AgroPortal: a vocabulary and ontology repository for agronomy

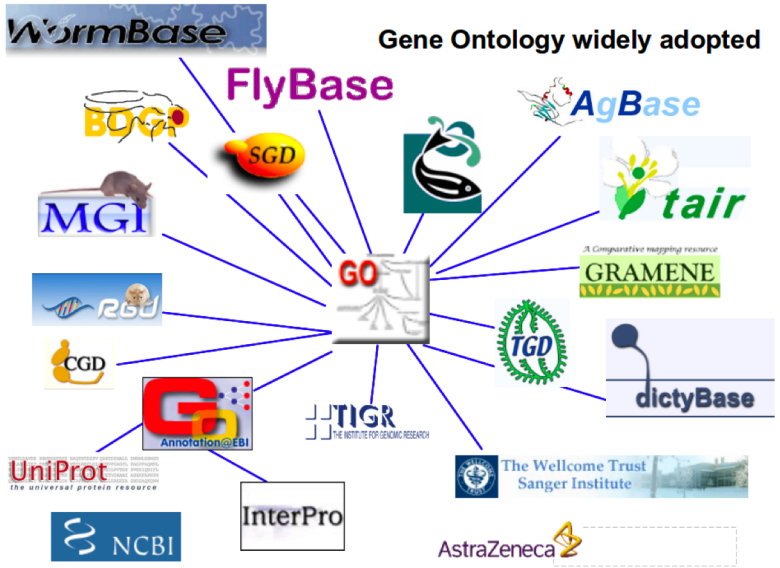
Clement Jonquet
LIRMM, University of Montpellier, France
… soon also at MISTEA, INRAE, Montpellier
jonquet@lirmm.fr

Journée Agro & IA, PFIA 2020
1st July 2020

PFIA 2020 ANGERS

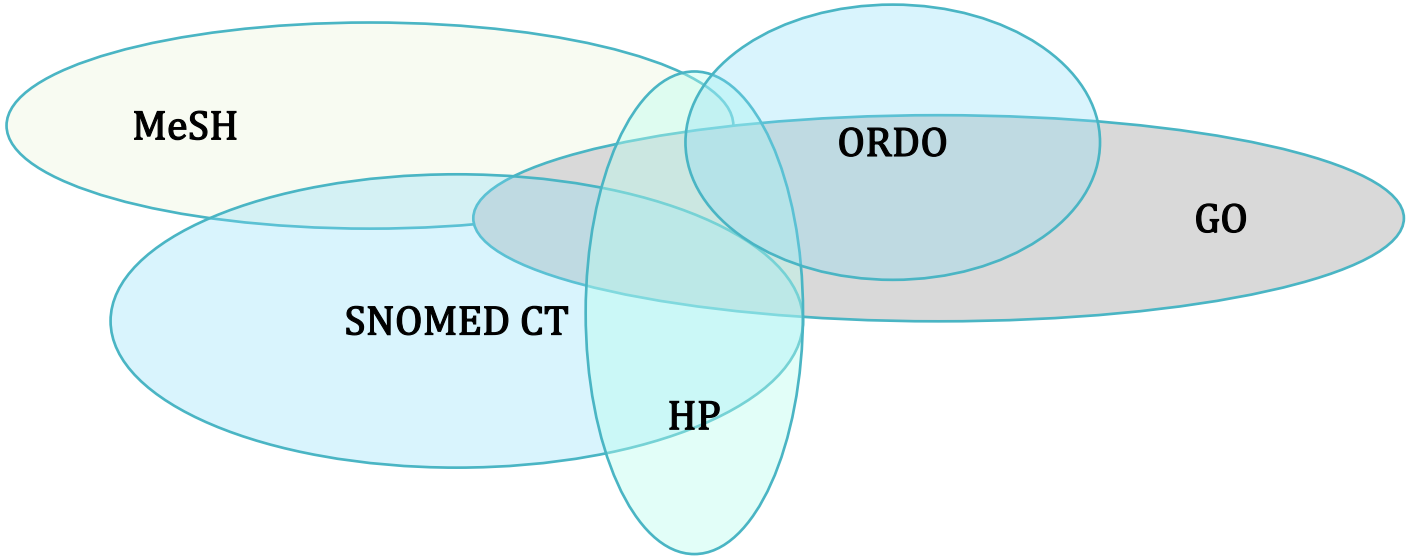
Why ontologies are important in science?

- To provide canonical representation and sharing of scientific knowledge
- To annotate experimental data to enable interpretation, comparison, and discovery across databases
- To facilitate knowledge-based applications for
  - Decision support, reasoning
  - Natural language-processing, text mining
  - Data integration

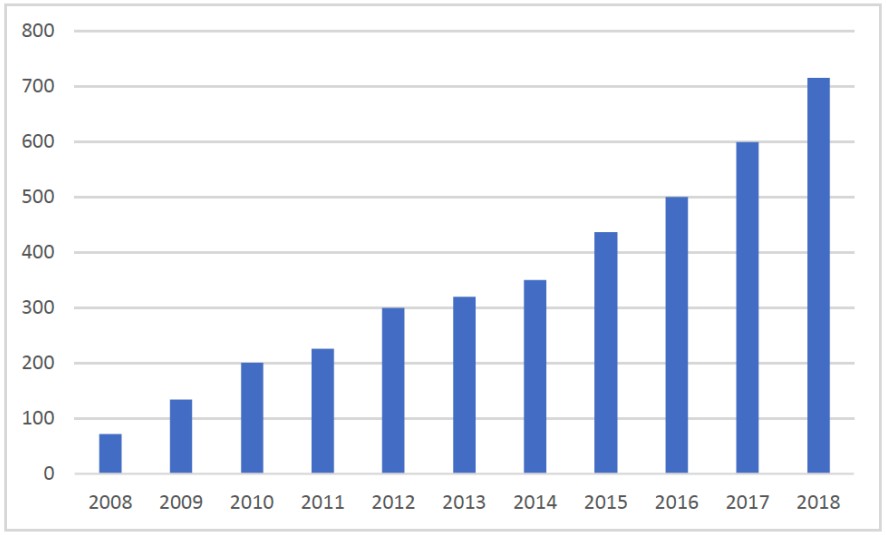


- But ontologies are: spread out, in different formats, of different size, with different structures

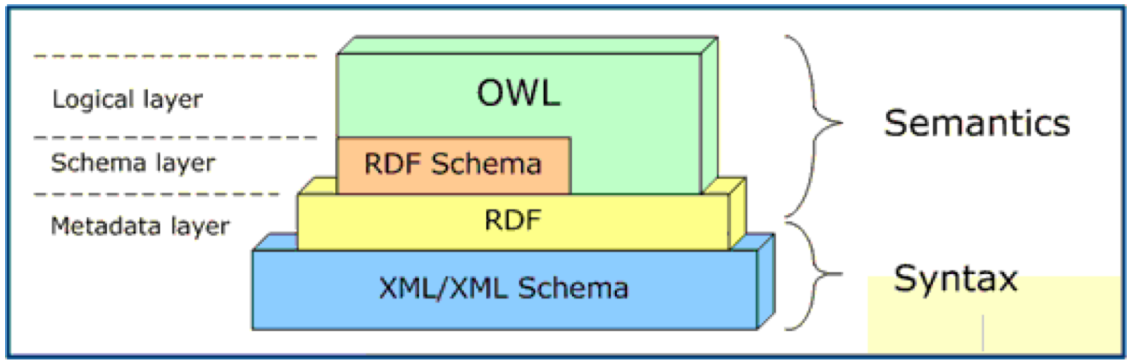
Other issues with ontologies



Overlapping ontologies



Number of ontologies in the NCBO BioPortal



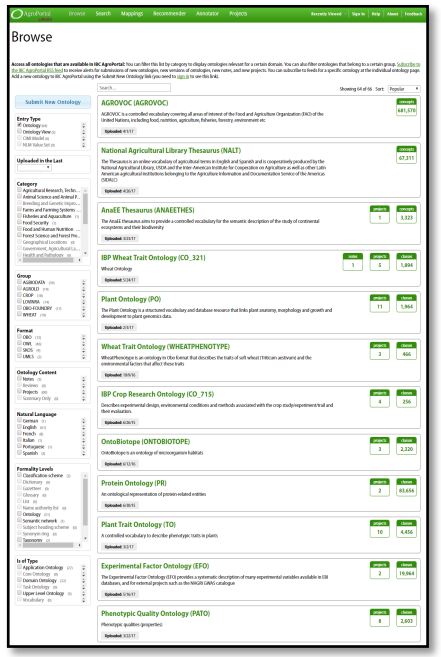
Variety of representation languages

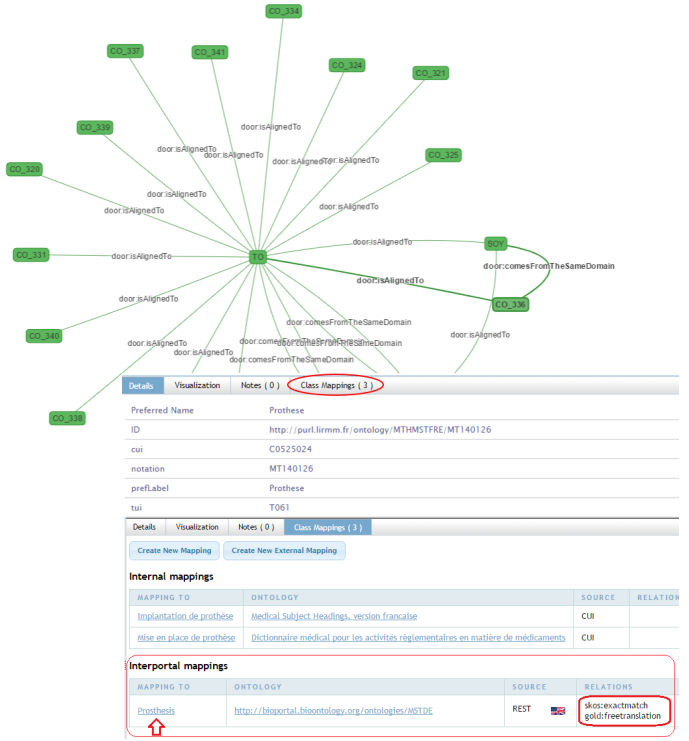
Why ontology repositories are important?

- 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 (or to specific terms)?
- How could you leverage your ontology to enable new science?
- How could you use ontologies without managing them ?

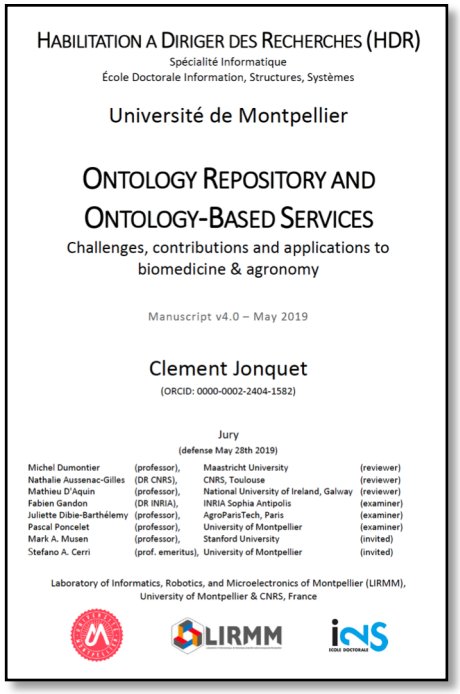
Ontology repositories help to make ontologies FAIR

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

LIRMM

Journée Agro & IA, PFIA 2020 – July 1st 2020

- Ontology metadata, evaluation and selection
- Multilingualism
- Ontology alignment
- Generic ontology-based services (especially for free text data)
- Annotations and linked data
- Scalability and interoperability

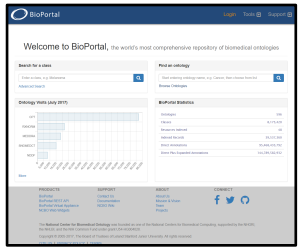
AgroPortal: a vocabulary and ontology repository for agronomy

http://agroportal.lirmm.fr

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

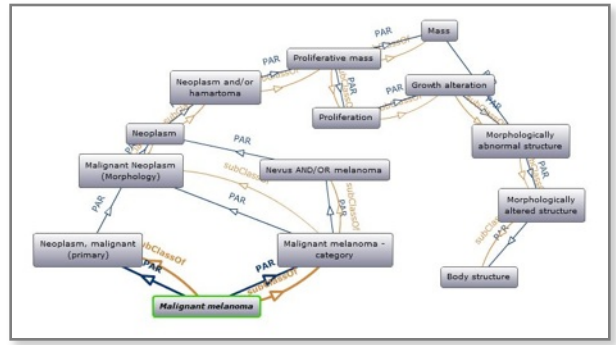
- Reusing the NCBO BioPortal technology
  - Avoid to re-implement what has been done, facilitate interoperability
  - Reusing the scientific outcomes, experience & methods of the biomedical domain

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

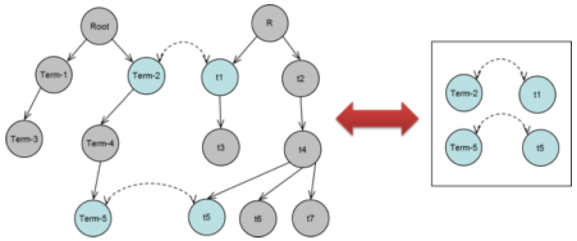


http://bioportal.bioontology.org

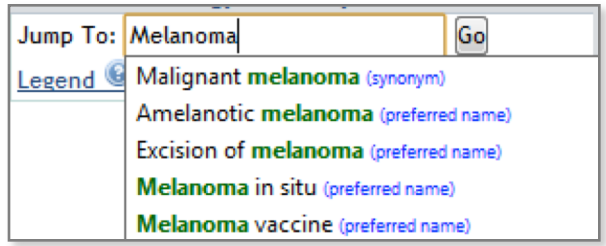
Ontology Services →
- Search
- Traverse
- Comment
- Download



