

Extraction, generation and analysis of alignments between ontologies on Agroportal

Elcio Abrahão

elcio.abraha@lirmm.fr

Séminaire réseau IN OVIVE - 25-26 juin 2018 - Montpellier

This presentation:

- Introduction
- Context background
- Context of Mappigns on Agroportal
- Project Phases / Methodology
- Preliminary Results
- Roadmap Summary

Elcio Abrahão

- Post Doc Researcher at Agroportal Project - LIRMM - Ontology Alignment
- PhD in Computer Engineering - Laboratory of Agriculture Automation - Polytechnic School from São Paulo University
- Agronomist Engineer, software developer and data architect
- Research area: ontologies, web semantics and linked data
- Agriculture Operations Task Ontology developer: <http://agropto.org>

Ontology repositories help to make ontologies FAIR

Findable

Accessible

Interoperable

Re-usable

API Documentation

General Usage

This API is comprised of a set of resources (Ontologies, Classes, etc) and related endpoints (Search, Annotator, Recommender) that are connected together via links, much like webpages. We recommend that you try browsing the API using a web browser (Chrome and Firefox work very well while IE does not) before you start writing code. For more information, please see the documentation on Media Types and Hypertext Links or view our sample code, available in Java, Python, Ruby and other languages (please email help@ontobio.org if you would like examples in another language).

Common Parameters

Parameter	Possible Values	Description
apikey	(your api key)	An API Key is required to access any API call. It can be provided in three ways: <ol style="list-style-type: none"> Using the <code>apikey</code> query string parameter Providing an <code>Authorization</code> header: <pre>Authorization: apikey token=your_apikey</pre> (replace <code>your_apikey</code> with your actual key) When using a web browser to explore the API, if you provide your API Key once using method 1, it will be stored in a cookie for subsequent requests. You can

SPARQL httpd server v1.1.5-122-;

