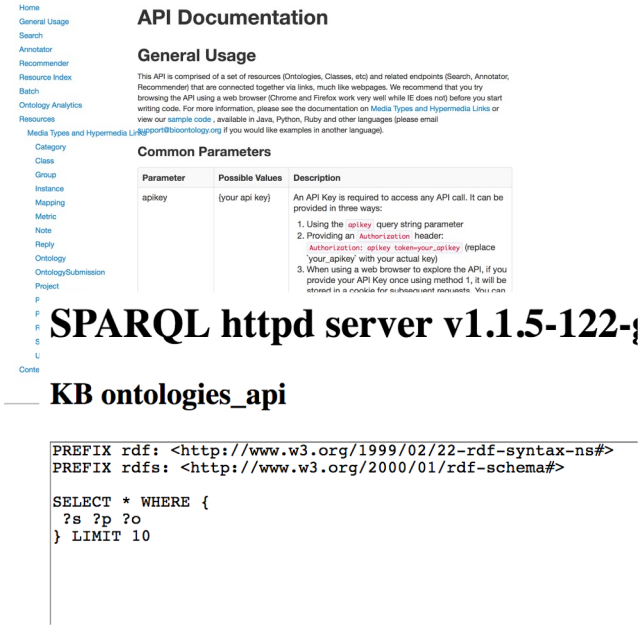
- You've built an ontology, how do you let the world know?
- You need an ontology, where do you go to get it?
- How do you know whether an ontology is any good?

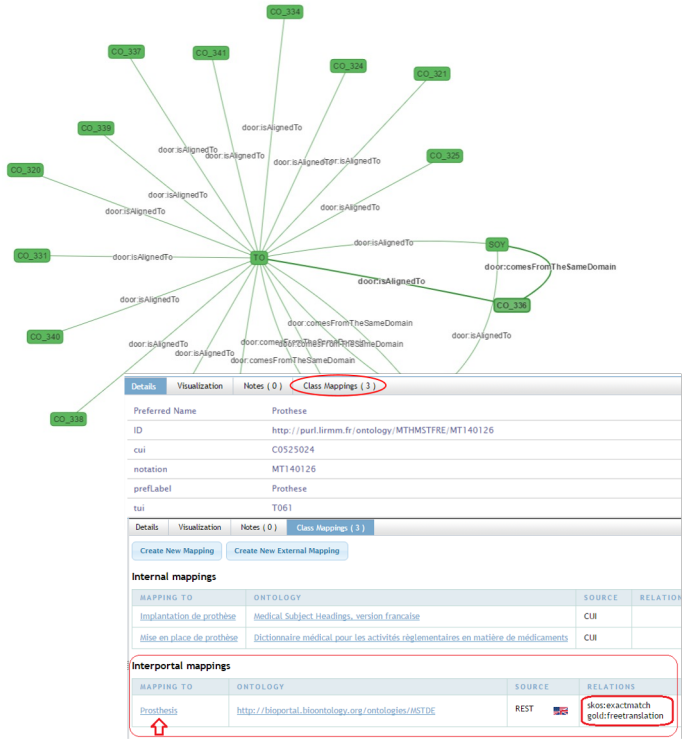
- How do you find data resources that are relevant to the domain of the ontology?
- How could you leverage your ontology to enable new science?
- How could you use ontologies without managing them ?

Ontology repositories help to make ontologies FAIR

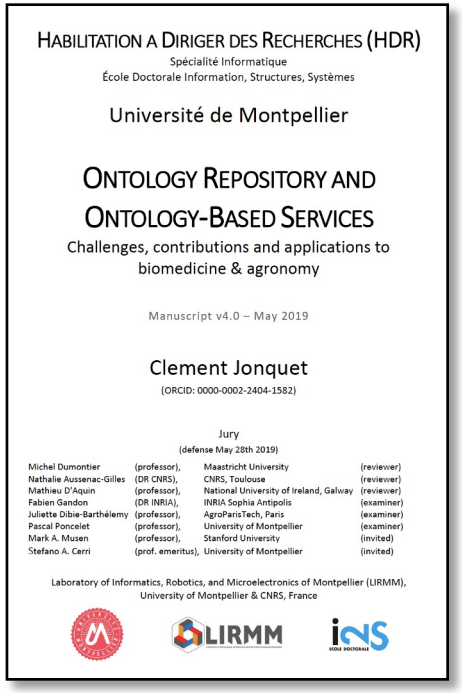
Findable **A**ccessible **I**nteroperable **R**e-usable







Challenges for ontology repositories



HABILITATION A DIRIGER DES RECHERCHES (HDR)
Spécialité Informatique
École Doctorale Information, Structures, Systèmes

Université de Montpellier

ONTOLOGY REPOSITORY AND ONTOLOGY-BASED SERVICES
Challenges, contributions and applications to biomedicine & agronomy

Manuscript v4.0 – May 2019

Clement Jonquet
(ORCID: 0000-0002-2404-1582)

Jury
(defense May 28th 2019)

Michel Dumontier	(professor),	Maastricht University	(reviewer)
Nathalie Aussenac-Gilles	(DR CNRS),	CNRS, Toulouse	(reviewer)
Mathieu D'Aquin	(professor),	National University of Ireland, Galway	(reviewer)
Fabien Gandon	(DR INRIA),	INRIA Sophia Antipolis	(examiner)
Juliette Dibie-Barthélemy	(professor),	AgroParisTech, Paris	(examiner)
Pascal Poncelet	(professor),	University of Montpellier	(examiner)
Mark A. Musen	(professor),	Stanford University	(invited)
Stefano A. Cerri	(prof. emeritus),	University of Montpellier	(invited)

Laboratory of Informatics, Robotics, and Microelectronics of Montpellier (LIRMM), University of Montpellier & CNRS, France

Ontology metadata, evaluation and selection

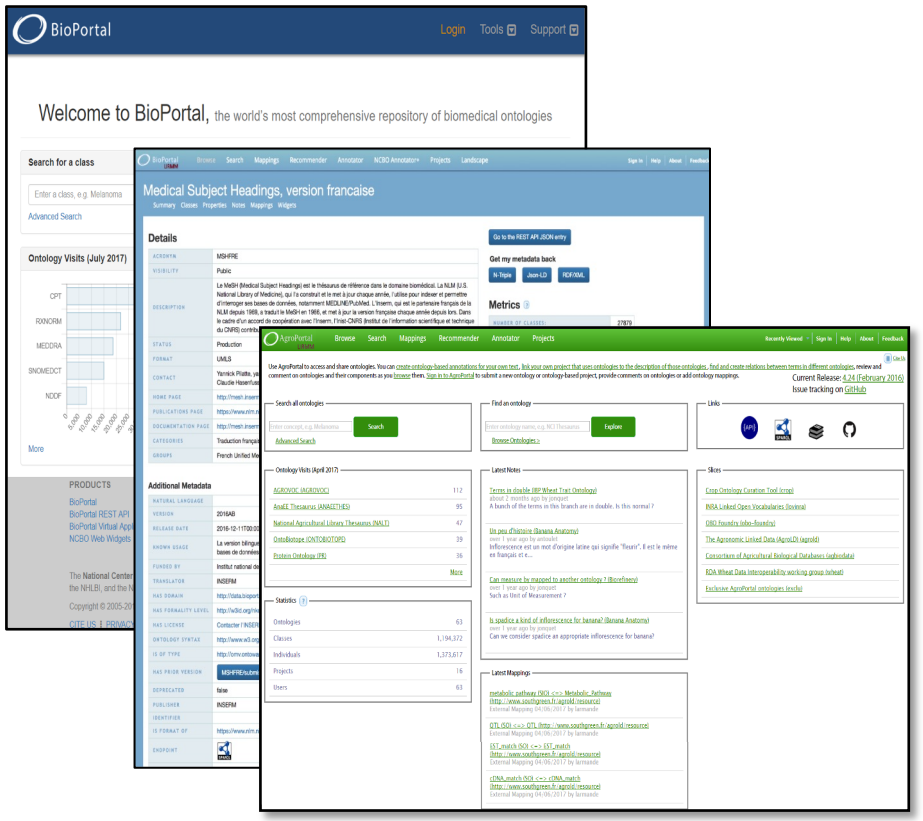
Multilingualism

Ontology alignment

Generic ontology-based services
(especially for free text data)