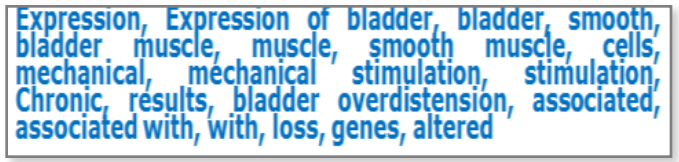
Mapping Services →
- Create
- Upload
- Download



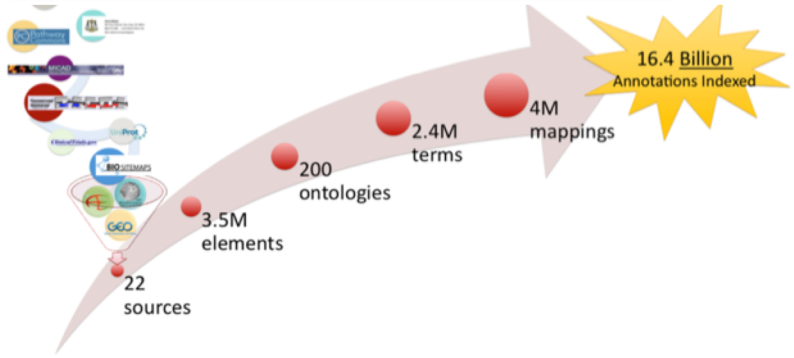
Widgets →
- Tree-view
- Auto-complete
- Graph-view



Annotation → **Term recognition**



Data Access → **Search "data" annotated with a given term**

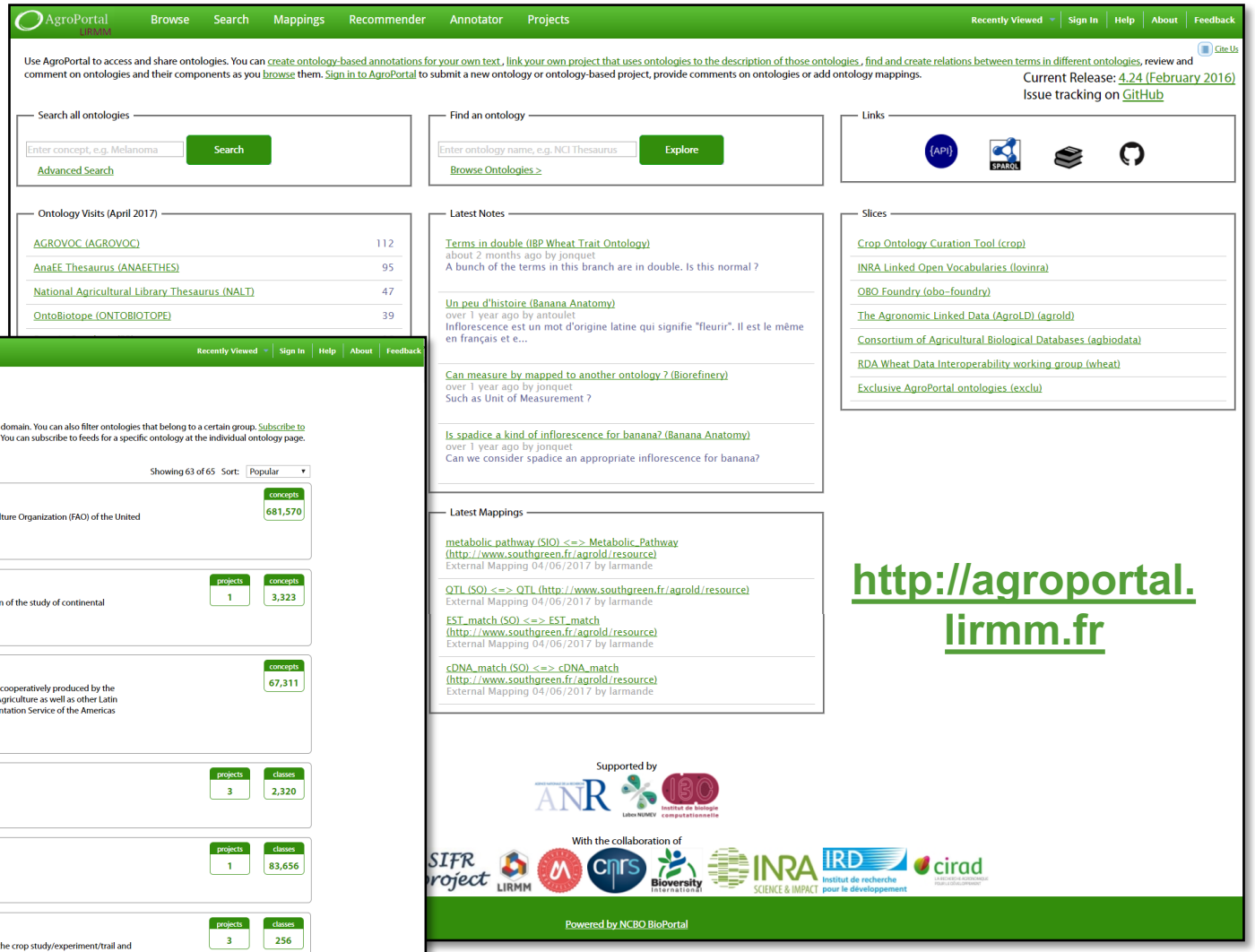


http://data.bioontology.org

AgroPortal an ontology repository

for agronomy, food, plant sciences & biodiversity

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



http://agroportal.lirmm.fr

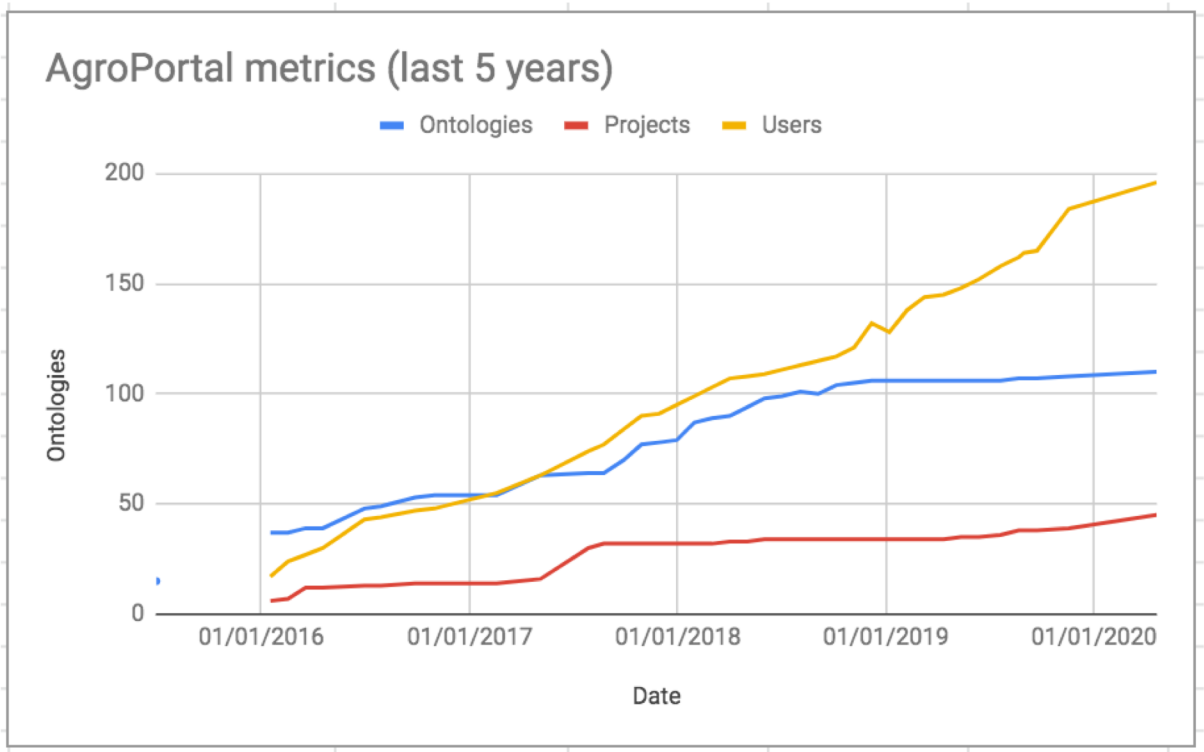
- 122 ontologies, 90 candidates
- 5 driving use cases
- ~200 registered users

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

2016: Five original driving agronomic use cases

- IBC Rice Genomics & AgroLD project
  - Data integration and knowledge management related to rice (P. Larmande)
- RDA Wheat Data Interoperability working group
  - Common framework for publishing wheat data (E. Dzalé-Yeumo)
- LovInra : INRA Linked Open Vocabularies
  - Vocabularies produced by INRA scientists (S. Aubin)
- Crop Ontology project
  - Ontologies for describing crop germplasm & traits (E. Arnaud)
- GODAN global map of agri-food data standards
  - VEST/AgroPortal MAP of standards (V. Pesce)

A growing interest in the community



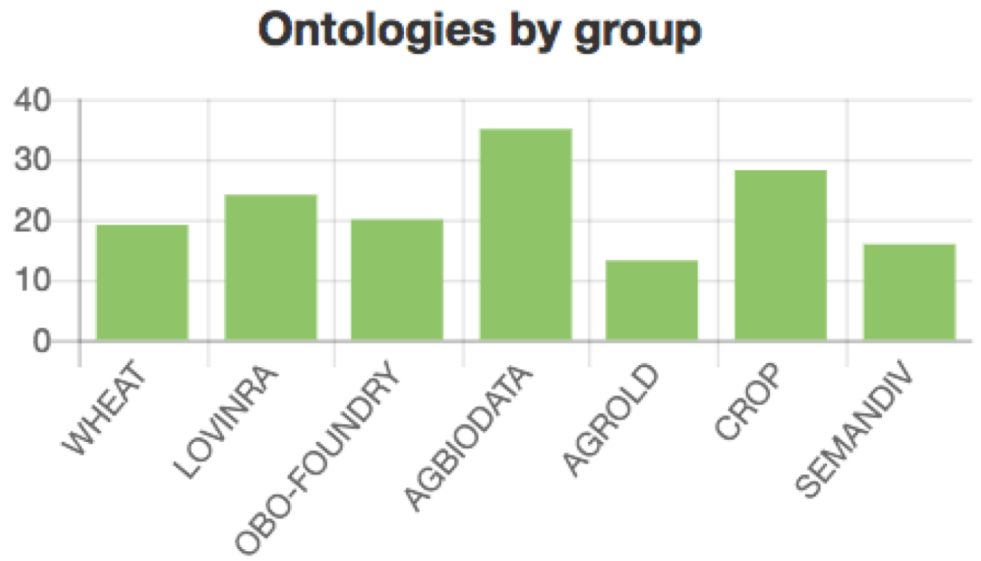
- RDA Agrismeantics WG
- GO-FAIR Food System Implementation Network
- D2KAB ANR project
- Adoptions by projects e.g. PHIS, AgroLD
- SemanDiv CNRS WG
- AgroHackathons
- Maybe:
  - IC-FOODS initiative
  - ELIXIR F&N community

Examples of ontologies uploaded in AgroPortal

Title	Format	Groups	Size
IBP Rice Trait Ontology (CO_320)	OWL	CROP, RICE	~2K
IBP Wheat Trait Ontology (CO_321)	OWL	CROP, WHEAT	~1K
IBP Wheat Anatomy Ontology (CO_121)	OBO	CROP, WHEAT	~80
IBP Crop Research (CO_715)	OBO	CROP	~250
Multi-Crop Passport Ontology (CO_020)	OBO	CROP	~90
Biorefinery (BIOREFINERY)	OWL	LOVINRA	~300
Matter Transfer (TRANSMAT)	OWL	LOVINRA	~1.1K
Plant Ontology (PO)	OWL	WHEAT, RICE, OBOF	~2K
Plant Trait Ontology (TO)	OWL	WHEAT, RICE, OBOF	~4.4K
Durum Wheat (DURUM_WHEAT)	OWL	LOVINRA	~130
Agricultural Experiments (AEO)	OWL	LOVINRA	~60
Environment Ontology (ENVO)	OWL	WHEAT, OBOF	~6.3K
NCBI Organismal Classification (NCBITAXON)	RRF	WHEAT	~900K
AnaEE Thesaurus (ANAEE)	SKOS	LOVINRA	~3.3K
French Crop Usage (CROPUSAGE)	SKOS	none	~300
Agrovoc (AGROVOC)	SKOS	none	~32K
Food Ontology (FOODON)	OWL	OBOF	~10K
National Agriculture Library Thesaurus (NALT)	SKOS	none	~67K
Global Agricultural Concept Scheme (GACS)	SKOS	none	~585K

Ontology groups and categories

Category	Number
Plant Phenotypes and Traits	31
Plant Anatomy and Development	4
Natural Resources, Earth and Environment	12
Animal Science and Animal Products	6
Agricultural Research, Technology and Engineering	15
Breeding and Genetic Improvement	1
Plant Science and Plant Products	7
Plant Genetic Resources	2
Food and Human Nutrition	7
Food Security	2
Taxonomic Classifications of Organisms	6
Farms and Farming Systems	5
Fisheries and Aquaculture	2
Forest Science and Forest Products	2
Biodiversity and Ecology	14