KB ontologies_api

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT * WHERE {
  ?s ?p ?o
} LIMIT 10
```

AgroPortal an ontology repository for agronomy, food, plant sciences & biodiversity

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

AgroPortal [Browse](#) [Search](#) [Mappings](#) [Recommender](#) [Annotator](#) [Projects](#) [Recently Viewed](#) [Sign In](#) [Help](#) [About](#) [Feedback](#)

Use AgroPortal to access and share ontologies. You can [create ontology-based annotations for your own text](#), [link your own project that uses ontologies to the description of those ontologies](#), [find and create relations between terms in different ontologies](#), review and comment on ontologies and their components as you [browse](#) them. [Sign in to AgroPortal](#) to submit a new ontology or ontology-based project, provide comments on ontologies or add ontology mappings.

Current Release: **4.24 (February 2016)**
Issue tracking on [GitHub](#)

Search all ontologies

[Advanced Search](#)

Find an ontology

[Browse Ontologies >](#)

Ontology Visits (April 2017)

AGROVOC (AGROVOC)	112
AnaEE Thesaurus (ANAETHES)	95
National Agricultural Library Thesaurus (NALT)	47
OntoBiotope (ONTOBIOTOPE)	39
Protein Ontology (PR)	36
More	

Latest Notes

[Terms in double \(IBP Wheat Trait Ontology\)](#)
about 2 months ago by jonquet
A bunch of the terms in this branch are in double. Is this normal ?

[Un peu d'histoire \(Banana Anatomy\)](#)
over 1 year ago by antoulet
Inflorescence est un mot d'origine latine qui signifie "fleurer". Il est le même en français et e...

[Concepts mapped to another ontology ? \(Biorefinery\)](#)
by jonquet
Measurement ?

[of inflorescence for banana? \(Banana Anatomy\)](#)
by jonquet
spadice an appropriate inflorescence for banana?

[y \(SIO\) <=> Metabolic_Pathway](#)
[hgreen.fr/agrold/resource](#)
04/06/2017 by larmande

[E1 \(http://www.southgreen.fr/agrold/resource\)](#)
04/06/2017 by larmande

[=> EST_match](#)
[hgreen.fr/agrold/resource](#)
04/06/2017 by larmande

[Y<=> cDNA_match](#)
[hgreen.fr/agrold/resource](#)
04/06/2017 by larmande

Slices

- [Crop Ontology Curation Tool \(crop\)](#)
- [INRA Linked Open Vocabularies \(lovinra\)](#)
- [OBO Foundry \(obo-foundry\)](#)
- [The Agronomic Linked Data \(AgroLD\) \(agrold\)](#)
- [Consortium of Agricultural Biological Databases \(agbiodata\)](#)
- [RDA Wheat Data Interoperability working group \(wheat\)](#)
- [Exclusive AgroPortal ontologies \(exclu\)](#)

AgroPortal [Browse](#) [Search](#) [Mappings](#) [Recommender](#) [Annotator](#) [Projects](#) [Recently Viewed](#) [Sign In](#) [Help](#) [About](#) [Feedback](#)