Annotations and linked data

Scalability and interoperability

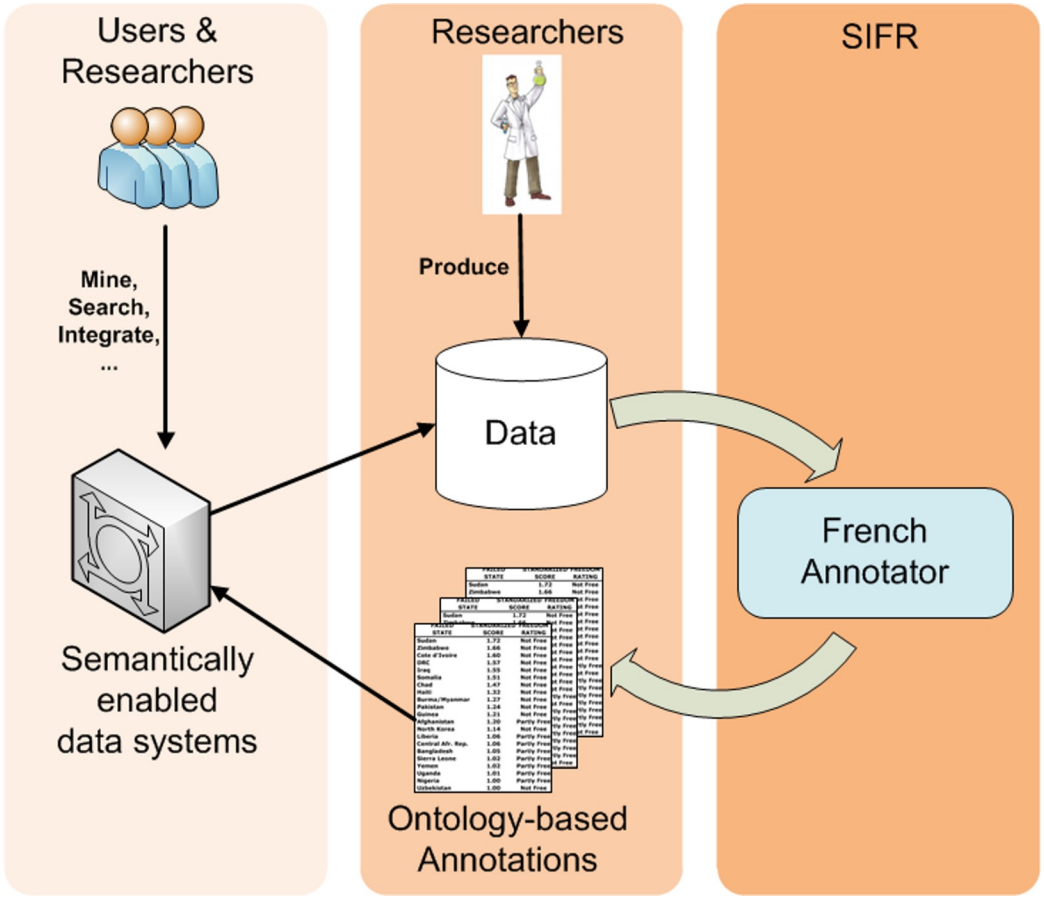


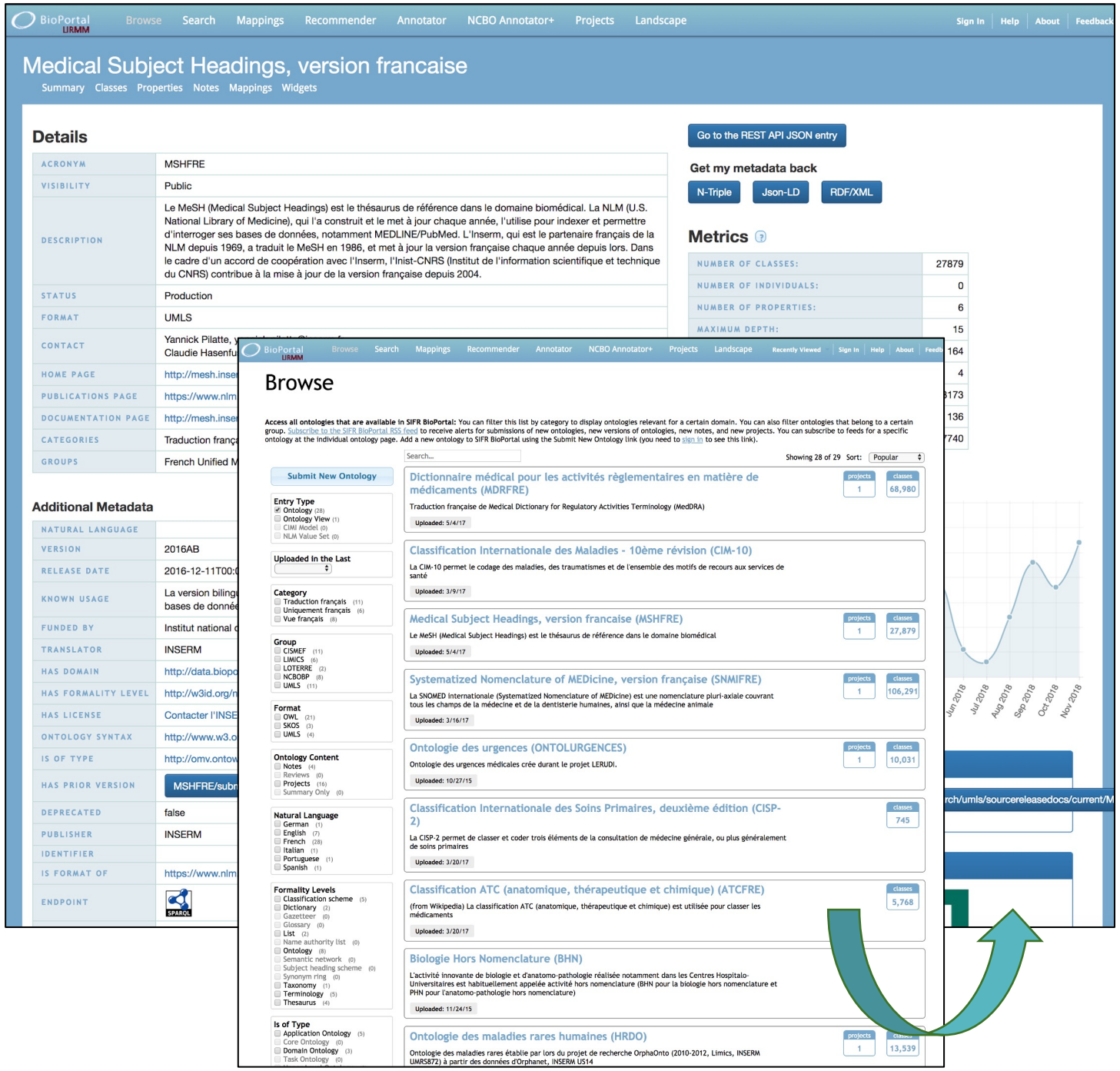
Collaborative projects on ontology-based services in biomedicine and agronomy

SIFR: Semantic Indexing of French Biomedical Data Resources

http://www.lirmm.fr/sifr

- Ontology-based services to index, mine and retrieve French biomedical data
- In France, there is already a reference repository for medical terminologies but nothing public for annotation
- Crucial need for tools & services for French biomedical data





A dedicated version of BioPortal for French ontologies

http://bioportal.lirmm.fr

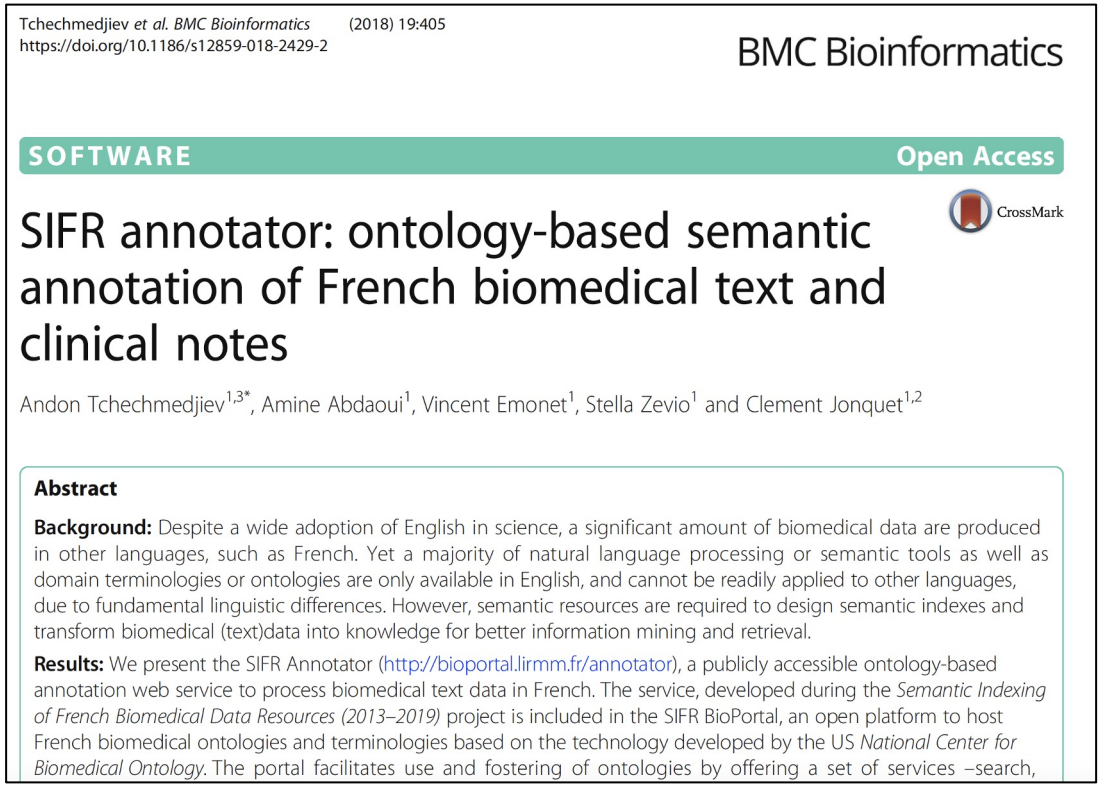
28 monolingual ontologies/terminologies

- From the UMLS or HeTOP or uploaded by users
- Cleaned and checked for annoation

C. Jonquet, A. Annane, K. Bouarech, V. Emonet & S. Melzi. **SIFR BioPortal: French biomedical ontologies and terminologies available for semantic annotation**, In *16th Journées Francophones d'Informatique Médicale, JFIM'16*. Genève, Suisse, July 2016.

SIFR Annotator

- Detect biomedical entities in French biomedical text
- Use semantics inside ontologies
- Performs comparably to other knowledge-based annotation approaches
  - 3 corpus (titles from French MEDLINE, EMEA drug labels and CépiDC death certificates)
  - Participate in CLEF eHealth 2017 competition



Tchechmedjiev *et al. BMC Bioinformatics* (2018) 19:405
https://doi.org/10.1186/s12859-018-2429-2

BMC Bioinformatics

SOFTWARE Open Access

SIFR annotator: ontology-based semantic annotation of French biomedical text and clinical notes

Andon Tchechmedjiev[1,3*], Amine Abdaoui[1], Vincent Emonet[1], Stella Zevio[1] and Clement Jonquet[1,2]

Abstract

Background: Despite a wide adoption of English in science, a significant amount of biomedical data are produced in other languages, such as French. Yet a majority of natural language processing or semantic tools as well as domain terminologies or ontologies are only available in English, and cannot be readily applied to other languages, due to fundamental linguistic differences. However, semantic resources are required to design semantic indexes and transform biomedical (text)data into knowledge for better information mining and retrieval.