Specific slices display to use only the ontologies of a group

http://crop.agroportal.lirmm.fr

http://inrae.agroportal.lirmm.fr/

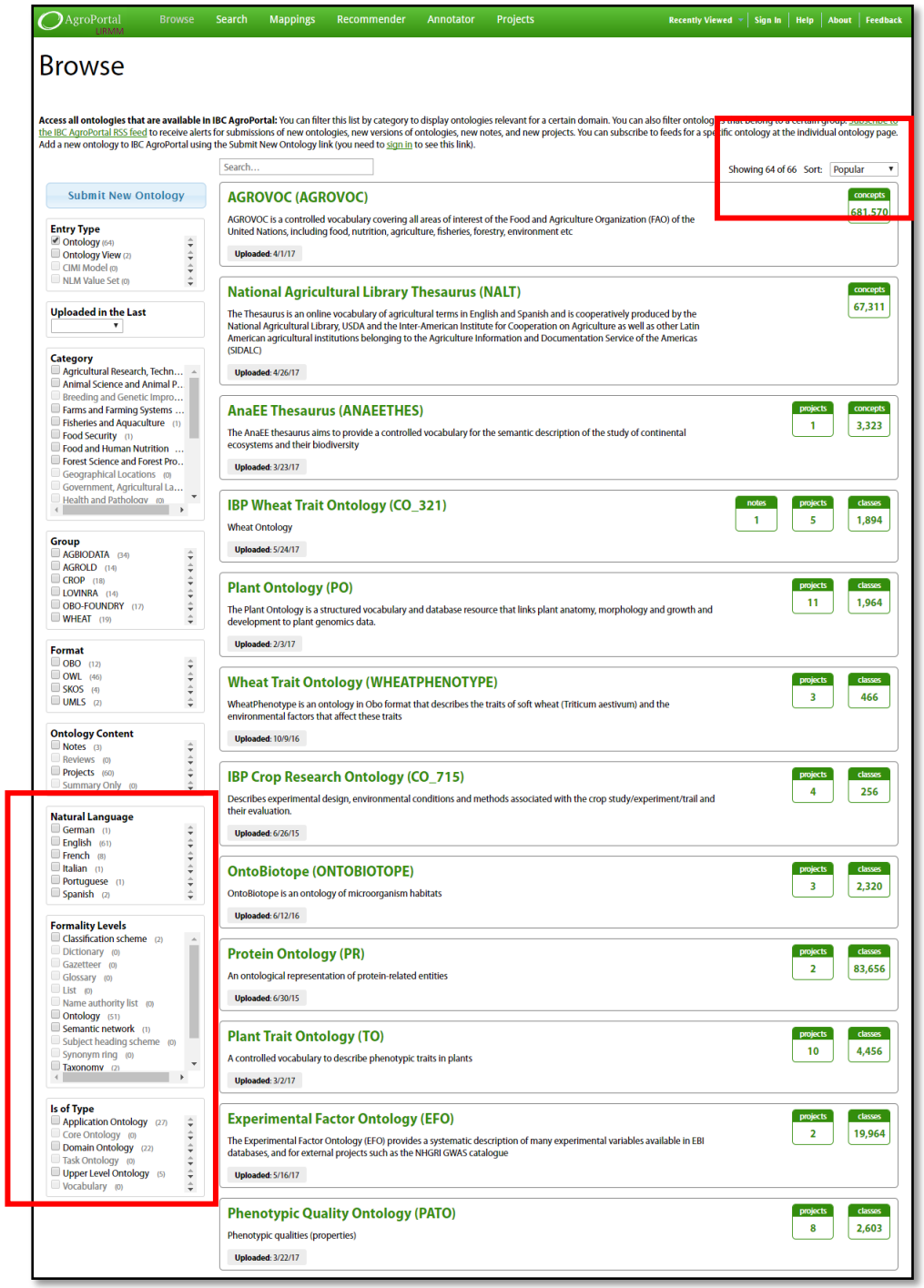
Describe ontologies with semantic metadata

- 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

Browse and select ontologies

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

- Filter ontologies in the list, sort the list (by name, released date, number of views).



Identify concepts to describe your data

AgroPortal LIRMM — Browse | Search | Mappings | Recommender | Annotator | Projects | Landscape | Recently Viewed | Sign In | Help | About | Feedback

Sustainable Development Goals Interface Ontology

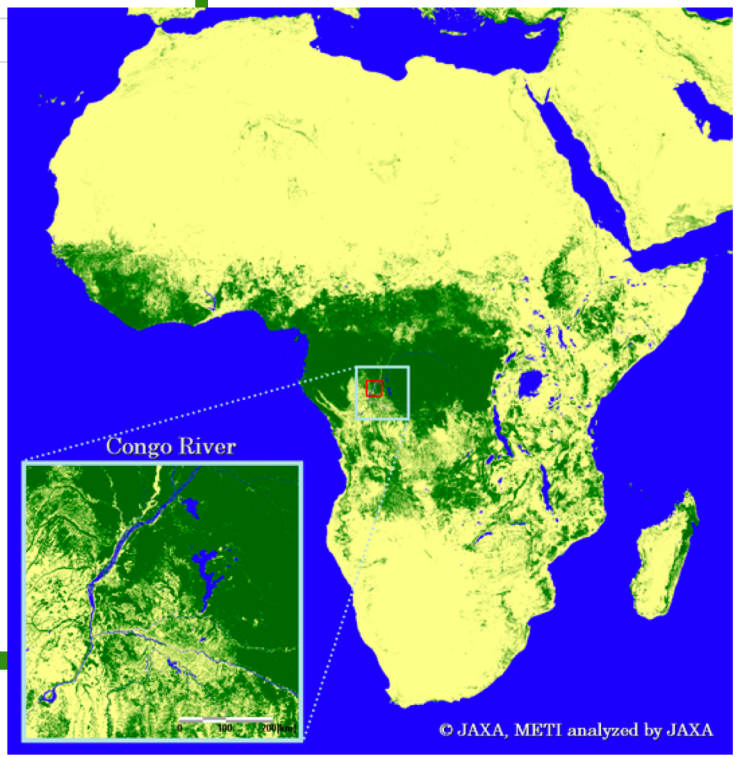
Summary | Classes | Properties | Notes | Mappings | Widgets

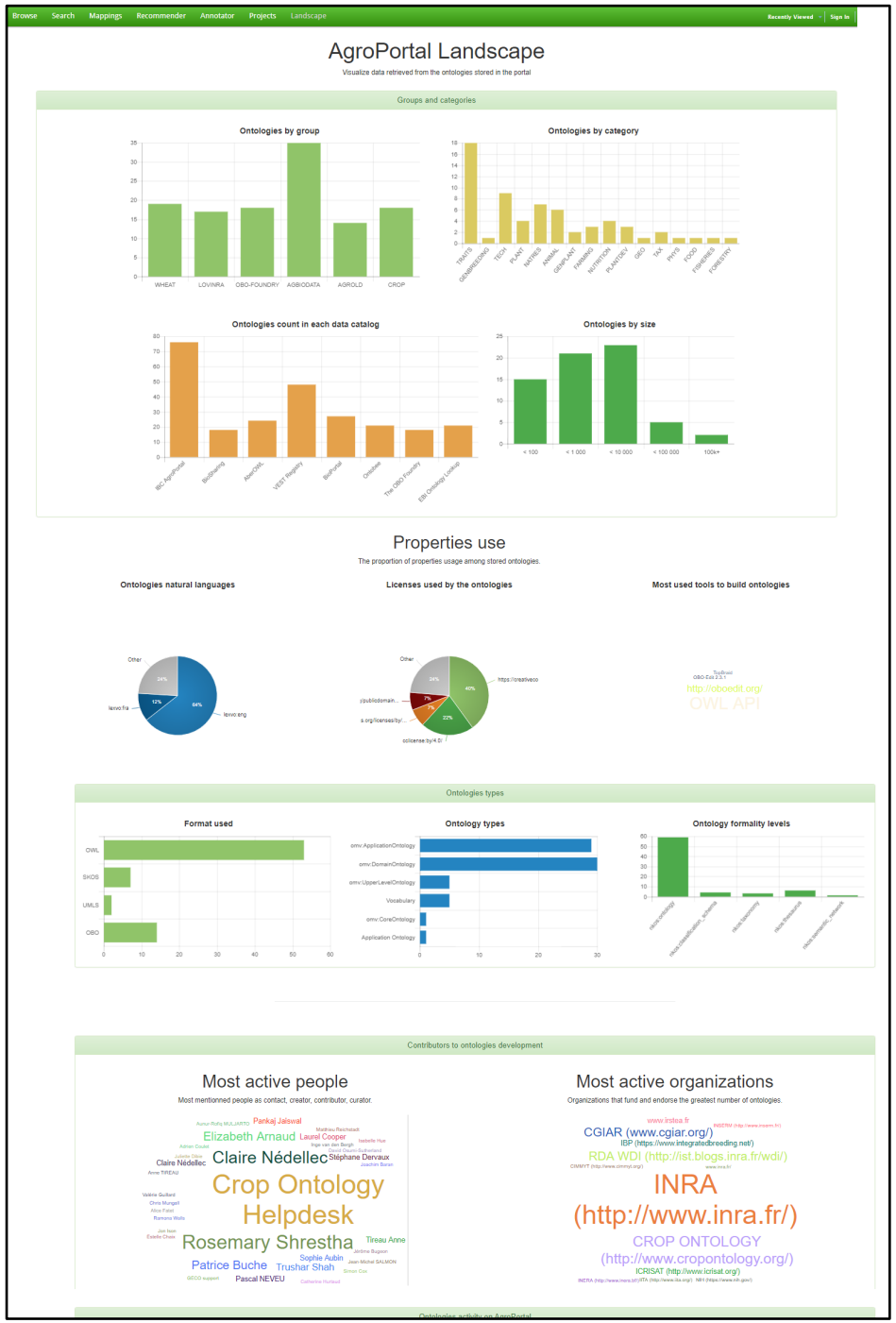
Jump To:

- entity
  - continuant
    - generically dependent continuant
      - information content entity
        - conclusion based on data
        - data item
          - data about an ontology part
          - data set
          - direct material input
          - energy intensity
          - material footprint
          - material footprint per capita
          - material input per capita
          - material intensity
          - material trade balance
          - material trade balance per capita
          - measurement datum
          - net permanent forest loss
          - resource consumption
          - sustainable development goal indicator value
            - Agricultural export subsidies
            - Aid for Trade commitments and disbursements
            - Amount of water- and sanitation-related official developr
            - Average marine acidity (pH) measured at agreed suite of
            - Change in the extent of water-related ecosystems over ti
            - Change in water-use efficiency over time
            - CO2 emission per unit of value added
            - Coverage by protected areas of important sites for mount
            - Death rate due to road traffic injuries
            - Debt service as a proportion of exports of goods and servi
            - Developing countries' and least developed countries' shar
            - Dollar value of financial and technical assistance (includir
            - Extent of use of country-owned results frameworks and pl
            - Extent to which (i) global citizenship education and (ii) ec
            - Financial Soundness Indicators
            - **Forest area as a proportion of total land area**
            - Global food loss index

Details | Visualization | Notes (0) | Class Mappings (0) | Access Class JSON

Preferred Name	Forest area as a proportion of total land area
ID	http://purl.unep.org/sdg/SDGIO_00020174
label	Forest area as a proportion of total land area
prefixIRI	sdg:SDGIO_00020174
prefLabel	Forest area as a proportion of total land area
UN SDG Indicator ID	15.1.1
UNSD SDG indicator code	C150101
subClassOf	sustainable development goal indicator value





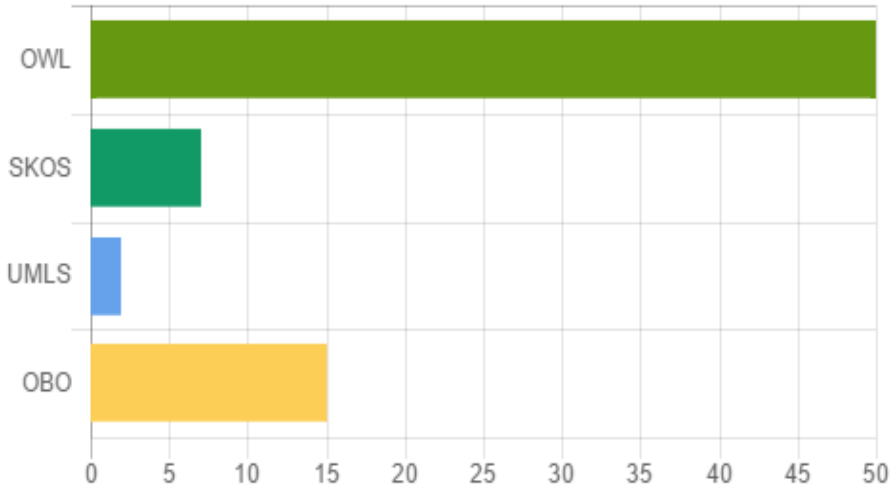
AgroPortal landscape page

- Display “per property”
  - Global presentation of the properties
  - Synthesis diagrams & listing

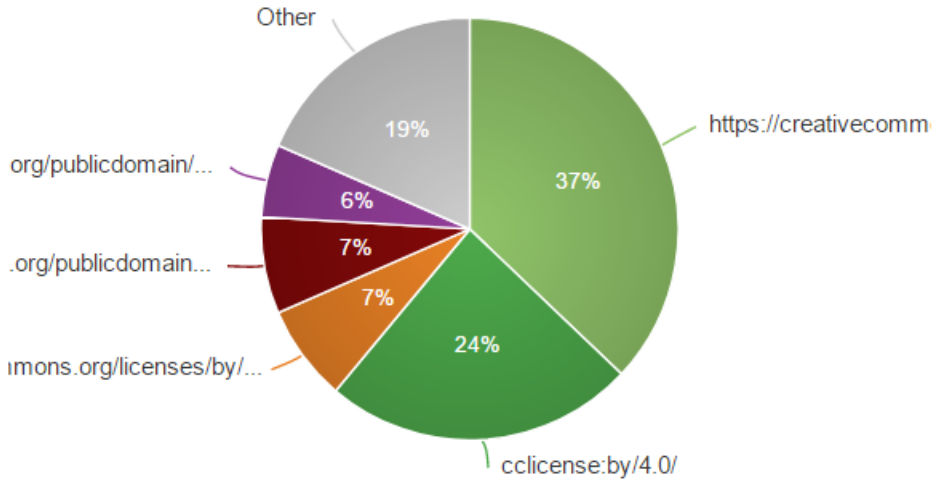
- Allows to explore the agronomical ontology landscape by automatically aggregating the metadata fields of each ontologies in explicit visualizations (charts, term cloud and graphs).