Browse

Access all ontologies that are available in IBC AgroPortal: You can filter this list by category to display ontologies relevant for a certain domain. You can also filter ontologies that belong to a certain group. [Subscribe to the IBC AgroPortal RSS feed](#) to receive alerts for submissions of new ontologies, new versions of ontologies, new notes, and new projects. You can subscribe to feeds for a specific ontology page. Add a new ontology to IBC AgroPortal using the Submit New Ontology link (you need to [sign in](#) to see this link).

Search... Showing 63 of 65 Sort: **Popular**

[Submit New Ontology](#)

Entry Type

- Ontology (63)
- Ontology View (2)
- CIMI Model (0)
- NLM Value Set (0)

Uploaded in the Last

Category

- Agricultural Research, Techn...
- Animal Science and Animal P...
- Breeding and Genetic Improv...
- Farms and Farming Systems ...
- Fisheries and Aquaculture (1)
- Food Security (1)
- Food and Human Nutrition (4)
- Forest Science and Forest Pro...
- Geographical Locations (0)
- Government, Agricultural Lab...
- Health and Pathology (0)

Group

- AGBIODATA (3)
- AGROLD (14)
- CROP (18)
- LOVINRA (14)
- OBO-FOUNDRY (17)
- WHEAT (19)

Format

- OBO (13)
- OWL (44)
- SKOS (4)
- UMLS (2)

Ontology Content

- Notes (3)
- Reviews (0)
- Projects (57)
- Summary Only (0)

Natural Language

- German (1)
- English (58)
- French (0)
- Italian (1)
- Portuguese (1)
- Spanish (2)

AGROVOC (AGROVOC) concepts 681,570

AGROVOC is a controlled vocabulary covering all areas of interest of the Food and Agriculture Organization (FAO) of the United Nations, including food, nutrition, agriculture, fisheries, forestry, environment etc.

Uploaded: 3/31/17

AnaEE Thesaurus (ANAETHES) projects 1 concepts 3,323

The AnaEE thesaurus aims to provide a controlled vocabulary for the semantic description of the study of continental ecosystems and their biodiversity

Uploaded: 3/23/17

National Agricultural Library Thesaurus (NALT) concepts 67,311

The Thesaurus is an online vocabulary of agricultural terms in English and Spanish and is cooperatively produced by the National Agricultural Library, USDA and the Inter-American Institute for Cooperation on Agriculture as well as other Latin American agricultural institutions belonging to the Agriculture Information and Documentation Service of the Americas (SIDALC)

Uploaded: 4/26/17

OntoBiotope (ONTOBIOTOPE) projects 3 classes 2,320

OntoBiotope is an ontology of microorganism habitats

Uploaded: 6/12/16

Protein Ontology (PR) projects 1 classes 83,656

An ontological representation of protein-related entities

Uploaded: 6/30/15