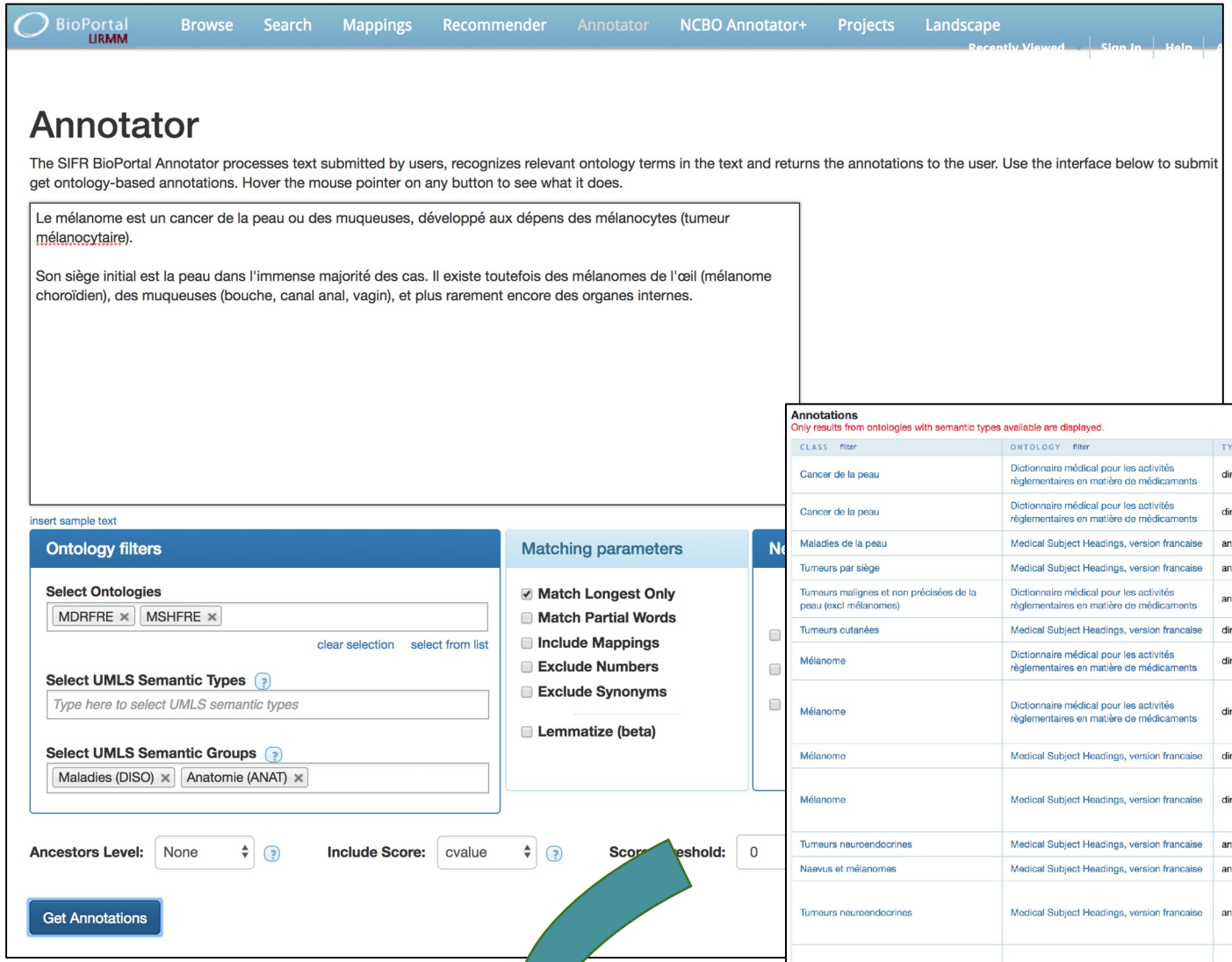
Results: We present the SIFR Annotator (http://bioportal.lirmm.fr/annotator), a publicly accessible ontology-based annotation web service to process biomedical text data in French. The service, developed during the *Semantic Indexing of French Biomedical Data Resources (2013–2019)* project is included in the SIFR BioPortal, an open platform to host French biomedical ontologies and terminologies based on the technology developed by the US *National Center for Biomedical Ontology*. The portal facilitates use and fostering of ontologies by offering a set of services –search,

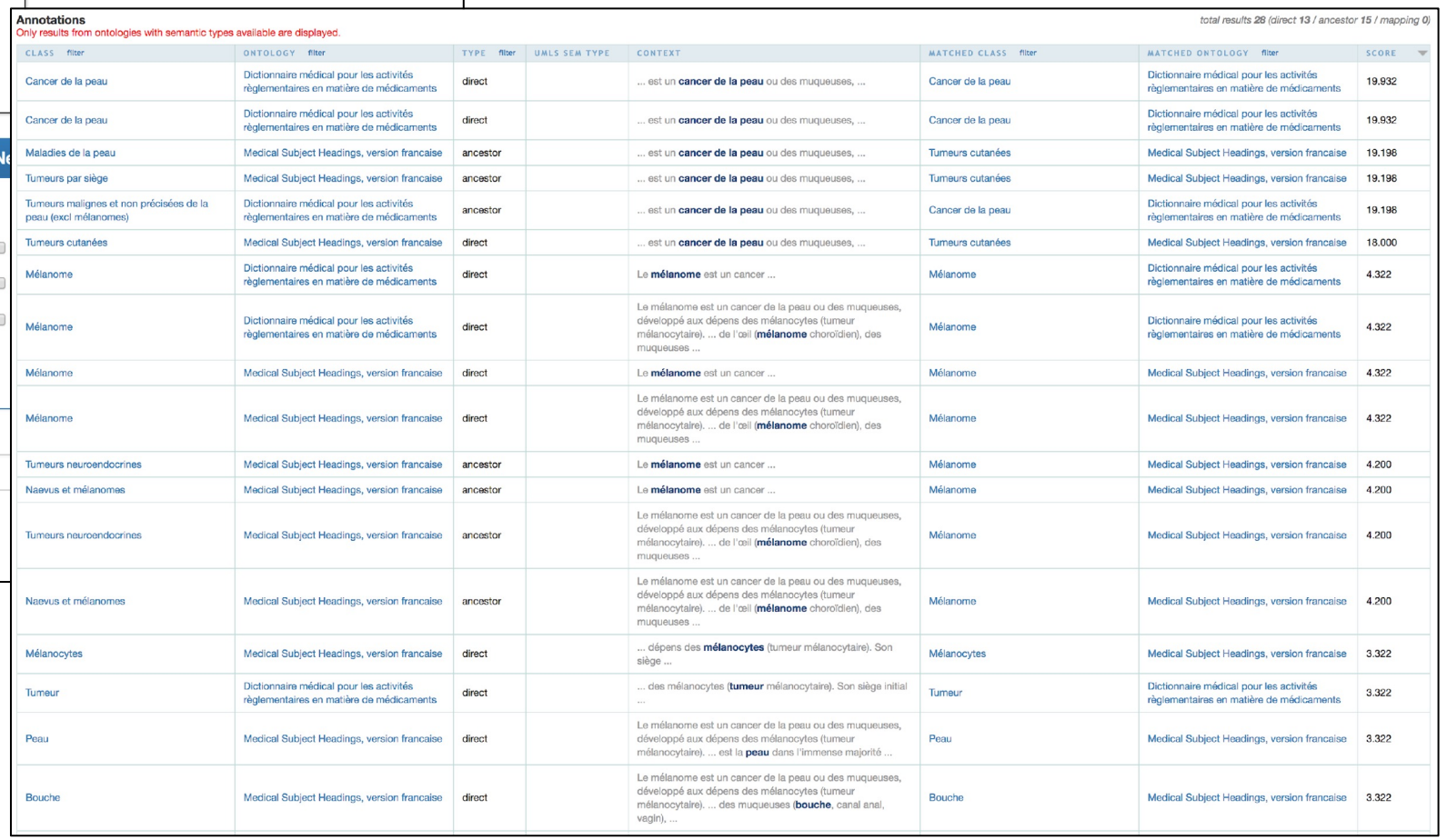
- Easy to use web service
  - Free and open access
  - Easy to plug-in external workflows
  - Annotations in several formats with concept URIs
  - Multiple parameters

French/SIFR Annotator

Detect biomedical entities in French biomedical text

http://bioportal.lirmm.fr/annotator





Annotations

Only results from ontologies with semantic types available are displayed.

total results 28 (direct 13 / ancestor 15 / mapping 0)

CLASS	ONTOLOGY	TYPE	UMLS SEM TYPE	CONTEXT	MATCHED CLASS	MATCHED ONTOLOGY	SCORE
Cancer de la peau	Dictionnaire médical pour les activités règlementaires en matière de médicaments	direct		... est un **cancer de la peau** ou des muqueuses, ...	Cancer de la peau	Dictionnaire médical pour les activités règlementaires en matière de médicaments	19.932
Cancer de la peau	Dictionnaire médical pour les activités règlementaires en matière de médicaments	direct		... est un **cancer de la peau** ou des muqueuses, ...	Cancer de la peau	Dictionnaire médical pour les activités règlementaires en matière de médicaments	19.932
Maladies de la peau	Medical Subject Headings, version francaise	ancestor		... est un **cancer de la peau** ou des muqueuses, ...	Tumeurs cutanées	Medical Subject Headings, version francaise	19.198
Tumeurs par siège	Medical Subject Headings, version francaise	ancestor		... est un **cancer de la peau** ou des muqueuses, ...	Tumeurs cutanées	Medical Subject Headings, version francaise	19.198
Tumeurs malignes et non précisées de la peau (excl mélanomes)	Dictionnaire médical pour les activités règlementaires en matière de médicaments	ancestor		... est un **cancer de la peau** ou des muqueuses, ...	Cancer de la peau	Dictionnaire médical pour les activités règlementaires en matière de médicaments	19.198
Tumeurs cutanées	Medical Subject Headings, version francaise	direct		... est un **cancer de la peau** ou des muqueuses, ...	Tumeurs cutanées	Medical Subject Headings, version francaise	18.000
Mélanome	Dictionnaire médical pour les activités règlementaires en matière de médicaments	direct		Le **mélanome** est un cancer ...	Mélanome	Dictionnaire médical pour les activités règlementaires en matière de médicaments	4.322
Mélanome	Dictionnaire médical pour les activités règlementaires en matière de médicaments	direct		Le mélanome est un cancer de la peau ou des muqueuses, développé aux dépens des mélanocytes (tumeur mélanocytaire). ... de l'œil (**mélanome** choroïdien), des muqueuses ...	Mélanome	Dictionnaire médical pour les activités règlementaires en matière de médicaments	4.322
Mélanome	Medical Subject Headings, version francaise	direct		Le **mélanome** est un cancer ...	Mélanome	Medical Subject Headings, version francaise	4.322
Mélanome	Medical Subject Headings, version francaise	direct		Le mélanome est un cancer de la peau ou des muqueuses, développé aux dépens des mélanocytes (tumeur mélanocytaire). ... de l'œil (**mélanome** choroïdien), des muqueuses ...	Mélanome	Medical Subject Headings, version francaise	4.322
Tumeurs neuroendocrines	Medical Subject Headings, version francaise	ancestor		Le **mélanome** est un cancer ...	Mélanome	Medical Subject Headings, version francaise	4.200
Naevus et mélanomes	Medical Subject Headings, version francaise	ancestor		Le **mélanome** est un cancer ...	Mélanome	Medical Subject Headings, version francaise	4.200
Tumeurs neuroendocrines	Medical Subject Headings, version francaise	ancestor		Le mélanome est un cancer de la peau ou des muqueuses, développé aux dépens des mélanocytes (tumeur mélanocytaire). ... de l'œil (**mélanome** choroïdien), des muqueuses ...	Mélanome	Medical Subject Headings, version francaise	4.200
Naevus et mélanomes	Medical Subject Headings, version francaise	ancestor		Le mélanome est un cancer de la peau ou des muqueuses, développé aux dépens des mélanocytes (tumeur mélanocytaire). ... de l'œil (**mélanome** choroïdien), des muqueuses ...	Mélanome	Medical Subject Headings, version francaise	4.200
Mélanocytes	Medical Subject Headings, version francaise	direct		... dépens des **mélanocytes** (tumeur mélanocytaire). Son siège ...	Mélanocytes	Medical Subject Headings, version francaise	3.322
Tumeur	Dictionnaire médical pour les activités règlementaires en matière de médicaments	direct		... des mélanocytes (**tumeur** mélanocytaire). Son siège initial ...	Tumeur	Dictionnaire médical pour les activités règlementaires en matière de médicaments	3.322
Peau	Medical Subject Headings, version francaise	direct		Le mélanome est un cancer de la peau ou des muqueuses, développé aux dépens des mélanocytes (tumeur mélanocytaire). ... est la **peau** dans l'immense majorité ...	Peau	Medical Subject Headings, version francaise	3.322
Bouche	Medical Subject Headings, version francaise	direct		Le mélanome est un cancer de la peau ou des muqueuses, développé aux dépens des mélanocytes (tumeur mélanocytaire). ... des muqueuses (**bouche**, canal anal, vagin), ...	Bouche	Medical Subject Headings, version francaise	3.322