Clement Jonquet, Anne Toulet, Biswanath Dutta, Vincent Emonet. Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal. *Journal on Data Semantics*, Springer, 2018, pp.1-31.

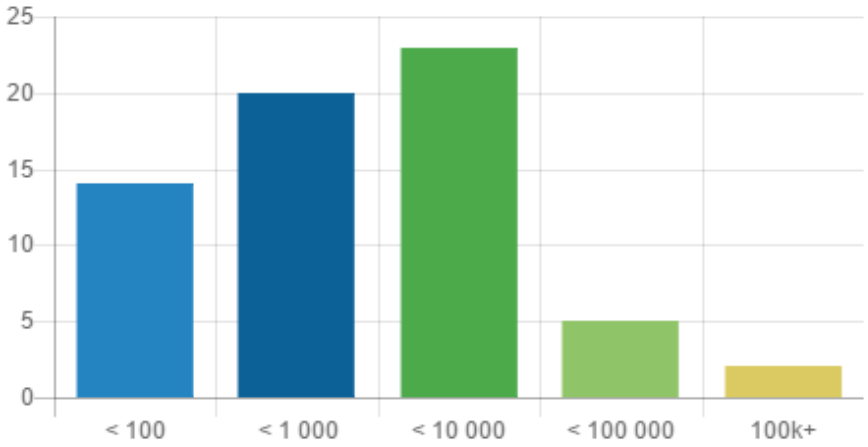
Format



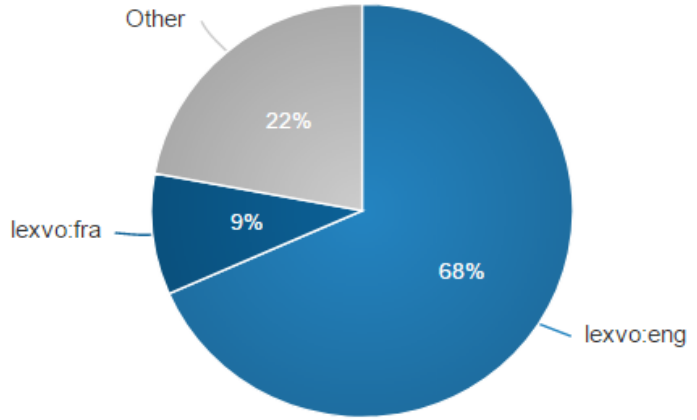
License



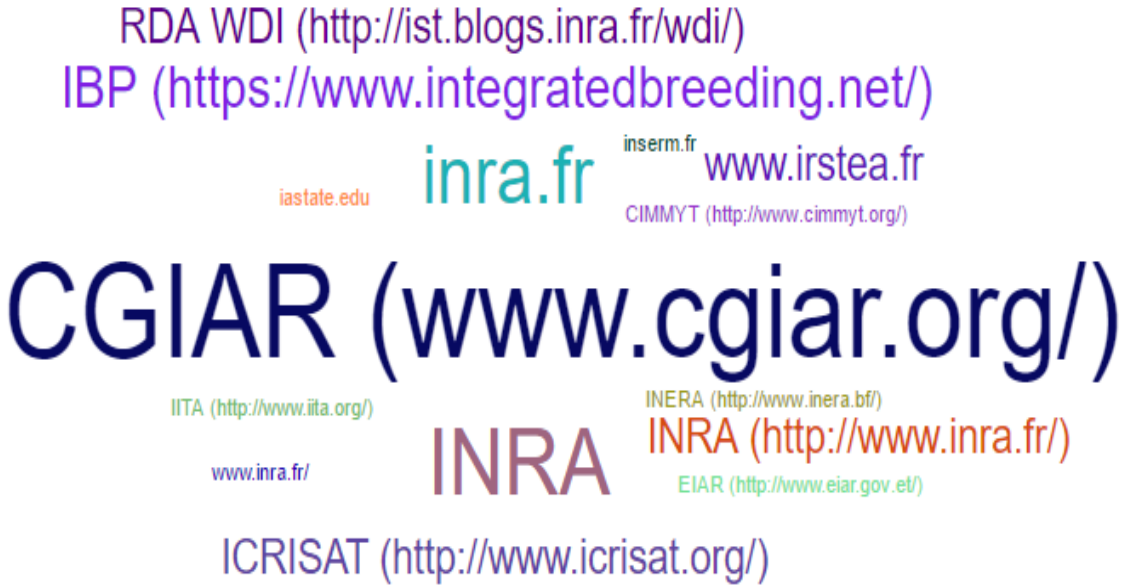
Size

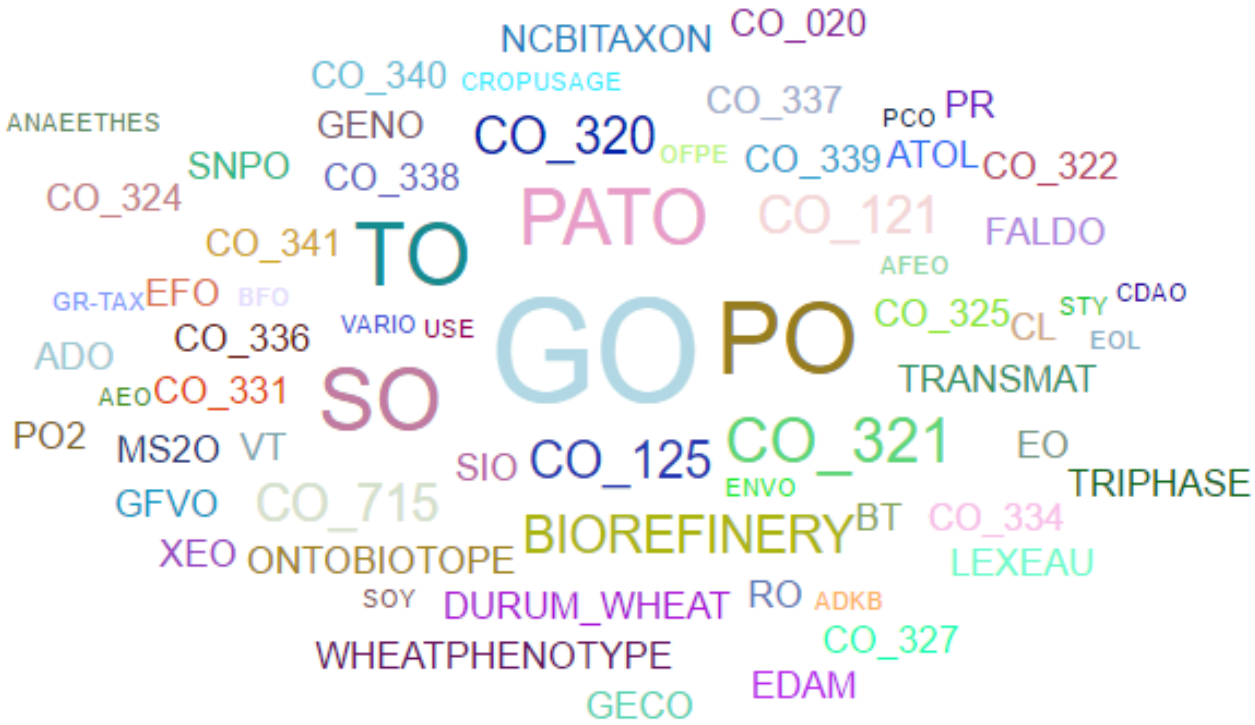


Language

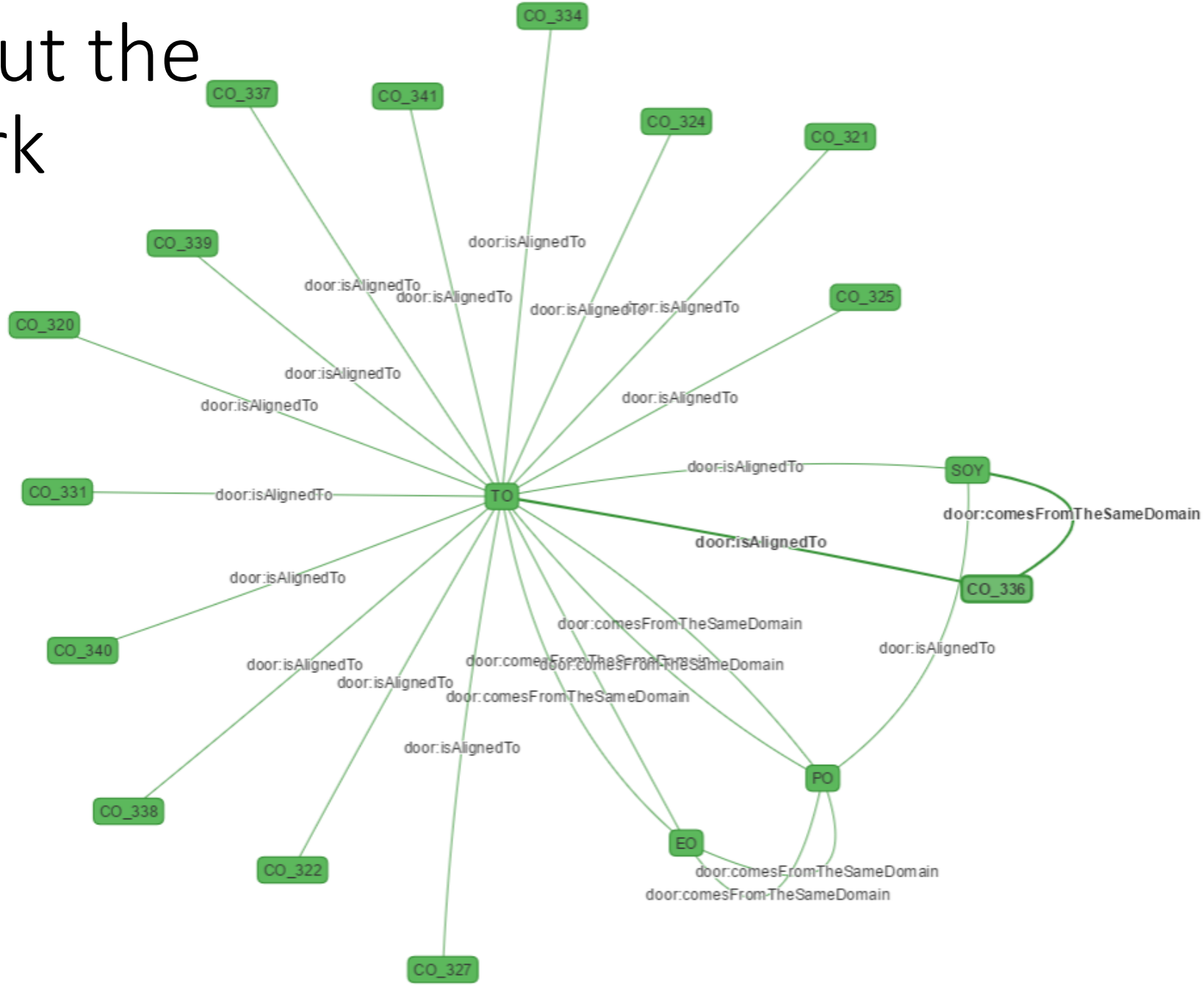


Information about the community





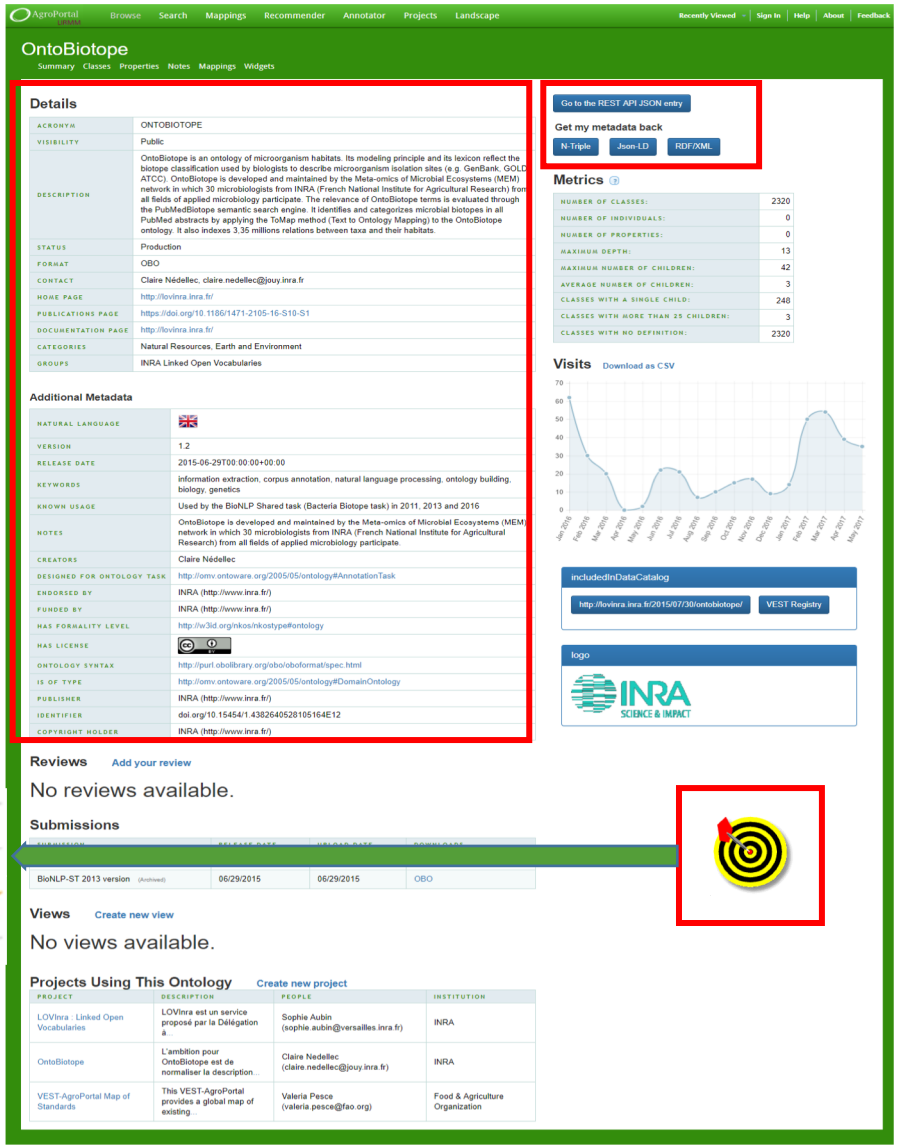
Information about the ontology network



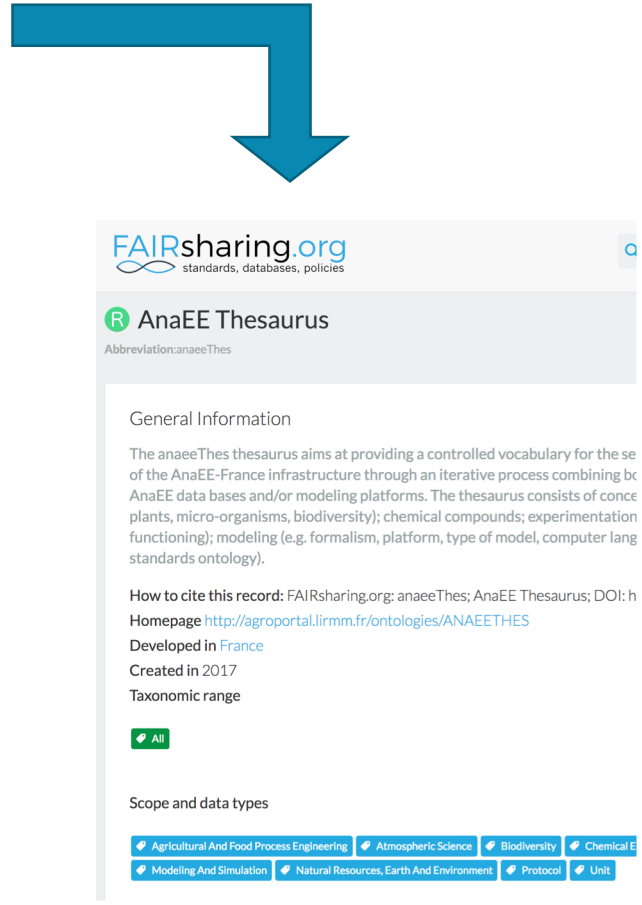
Our objective now: automatic FAIRness assessment of an ontology

1. within AgroPortal
2. outside of AgroPortal

- Enhance the level FAIRness of ontologies.
- Help users respect the I2 FAIR principle.
- Help users in identifying FAIR ontologies.
- Provide useful analysis of the semantic agronomic landscape.

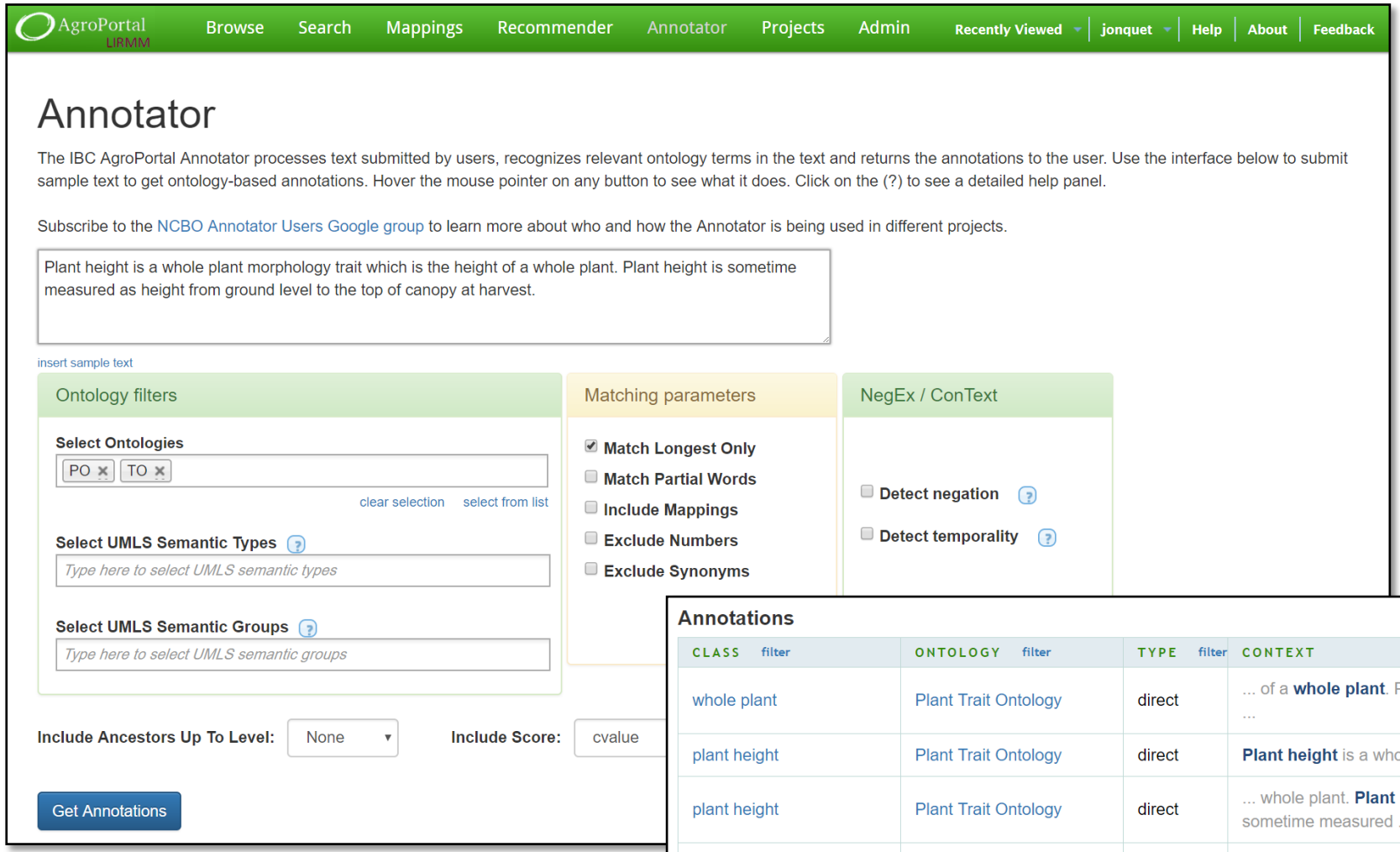


Harvesting AgroPortal ontologies and vocabularies into FAIRsharing



Both manually curate the metadata … better synchronization of the fields to come….

AgroPortal has a new metadata model of 127 properties to describe ontologies & vocabularies



AgroPortal Annotator

identifies ontology concepts within plain text for semantic indexing

Annotations *total results 7 (direct 7 / ancestor 0 / mapping 0)*

CLASS filter	ONTOLOGY filter	TYPE filter	CONTEXT	MATCHED CLASS filter	MATCHED ONTOLOGY filter	SCORE
whole plant	Plant Trait Ontology	direct	... of a **whole plant**. Plant height is ...	whole plant	Plant Trait Ontology	10.000
plant height	Plant Trait Ontology	direct	**Plant height** is a whole ...	plant height	Plant Trait Ontology	8.644
plant height	Plant Trait Ontology	direct	... whole plant. **Plant height** is sometime measured ...	plant height	Plant Trait Ontology	8.644
whole plant morphology trait	Plant Trait Ontology	direct	... is a **whole plant morphology trait** which is the ...	whole plant morphology trait	Plant Trait Ontology	6.644
whole plant	Plant Ontology	direct	... of a **whole plant**. Plant height is ...	whole plant	Plant Ontology	6.644
height	Plant Trait Ontology	direct	... is the **height** of a whole ...	height	Plant Trait Ontology	4.322
height	Plant Trait Ontology	direct	... measured as **height** from ground level ...	height	Plant Trait Ontology	4.322

Format Results As: JSON

Sprouting
Initial Vigor
Color of unexpanded apical leaves
Color of first fully expanded leaf
Leaf vein color
Apical Pubescence
Length of stipules
Number of leaf lobes
Leaf lobe position
Angle of petiole insertion
Petiole length
Petiole color
Anthocyanin pigmentation
Growth habit of young stem
Pubescence of young stem
Stem color
Leaf scar prominence
Apical branching
Branching levels
Branching Angle
Height of first apical branch
Height of plant
Total fresh weight foliage and stems
Total fresh weight foliage and stems
Number harvested

Root number
Fresh weight of storage root
Fresh root yield
Dry yield
Harvest index
Proportion of lodged plants
Leaf retention
Plant architecture
Flowers (50%)
Sepal Color
Disc Color
Sigma color
Ovary color
Anther color
Female stamenoids
Male Sterile
Days to Flower
Fruit set
Fruit Exocarp
Ploidy
Seed oclor

AgroPortal LIRMM Browse Search Mappings

Annotator

The IBC AgroPortal Annotator processes text submitted by user
on any button to see what it does. Click on the (?) to see a detai

Subscribe to the NCBO Annotator Users Google group to learn

Plant architecture
Flowers (50%)
Sepal Color
Disc Color

Cassava Trait Ontology

Ontology filters

Select Ontologies

CO_334 ×

clear selection select from list


```
- {
  - annotatedClass: {
      @id: "http://www.cropontology.org/rdf/CO_334:0000386",
      @type: "http://www.w3.org/2002/07/owl#Class"
    },
    hierarchy: [ ],
  - annotations: [
    - {
        from: 11,
        to: 23,
        matchType: "PREF",
        text: "INITIAL VIGOR"
      }
    ],
```