Tchechmedjiev, A., Abdaoui, A., Emonet, V., Zevio, S., & Jonquet, C. (2018). SIFR annotator: ontology-based semantic annotation of French biomedical text and clinical notes. *BMC bioinformatics*, *19*(1), 1-26.

AgroPortal: a vocabulary and ontology repository for agronomy

http://agroportal.lirmm.fr

- Develop and support a reference ontology repository
  - Primary focus on the agronomy & close related domains (plant sciences, food and biodiversity)

- Reusing the NCBO BioPortal technology
  - Avoid to re-implement what has been done, facilitate interoperability
  - Reusing the scientific outcomes, experience & methods of the biomedical domain

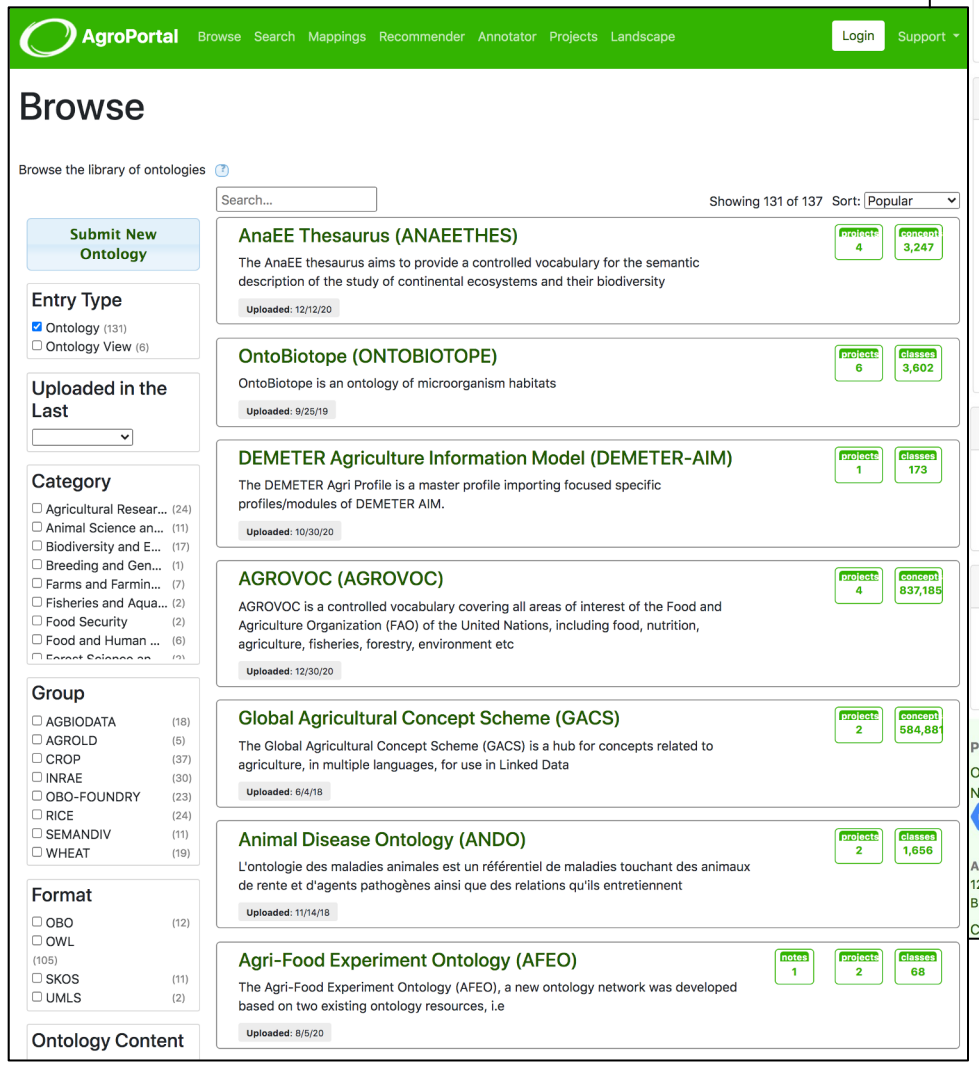
- Enable straightforward use of agronomic related ontologies
  - Respect the requirements & specificities of the agronomic community
  - Fully semantic web compliant infrastructure
  - Enable new science

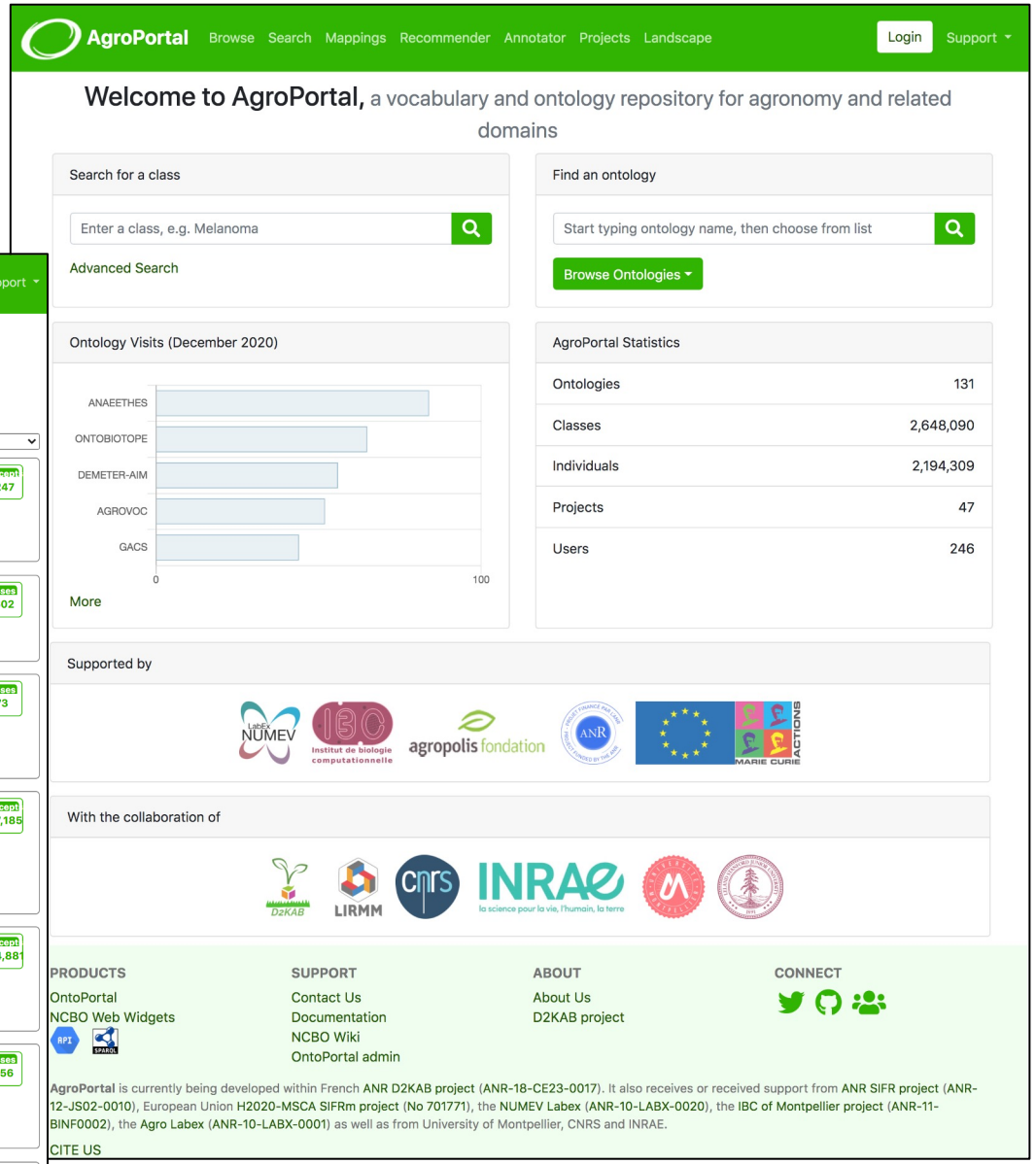
AgroPortal an ontology repository

for agronomy, food, plant sciences & biodiversity

http://agroportal.lirmm.fr

- Publish, search, download
- Browse, visualize
- Peer review
- Versioning
- Annotation
- Recommendation
- Mapping
- Notes
- Projects