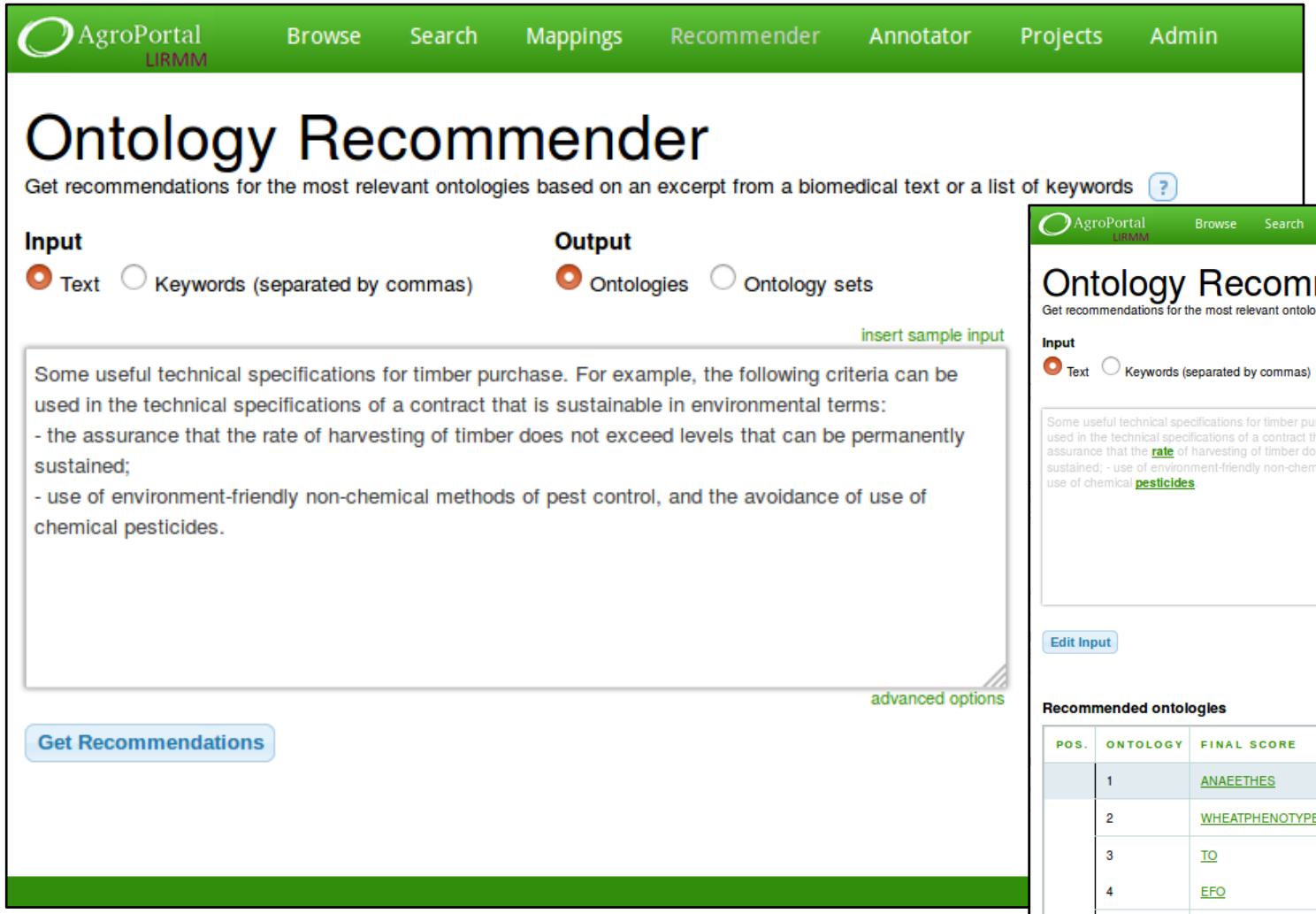

Cassava Trait Ontology

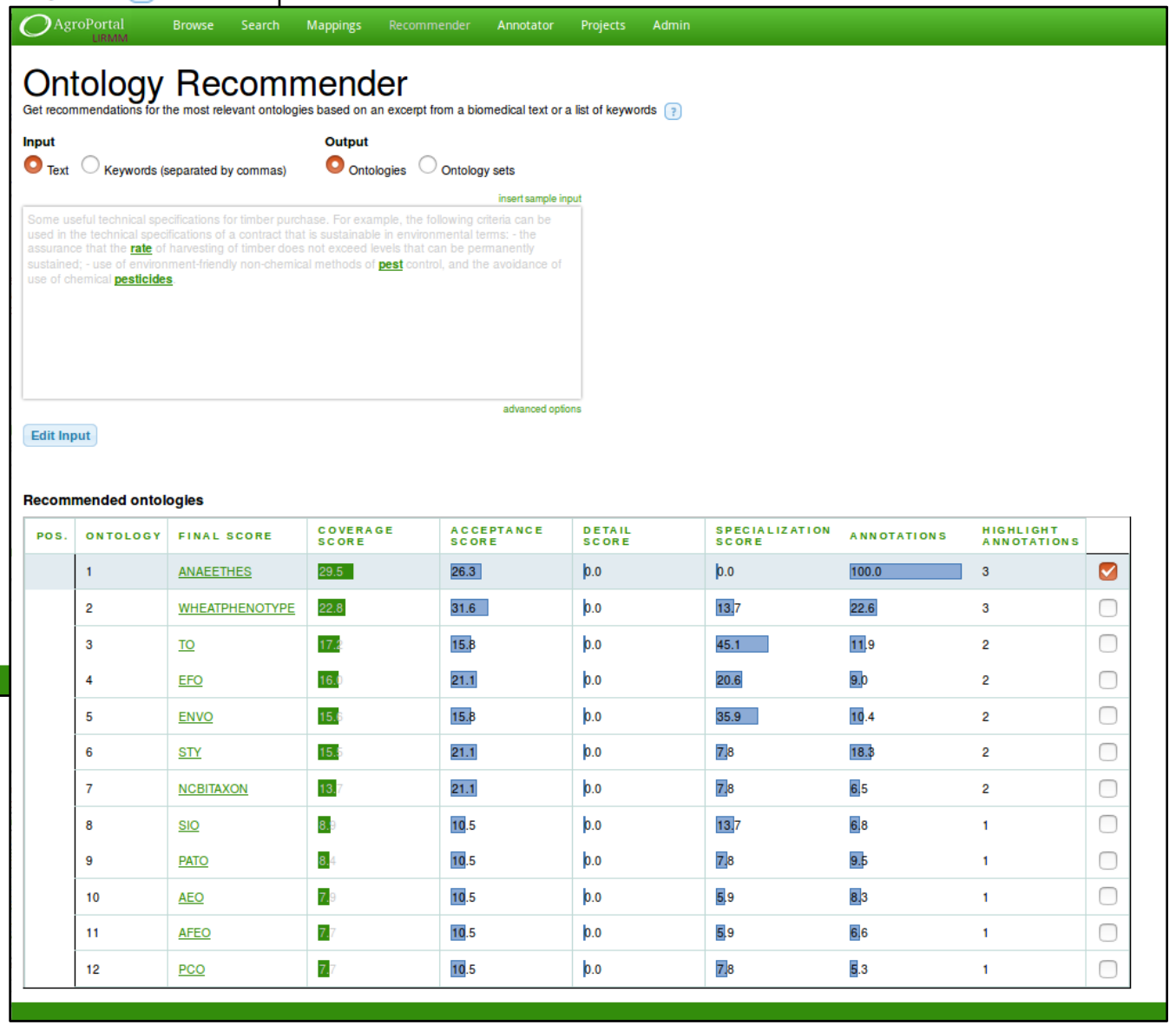
Summary Classes Properties Notes Mappings Widgets

Jump To:

- Cassava trait
  - Agronomical trait
    - Anthocyanin Pigmentation
    - Ease of Harvest
    - Female Stamenoids
    - Fresh Shoot Weight
    - Fruit Exocarp Texture
    - Fruit set presence
    - Initial Vigor
    - Leaf weight
    - Male Sterile
    - Marketable root number

Preferred Name	Initial Vigor
Synonyms	Initial plant vigor
Definitions	Initial plant vigor at one month after planting





POS.	ONTOLOGY	FINAL SCORE	COVERAGE SCORE	ACCEPTANCE SCORE	DETAIL SCORE	SPECIALIZATION SCORE	ANNOTATIONS	HIGHLIGHT ANNOTATIONS	
	1	ANAEETHES	29.5	26.3	0.0	0.0	100.0	3	☑
	2	WHEATPHENOTYPE	22.8	31.6	0.0	13.7	22.6	3	☐
	3	TO	17.4	15.8	0.0	45.1	11.9	2	☐
	4	EFO	16.0	21.1	0.0	20.6	9.0	2	☐
	5	ENVO	15.5	15.8	0.0	35.9	10.4	2	☐
	6	STY	15.5	21.1	0.0	7.8	18.3	2	☐
	7	NCBITAXON	13.7	21.1	0.0	7.8	6.5	2	☐
	8	SIO	8.9	10.5	0.0	13.7	6.8	1	☐
	9	PATO	8.4	10.5	0.0	7.8	9.5	1	☐
	10	AEO	7.9	10.5	0.0	5.9	8.3	1	☐
	11	AFEO	7.7	10.5	0.0	5.9	6.6	1	☐
	12	PCO	7.7	10.5	0.0	7.8	5.3	1	☐

AgroPortal Recommender

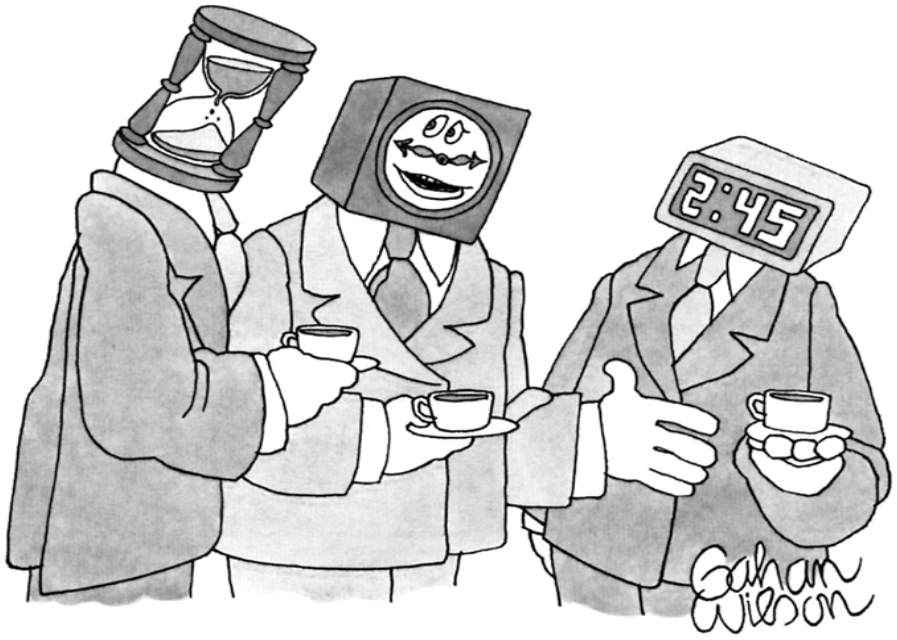
get the most relevant ontologies for your data

Ontology alignment

- Ontologies, vocabularies, and terminologies inevitably overlap in coverage

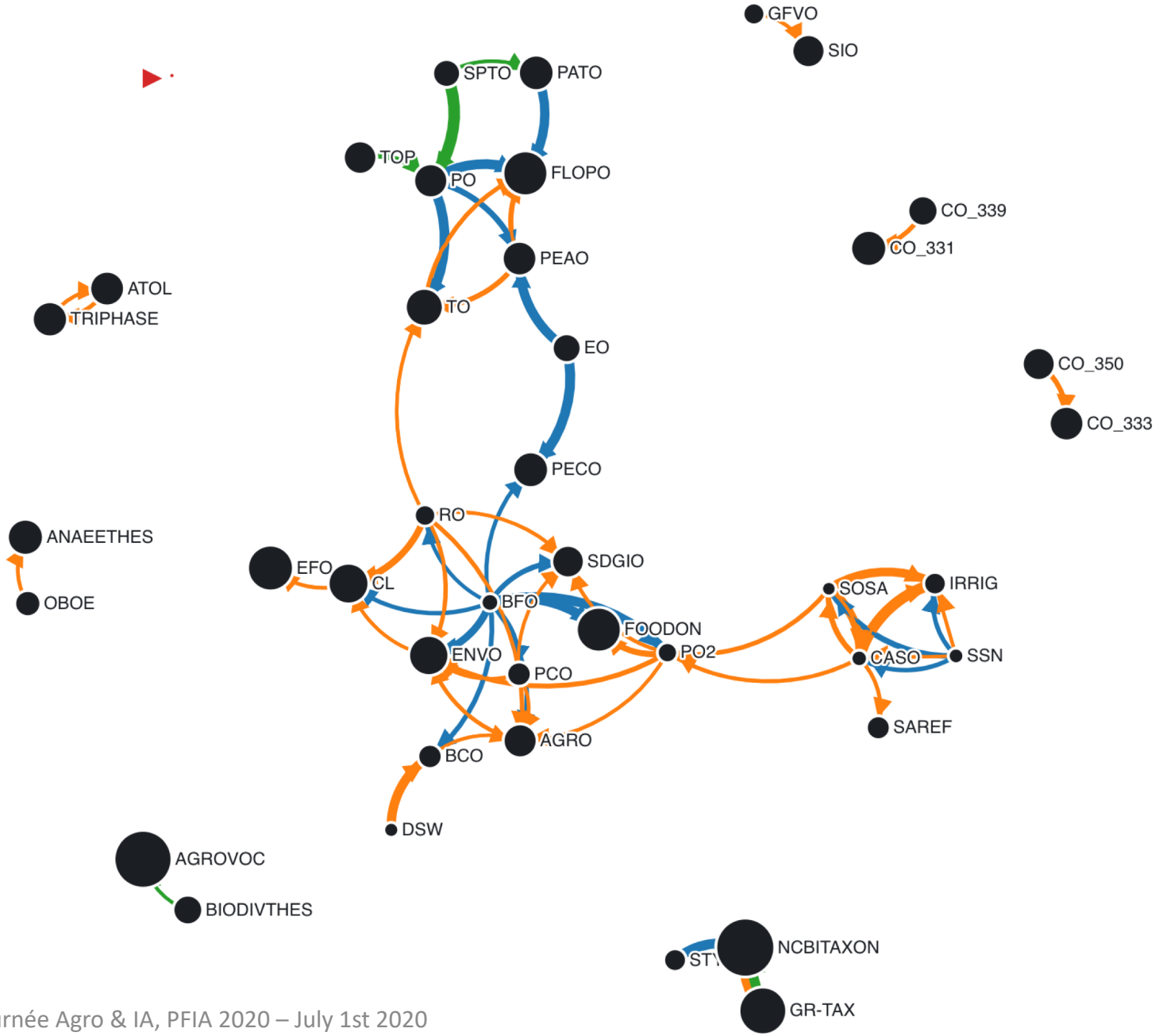
- Mappings do not always belong to an ontology
  - The community needs a place to store and retrieve them
  - That's the role of the ontology repository

- Dealing with mappings is a technical, data and scientific challenge
  - Capture the whole mapping lifecycle
  - Semantically described with plenty of provenance information