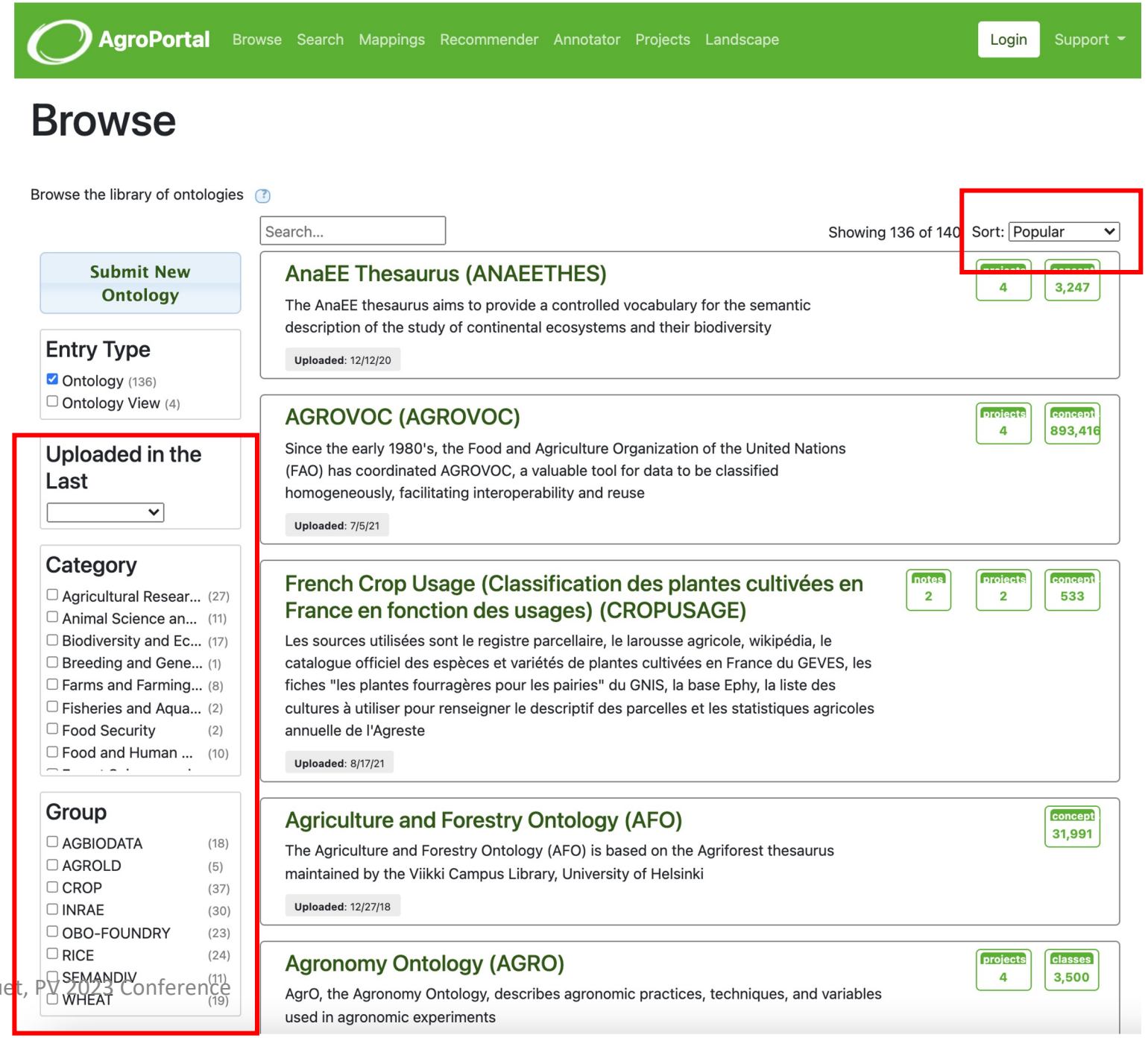


- 131 ontologies, 90 candidates
- 5 driving use cases
- ~240 registered users

C. Jonquet, A. Toulet, (…) P. Larmande. **AgroPortal: an ontology repository for agronomy**, *Computers and Electronics in Agriculture*. Jan 2018. 144, pp.126-143. Elsevier.

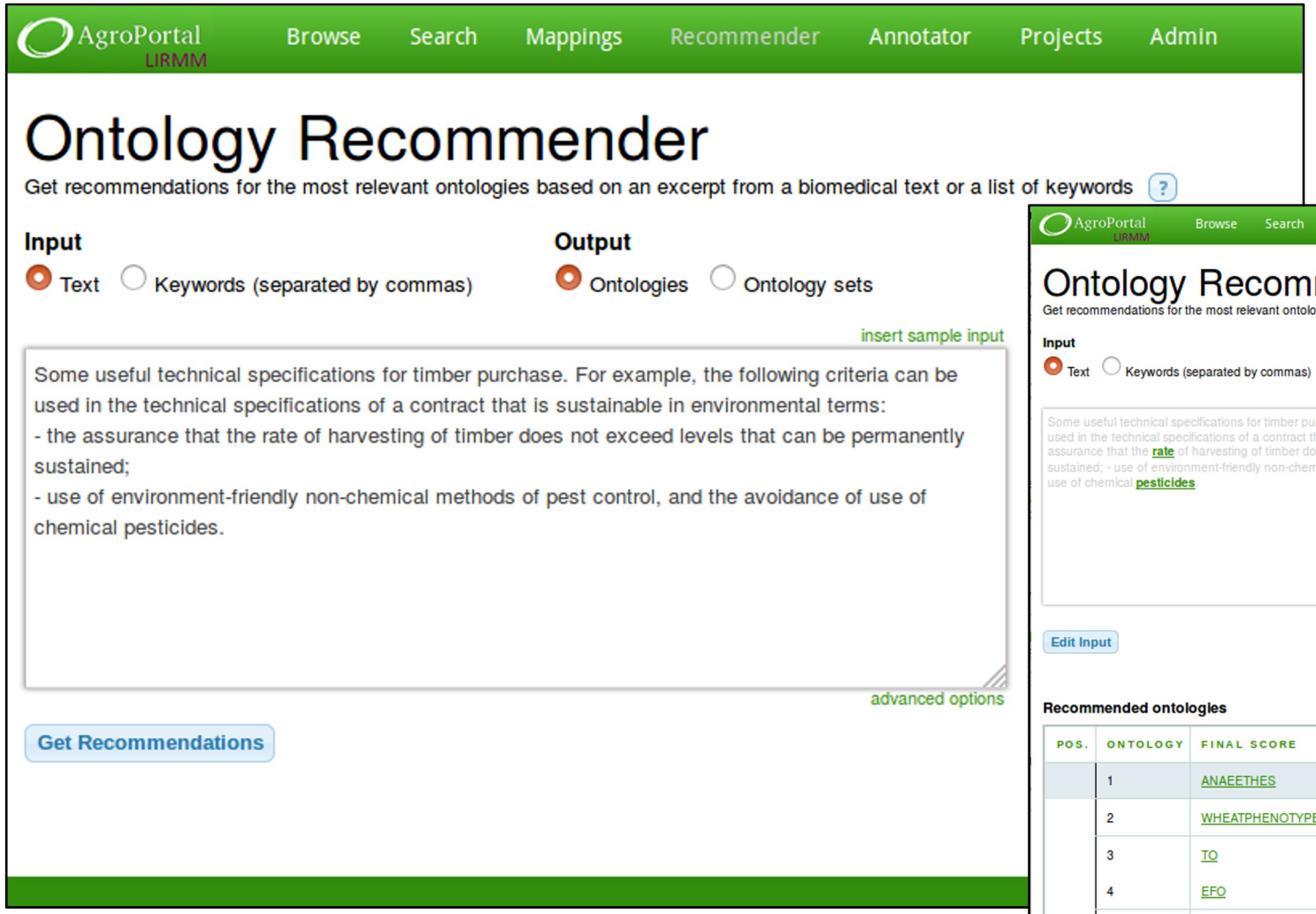
Browse and select ontologies

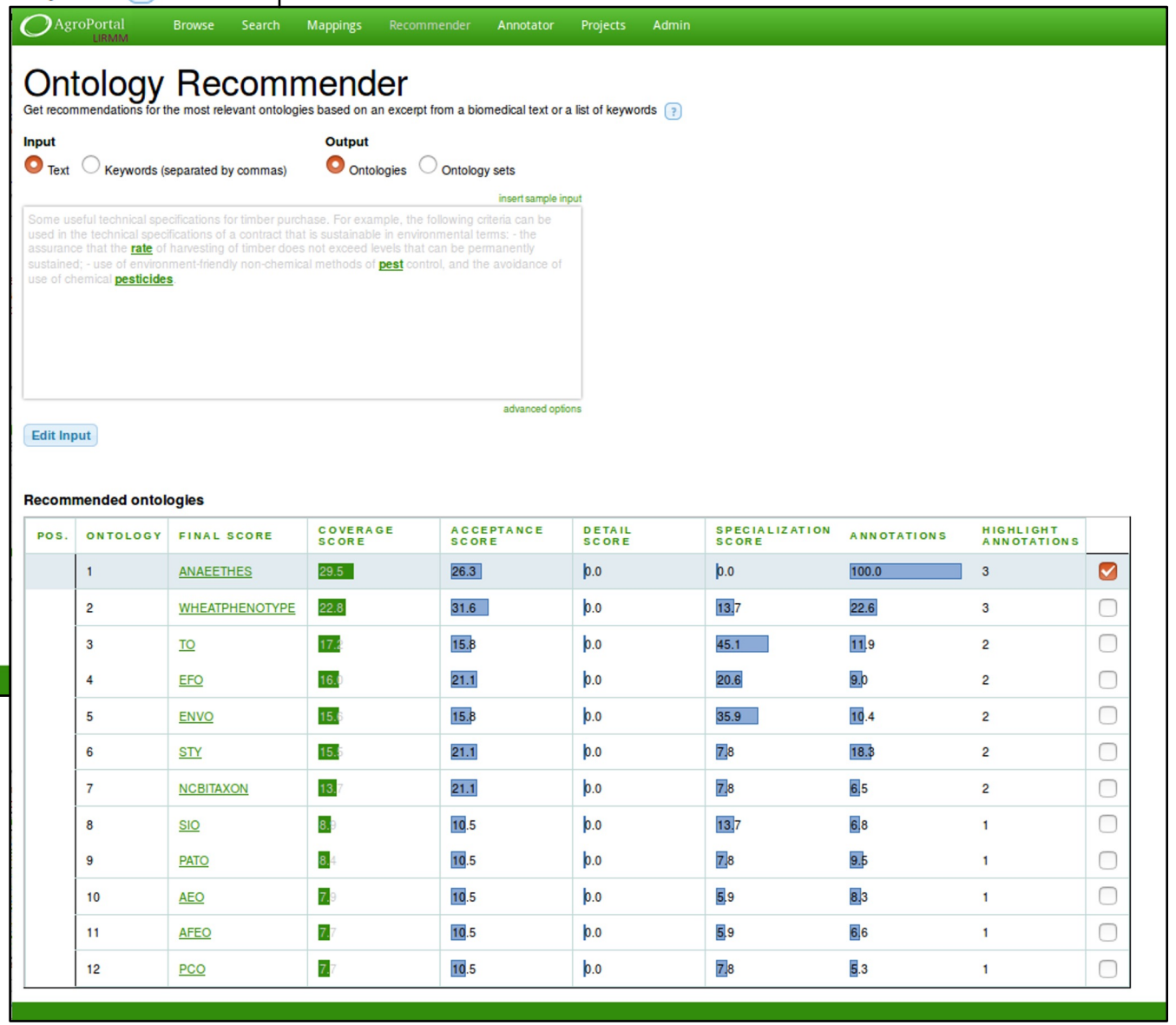
- Allows to search, order and select ontologies using a facetted search approach, based on the metadata