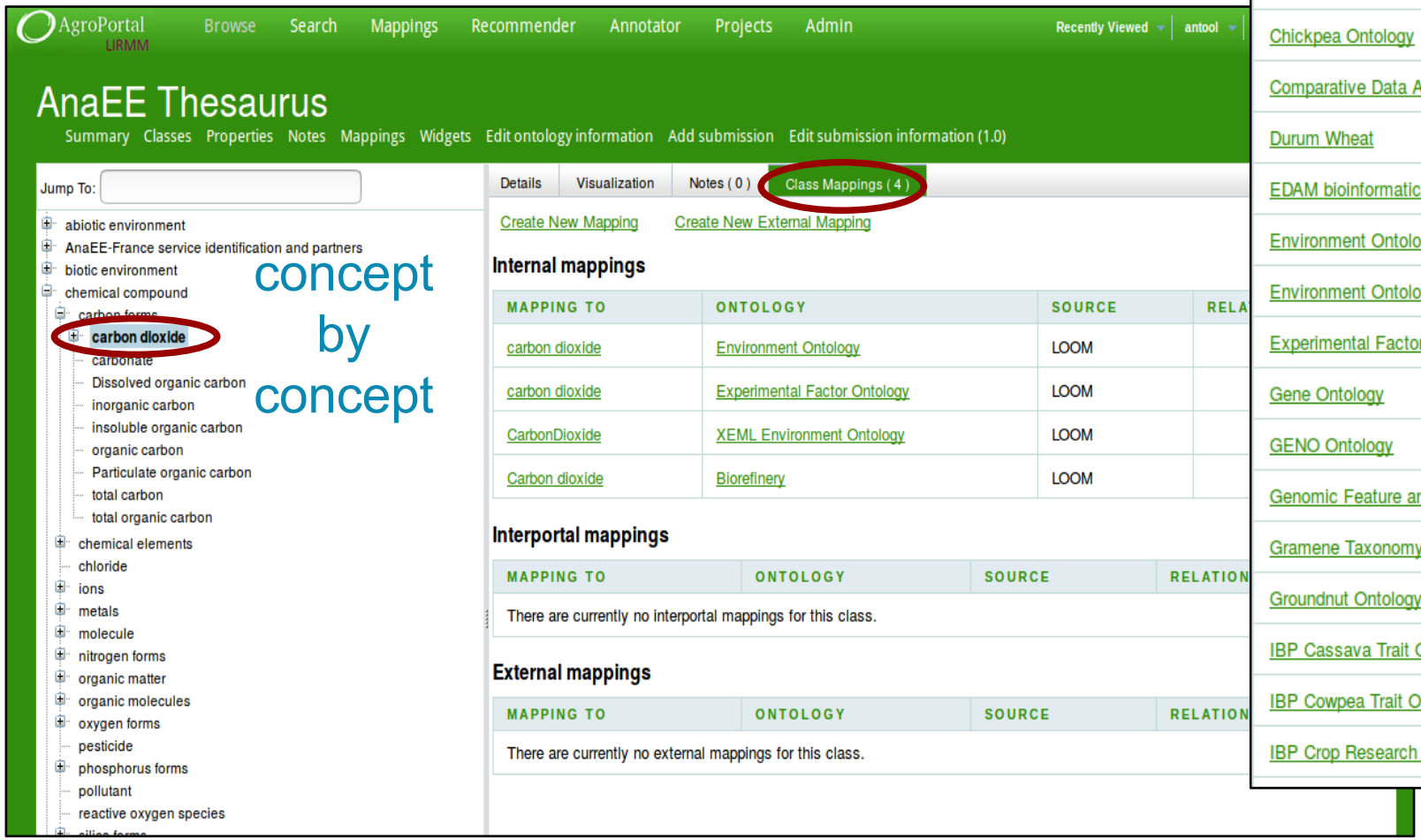


"Basically, we're all trying to say the same thing."

Term reuse, term overlap and extracted mappings



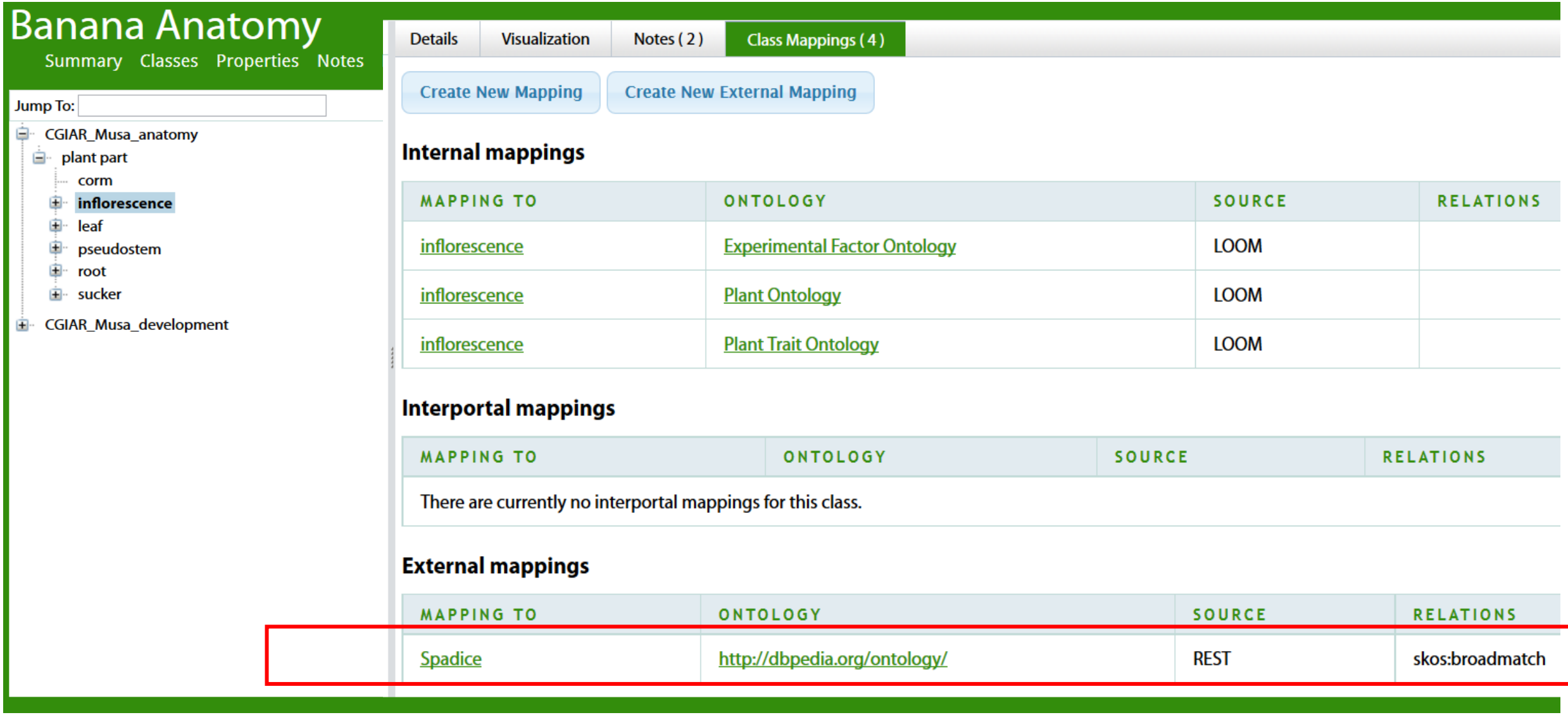
Align ontologies one another



Mappings

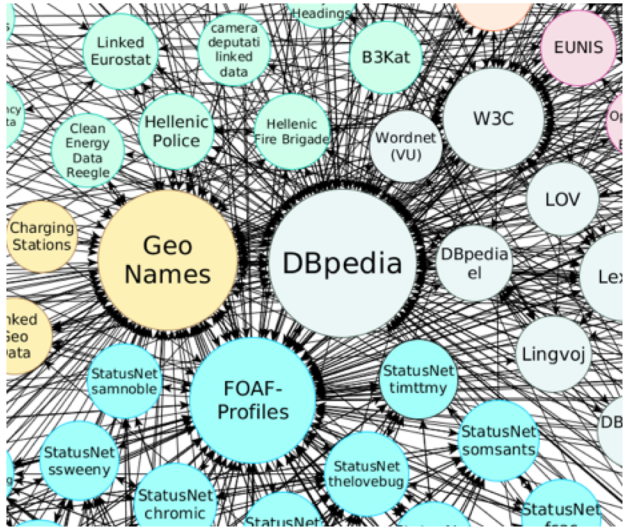
ONTOLOGY	MAPPINGS
Agri-Food Experiment Ontology	1
Agricultural Experiments Ontology	5
Banana Anatomy	2
Basic Formal Ontology	1
Biorefinery	13
Cell Ontology	4
Chickpea Ontology	14
Comparative Data Analysis Ontology	3
Durum Wheat	2
EDAM bioinformatics operations, data types, formats, identifiers and topics	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

Enable to store external mappings
i.e., mappings that only one part is in BioPortal



Mappings to external resources were also extracted

- e.g.,



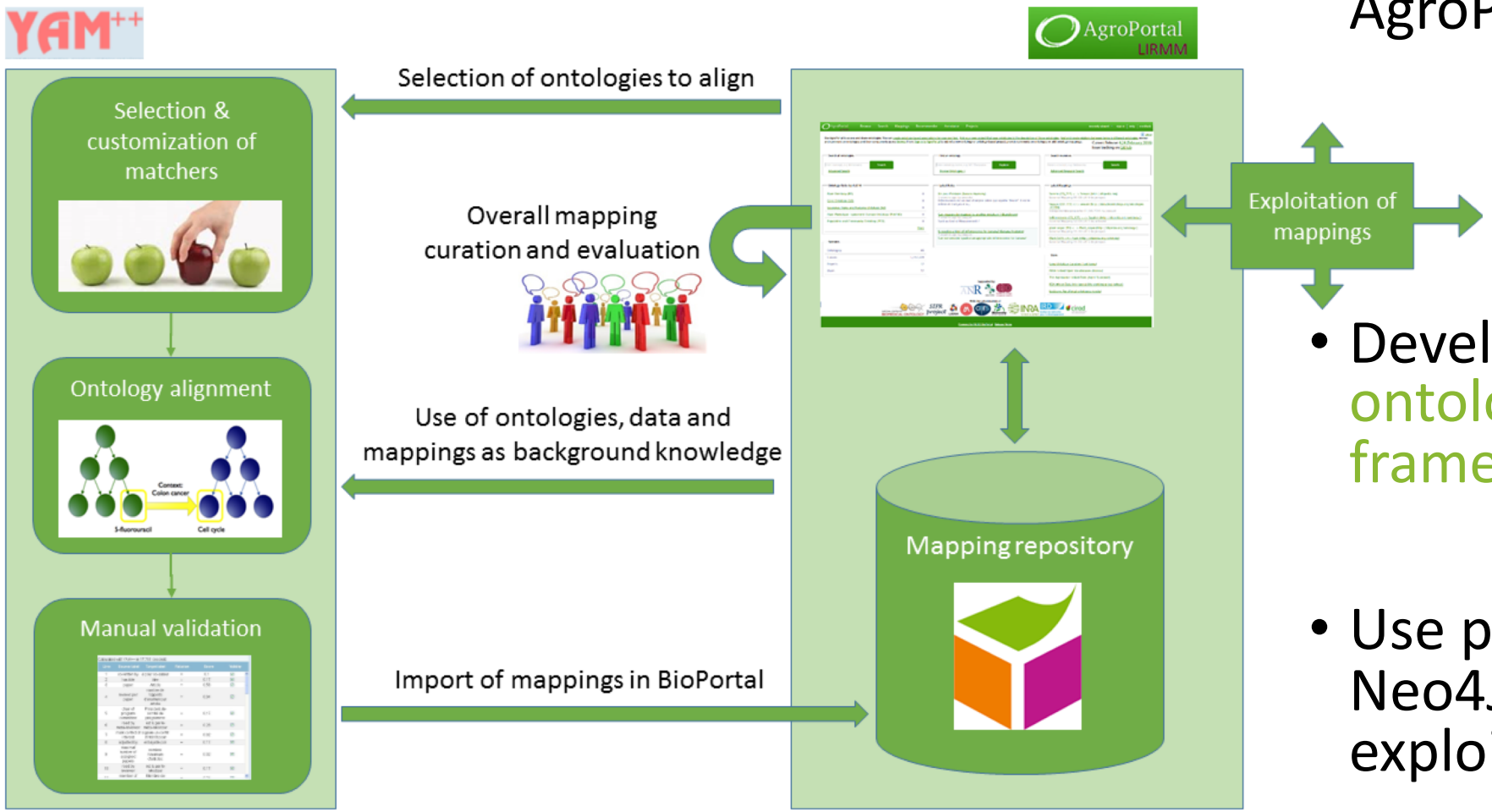
AGROVOC - AGROVOC (84,030)

Table Visualization

ONTOLOGY	MAPPINGS
Agri-Food Experiment Ontology	1
Agricultural and Nutrition Technology Ontology	2
Agriculture and Forestry Ontology	22
AGRORDF	2
AnaEE Thesaurus	333
Animal Disease Ontology	17
Animal Trait Ontology for Livestock	3
Biodiversity Thesaurus	255
Biological Collections Ontology	1
Biorefinery	2
Brachiaria Ontology	1
Brassica Ontology	2
Cassava Trait Ontology	1
Castor Bean Ontology	1
Cell Ontology	19
Chickpea Ontology	2
Common bean Ontology	1
Cowpea Trait Ontology	1
Darwin-SW	1
EDAM bioinformatics operations, data types, formats, identifiers and topics	6
Environment Ontology	29
Experimental Factor Ontology	85
External Mappings	47,809
Flora Phenotype Ontology	5
FoodOn	192