Describe ontologies with semantic metadata

- Display “per ontology”
  - Ontology specific properties => viewable and editable within the ontology specific page
- Everything you need to know about an ontology
- URIs used in the backend to store the information
  - e.g., CC-BY => https://creativecommons.org/licenses/by-nd/4.0/
- “Get my metadata back” buttons

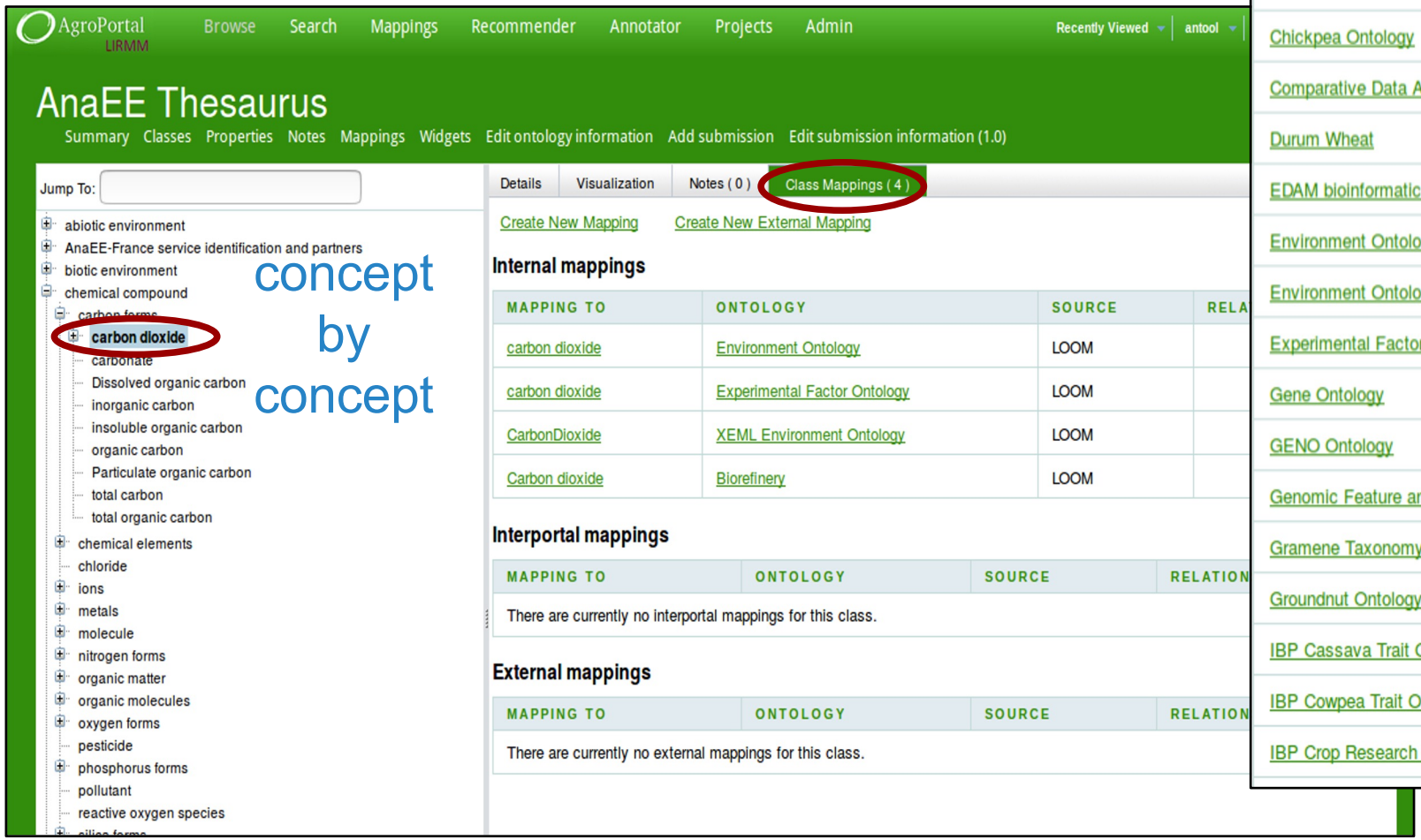




AgroPortal Recommender

get the most relevant ontologies for your data

Align ontologies one another



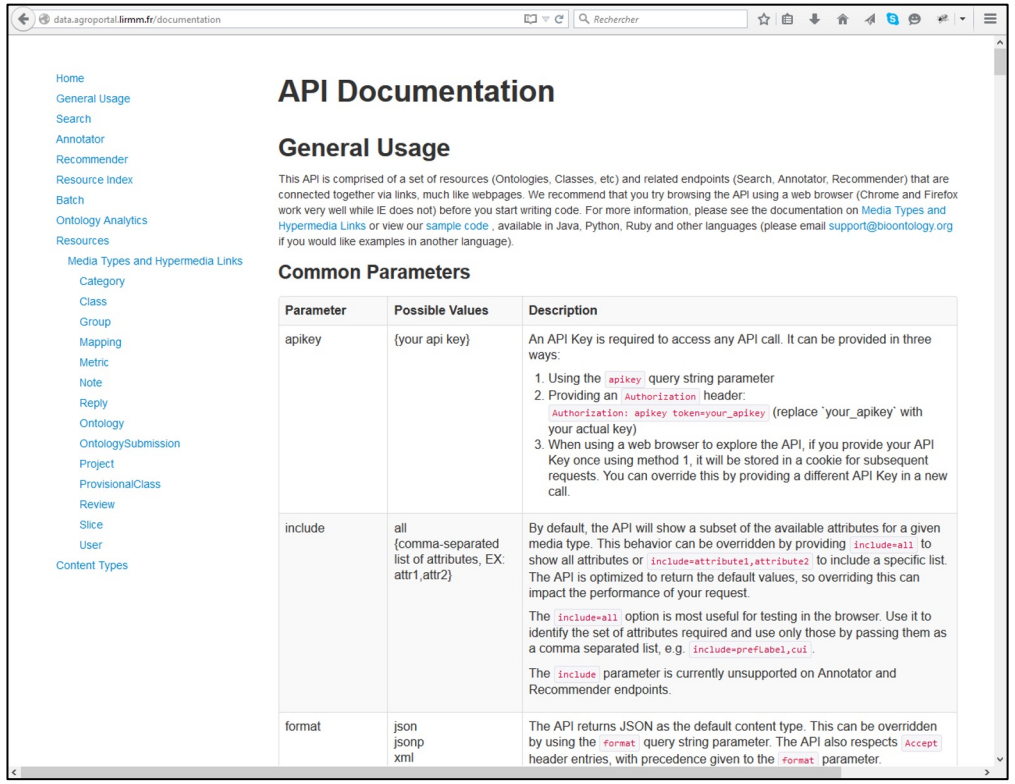
concept by concept

Mappings

ONTOLOGY	MAPPINGS
Agri-Food Experiment Ontology	1
Agricultural Experiments Ontology	5
Banana Anatomy	2
Basic Formal Ontology	1
Biorefinery	13
Cell Ontology	4
Chickpea Ontology	14
Comparative Data Analysis Ontology	3
Durum Wheat	2
EDAM bioinformatics operations, data types, formats, identifiers and topics	25
Environment Ontology	72
Environment Ontology for Livestock	10
Experimental Factor Ontology	93
Gene Ontology	5
GENO Ontology	5
Genomic Feature and Variation Ontology	5
Gramene Taxonomy Ontology	3
Groundnut Ontology	16
IBP Cassava Trait Ontology	23
IBP Cowpea Trait Ontology	25
IBP Crop Research Ontology	22

REST Service API:

http://data.agroportal.lirmm.fr/documentation



SPARQL endpoint:

http://sparql.agroportal.lirmm.fr

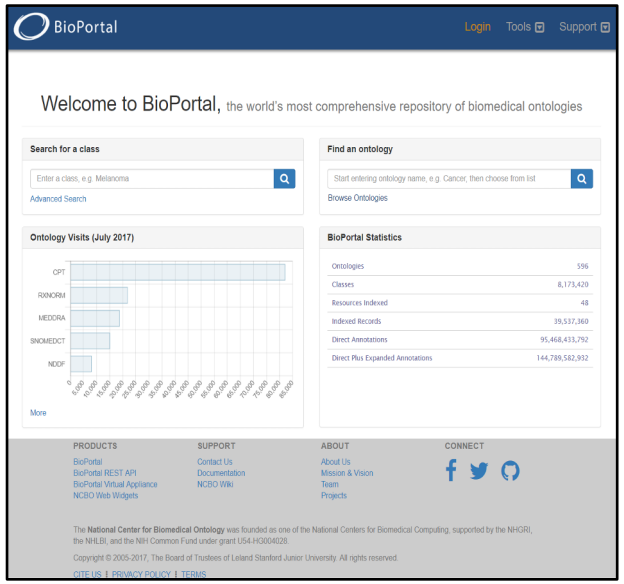
OntoPortal Alliance: Synchronizing and mutualizing research and development efforts

Representing OntoPortal adopters and end users

- to maximize OntoPortal value (state-of-the-art service portfolio)
- to improve OntoPortal software while managing several parallel and different installations
- to increase semantic uptake in science communities and facilitate adoption of the FAIR principles
- to increase the ecosystem's long term operational and financial health

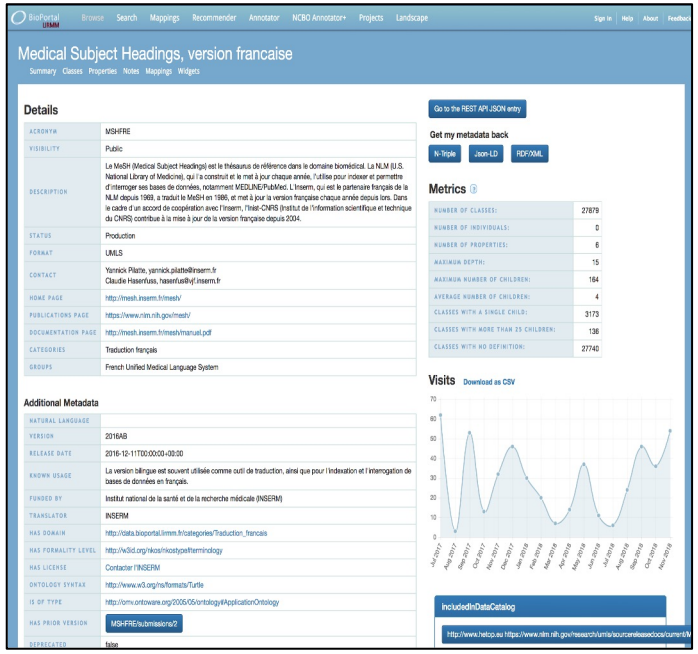
We develop and maintain ontology repositories in the OntoPortal Alliance (1/3)

BioPortal



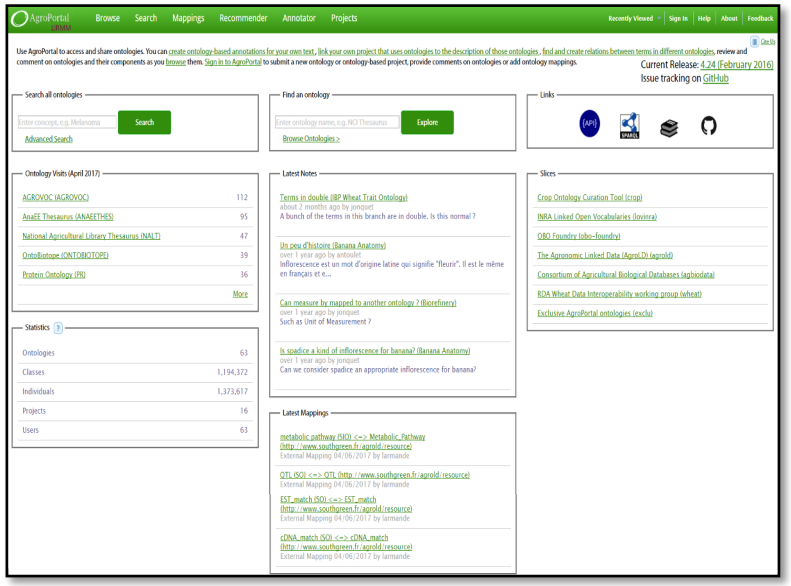
http://bioportal.bioontology.org

SIFR BioPortal



https://bioportal.lirmm.fr

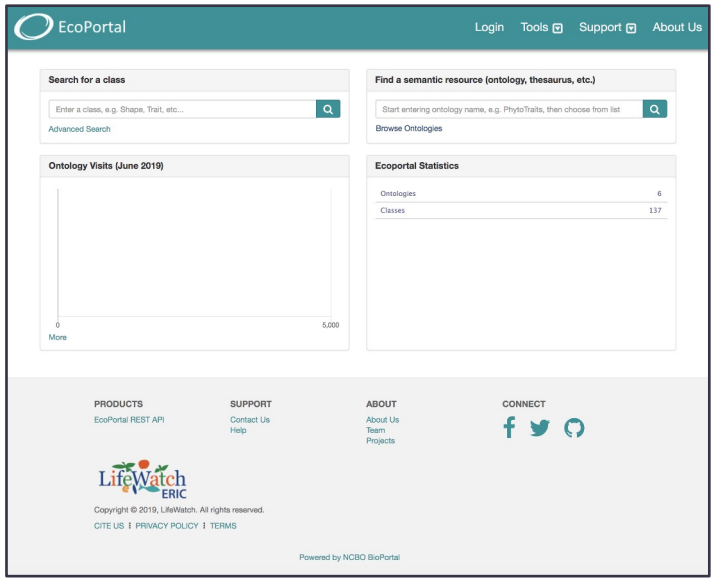
AgroPortal



http://agroportal.lirmm.fr

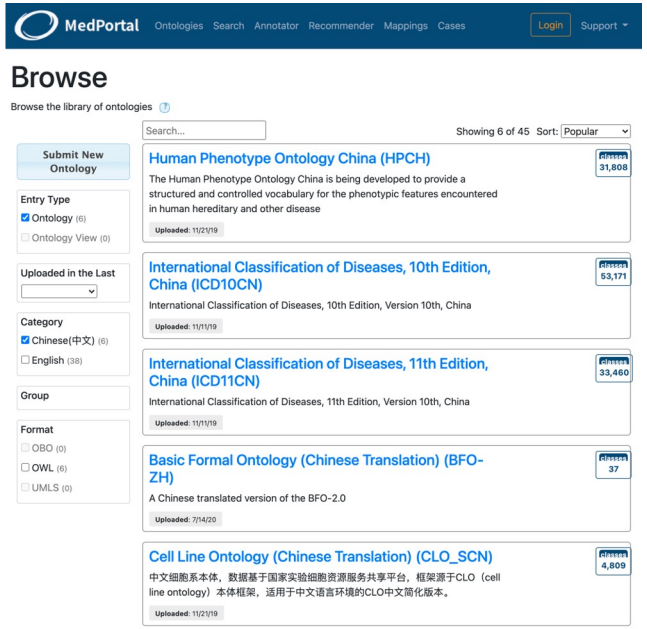
We develop and maintain ontology repositories in the OntoPortal Alliance (2/3)

EcoPortal



http://ecoportal.lifewatchitaly.eu

MedPortal



http://medportal.bmicc.cn/

And other installations with

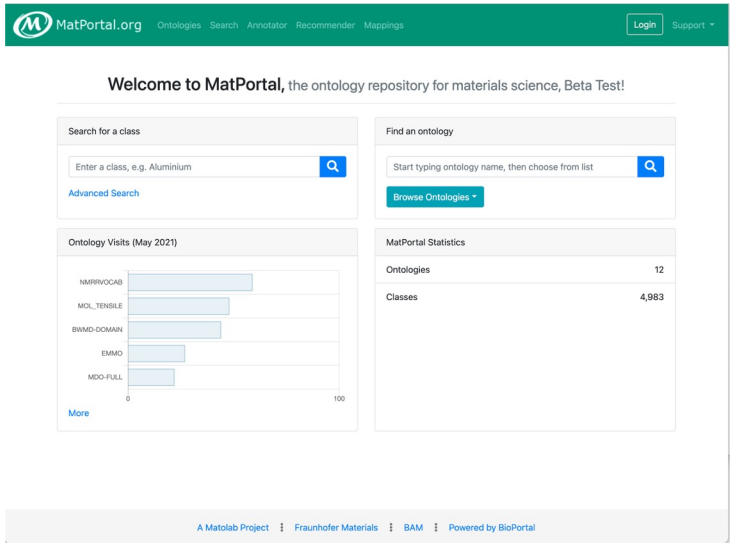
Cogni.zone

We develop and maintain ontology repositories in the OntoPortal Alliance (3/3)