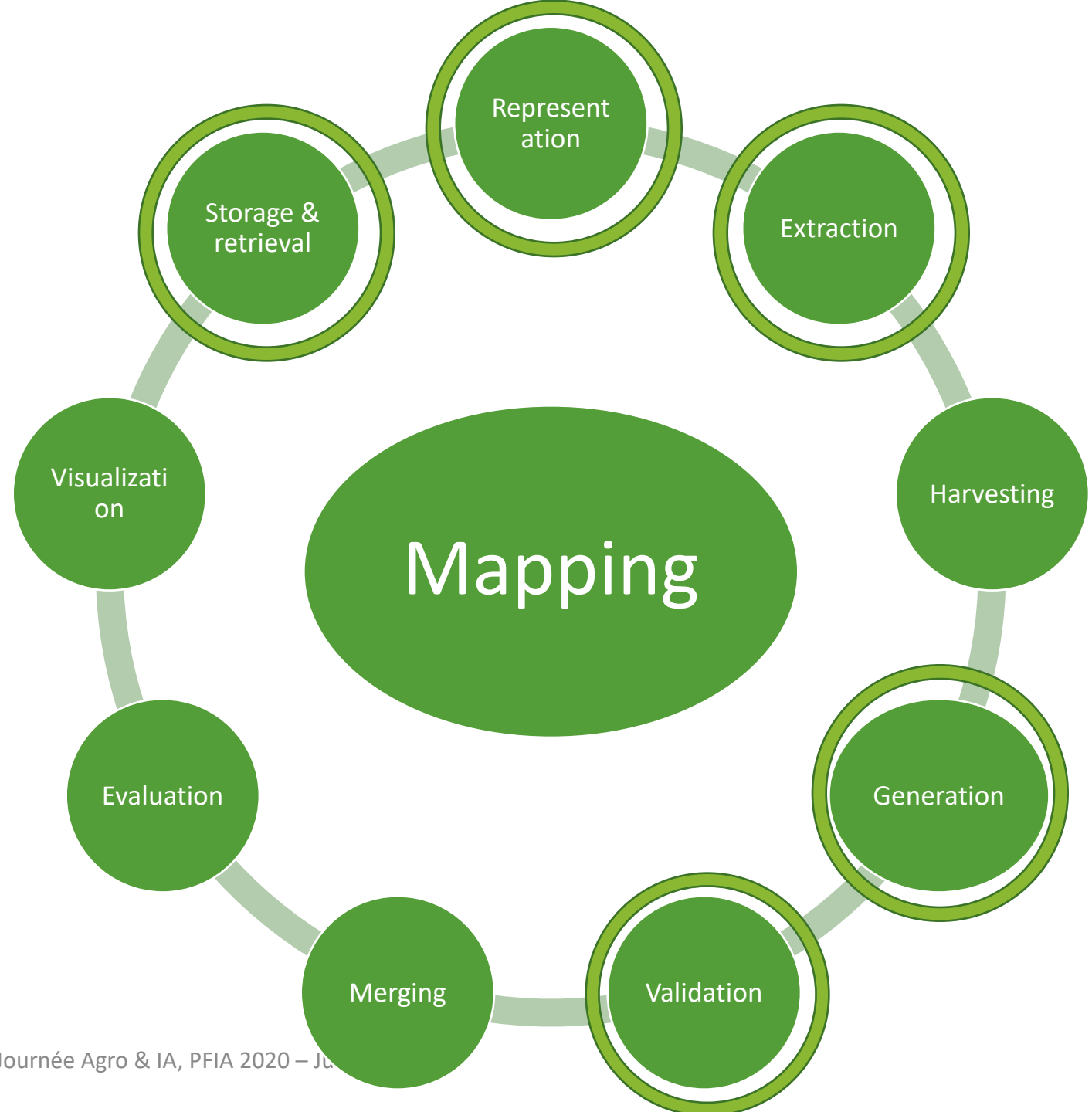
What to do now: analysis and feedback to community to improve the dataset

What's next for ontology alignment?



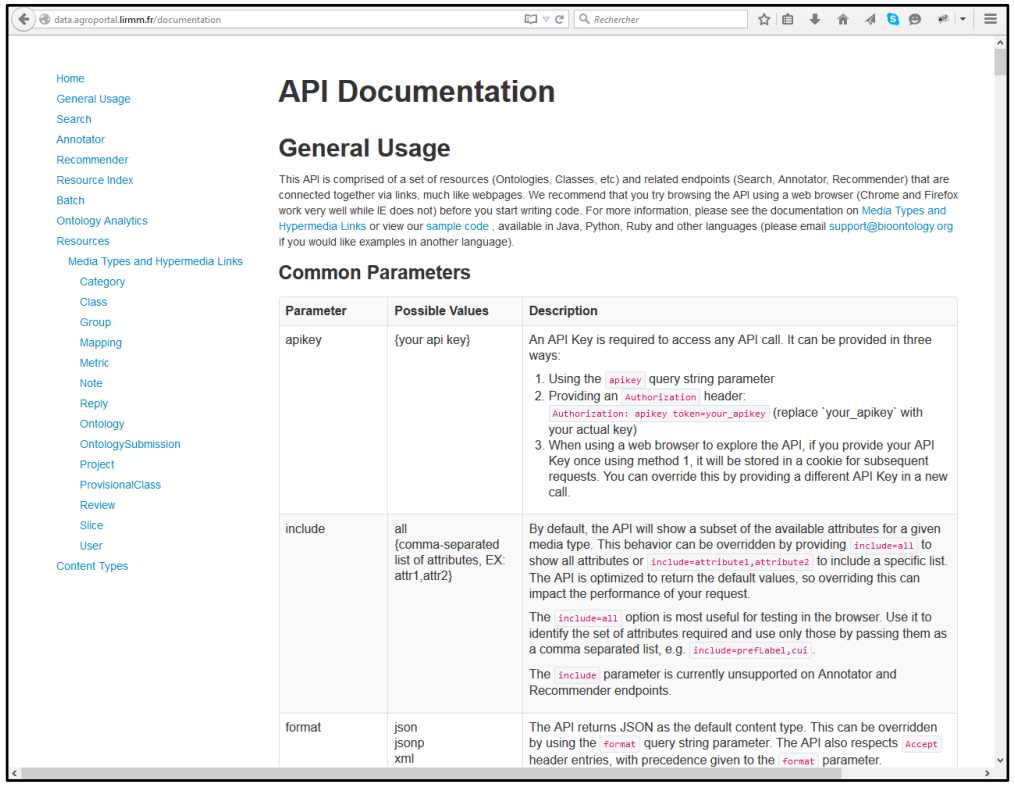
- **Align** the ontologies within AgroPortal
- Developing a state-of-the-art ontology alignment framework (all aspects)
- Use property graphs (e.g., Neo4J) for mapping exploitation

All aspects of ontology alignments



REST Service API:

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



SPARQL endpoint:

http://sparql.agroportal.lirmm.fr

Conclusions

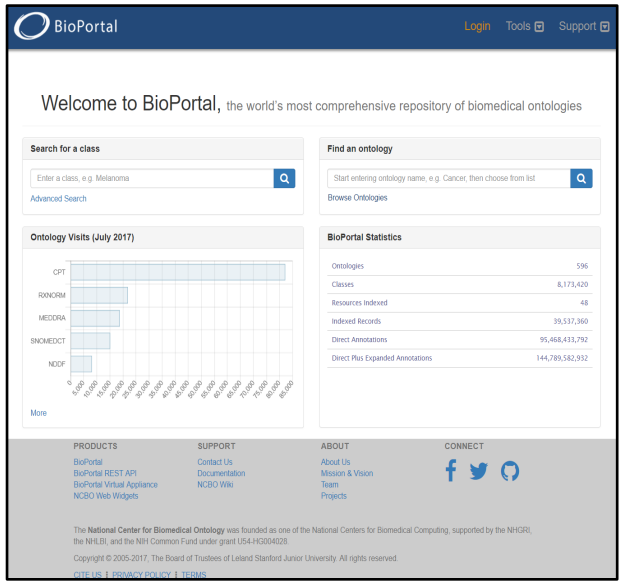
ANR Project D2KAB: Data to Knowledge in Agronomy and Biodiversity (2019-2023)

*Create a framework to **turn agronomy and biodiversity data into knowledge –semantically described, interoperable, actionable, open–** and investigate scientific methods and tools to exploit this knowledge for applications in science & agriculture*

- How: Ontologies & Linked Open Data
  - 1 work-package on building and harnessing knowledge graphs
  - 2 work-packages of driving ag & biodiv projects (food packaging, agro-agri linked data, wheat phenotype, ecosystems & plant biogeography)

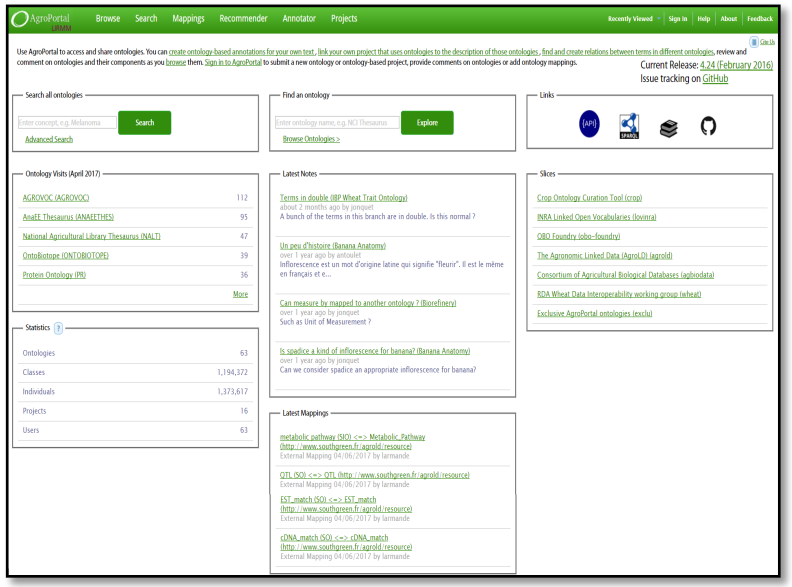
We all develop and maintain ontology repositories in the OntoPortal Alliance

BioPortal



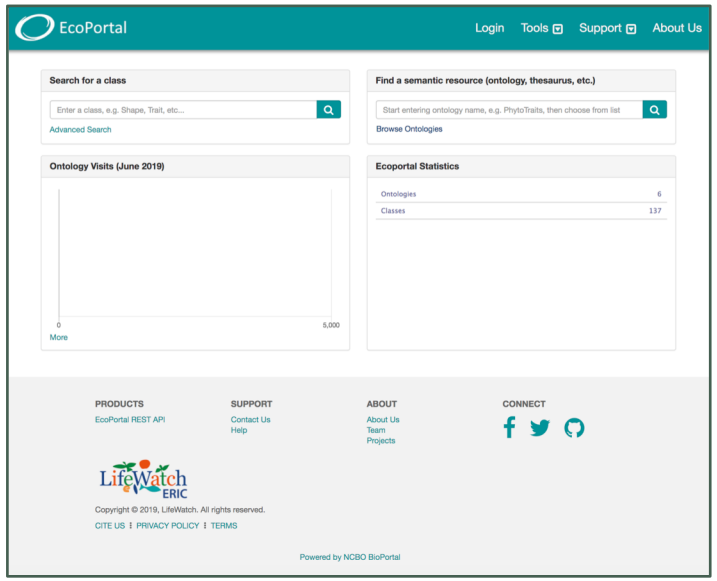
http://bioportal.bioontology.org

AgroPortal



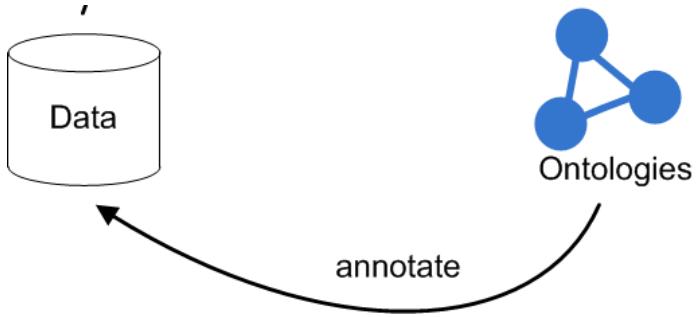
http://agroportal.lirmm.fr

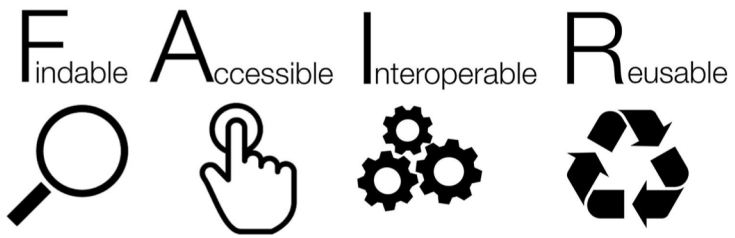
EcoPortal



http://ecoportal.lifewatchitaly.eu

Take home message





Credits (people & support)

- LIRMM
  - Vincent Emonet
  - Anne Toulet
  - Andon Tchechmedjiev
  - Amine Abdaoui
  - Juan-Antonio Lossio
  - Elcio Abrahao
  - Amir Laadhar
  - Emna Amdouni
  - Jerome Lamarque
  - Zohra Bellahsene
  - Amina Annane (ESI Algeria)
  - Mathieu Roche (CIRAD)
  - Sandra Bringay
  - Few MSc students / year

- Collaborators
  - Pierre Larmande (IRD)
  - Mark Musen (Stanford)
  - John Graybeal (Stanford)
  - Stefan Darmoni (CISMEF)
  - Maguelonne Teisseire (IRSTEA)
  - Sebastien Harispe (LGI2P)
  - Adrien Coulet (LORIA)
  - Elizabeth Arnaud (CGIAR)
  - S. Aubin, O. Hologne, E. Dzalé, P. Neveu, C. Pommier, C. Nédellec … (INRAE)
  - ….

Questions ?

Clement Jonquet

jonquet@lirmm.fr