BAM

Fraunhofer MATERIALS

MatPortal

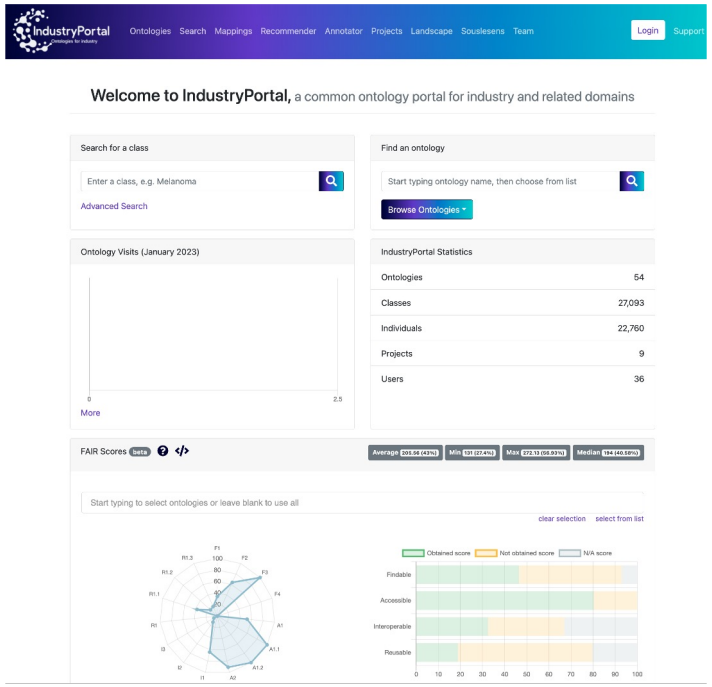


https://matportal.org/

ONTO COMMONS
ONTOLOGY-DRIVEN DATA DOCUMENTATION FOR INDUSTRY COMMONS

INP TOULOUSE ENIT ÉCOLE NATIONALE D'INGÉNIEURS DE TARBES

IndustryPortal



C. Jonquet, PV 2023 Conference

NFDI 4 BIODIVERSITY

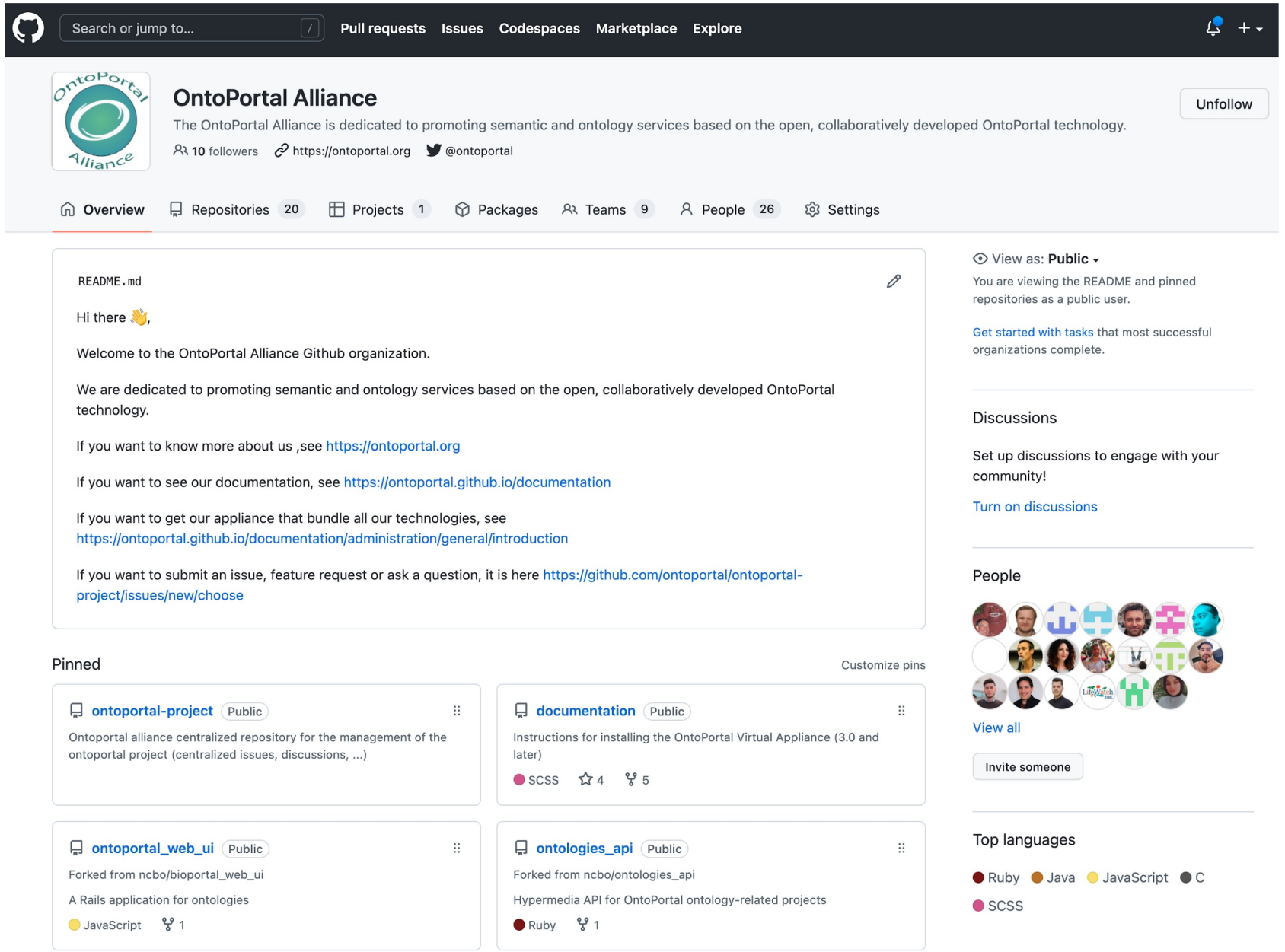
BiodivPortal

DATA TERRA

cnrs

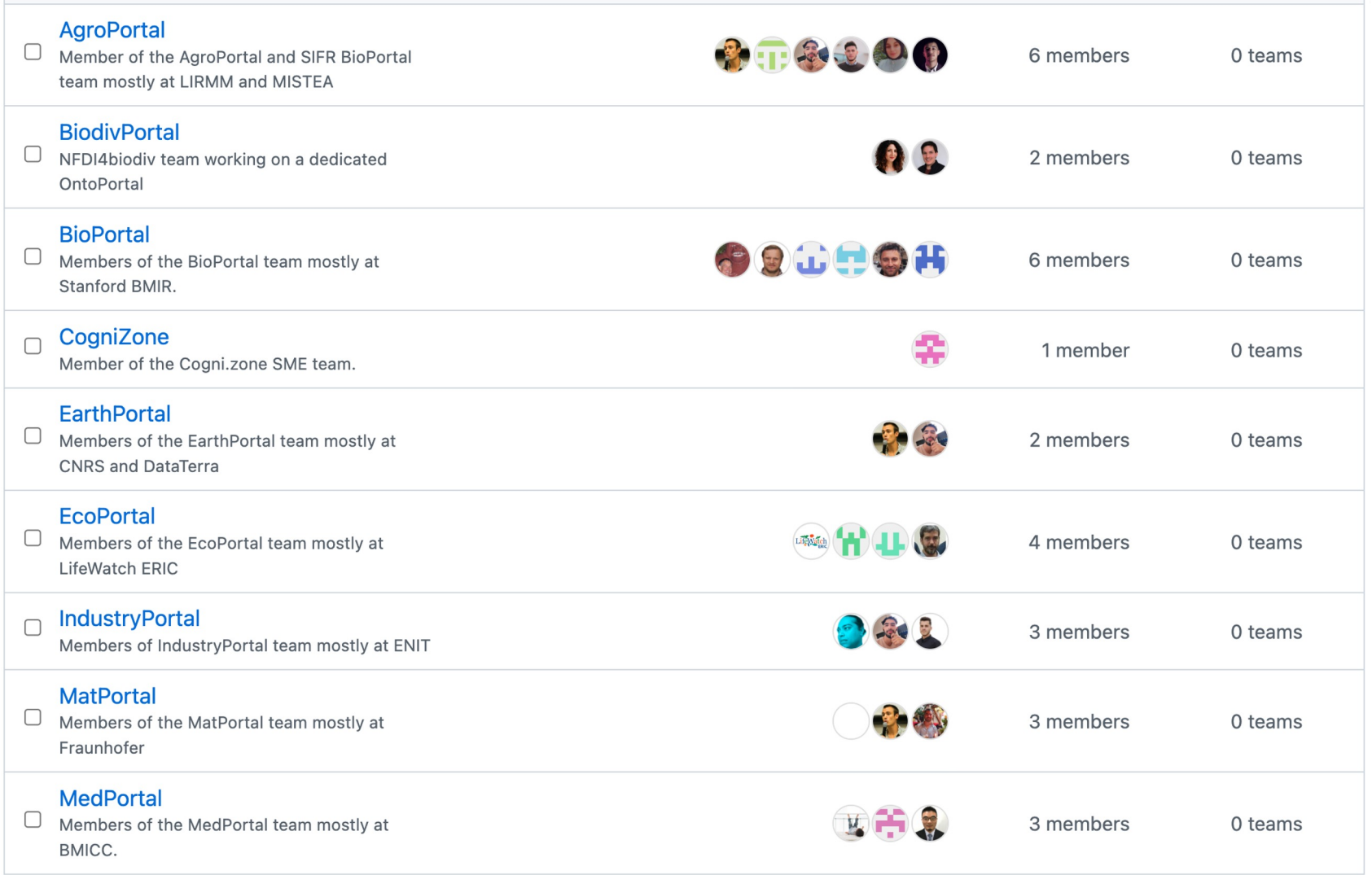
EarthPortal

Making OntoPortal a real open source project



Membership increasing

- 7 existing public repositories. 2 others in the pipe
- 1 active commercial participant
- Multiple interested parties beyond that
  - 60 installation of the appliance

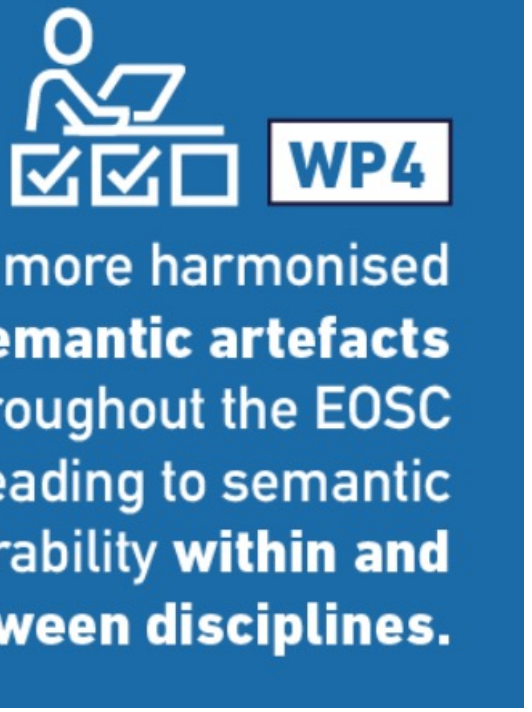


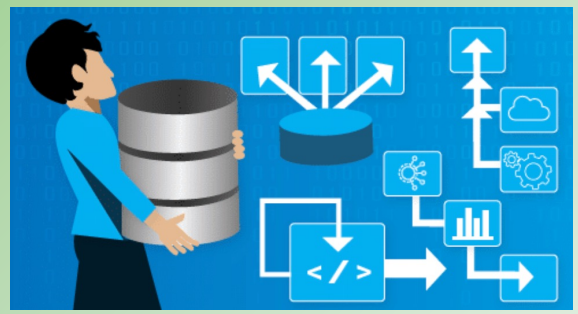
AgroPortal Member of the AgroPortal and SIFR BioPortal team mostly at LIRMM and MISTEA		6 members	0 teams
BiodivPortal NFDI4biodiv team working on a dedicated OntoPortal		2 members	0 teams
BioPortal Members of the BioPortal team mostly at Stanford BMIR.		6 members	0 teams
CogniZone Member of the Cogni.zone SME team.		1 member	0 teams
EarthPortal Members of the EarthPortal team mostly at CNRS and DataTerra		2 members	0 teams
EcoPortal Members of the EcoPortal team mostly at LifeWatch ERIC		4 members	0 teams
IndustryPortal Members of IndustryPortal team mostly at ENIT		3 members	0 teams
MatPortal Members of the MatPortal team mostly at Fraunhofer		3 members	0 teams
MedPortal Members of the MedPortal team mostly at BMICC.		3 members	0 teams

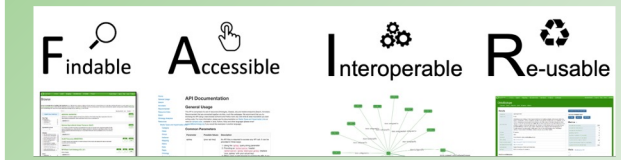
Semantic artefacts are a key elements to achieving FAIR and these artefacts and their catalogues have to be FAIR too

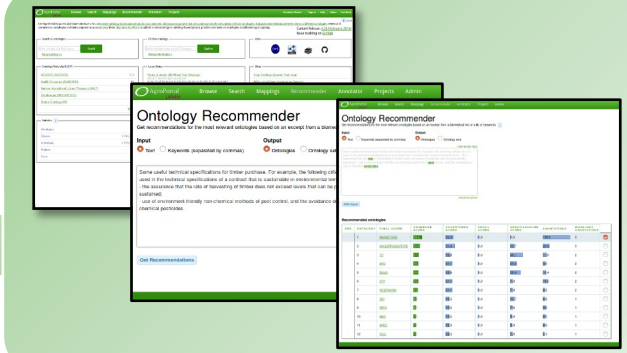
eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC

- AgroPortal
- EcoPortal
- EarthPortal









Summary

Questions ?