IBP Crop Research Ontology (CO_715) projects 3 classes 256

Describes experimental design, environmental conditions and methods associated with the crop study/experiment/trail and their evaluation.

Uploaded: 6/26/15

Process and Observation Ontology (PO2) projects 2 classes 4,449

A core ontology for modeling transformation processes and their observations.

Uploaded: 3/29/17

IBP Wheat Trait Ontology (CO_321) notes 1 projects 5 classes 1,023

Wheat Ontology

Uploaded: 9/19/16

<http://agroportal.lirmm.fr>

Supported by     

With the collaboration of      

Powered by [NCBO BioPortal](#)

- ▶ 106 ontologies
- ▶ 5 driving use cases

Ontology Alignment Context

What is an ontology?

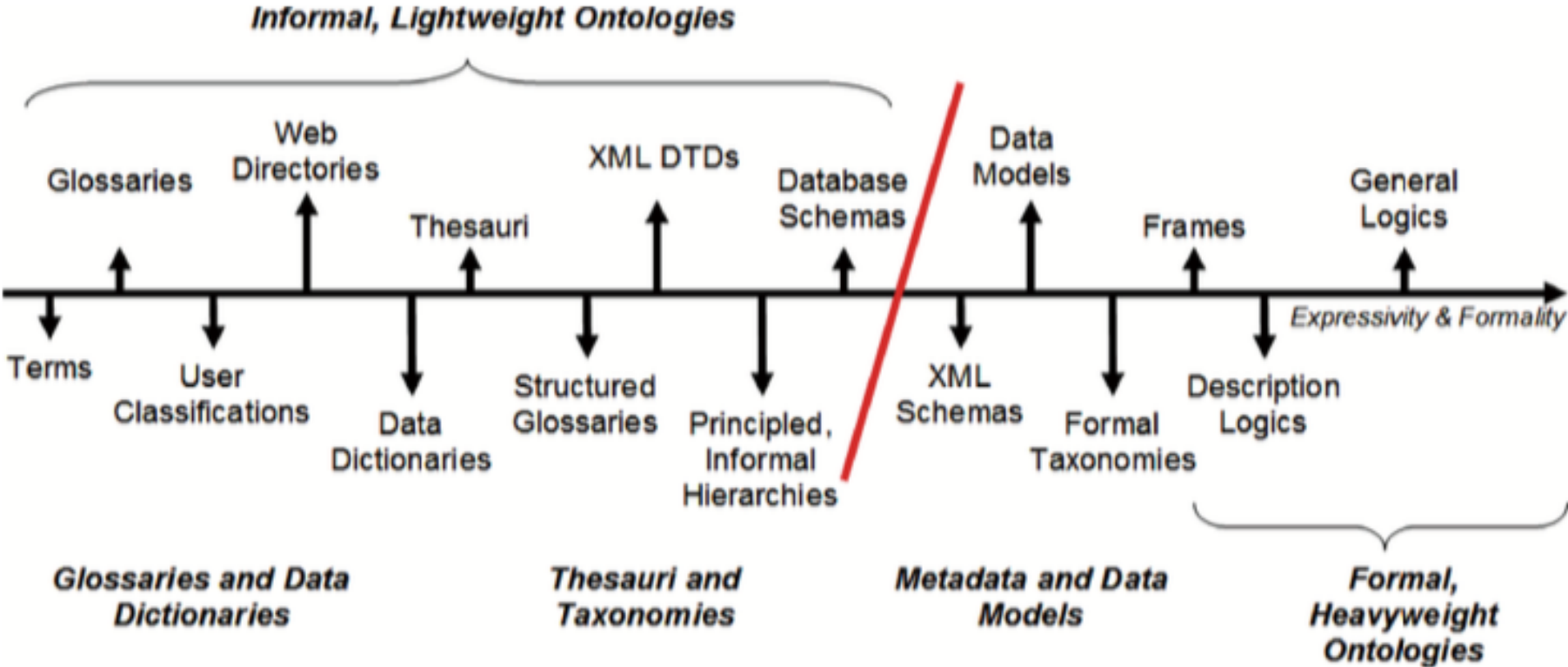
An ontology typically provides a **vocabulary** describing a domain of interest and a **specification of the meaning** of terms in that vocabulary.

Types of ontologies:

- Top Level Ontologies (Foundational Ontologies)
- Domain Ontology: Ex: biology, vehicle, cooking;
- Task Ontology: Ex: gene sequenceation;
- Application ontology: Ex: gene ontology;

Ontology Alignment Context

Forms of Ontology



Ontology Alignment Context

Ontology Entities:

- Classes and concepts
- Individuals or objects or instances
- Relations
- Data types
- Data values

Entities may be connected by various kinds of relations:

- Specialization, or subsumption
- Exclusion, or disjointness
- Instantiation, or typing

Ontology Alignment Context

Challenge

In **open or evolving** systems such as the semantic web different parties would, in general, **adopto different ontologies**. Merely using ontologies, like using XML, **does not reduce heterogeneity**: it raises heterogeneity problems to a higher level.

Ontology Alignment Context

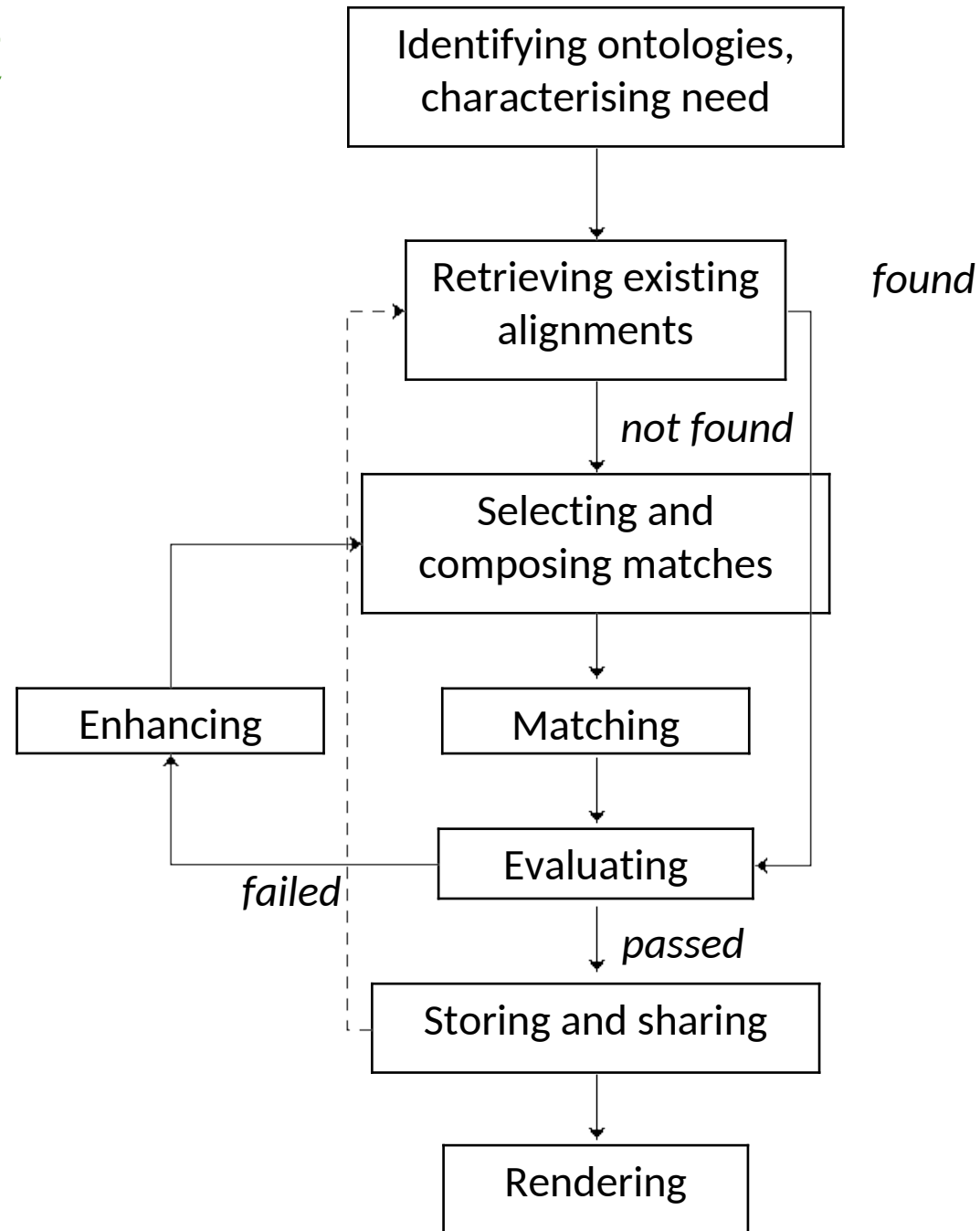
Types of Heterogeneity

- **Syntactic heterogeneity** --> different ontology languages (OWL, RDF, SKOS, OBO, etc)
- **Terminological heterogeneity** --> Different terms for the same concept (most common)
- **Conceptual heterogeneity / semantic heterogeneity**
 - *Difference in coverage* --> ontologies that covers different parts of a domain, Ex: plant disease and plant nutrition;
 - *Difference in granularity* --> different levels of detail on the same domain, Ex: geometry: point x circle;
 - *Difference in perspective* --> each ontology view one different aspect of the same concept, Ex: PLANT concept on the BIOLOGY perspective and on CROP PRODUCTION perspective;
- **Semiotic heterogeneity / pragmatic heterogeneity** --> difference on the human interpretation;

Ontology Alignment Context

Ontology matching methodology life cycle

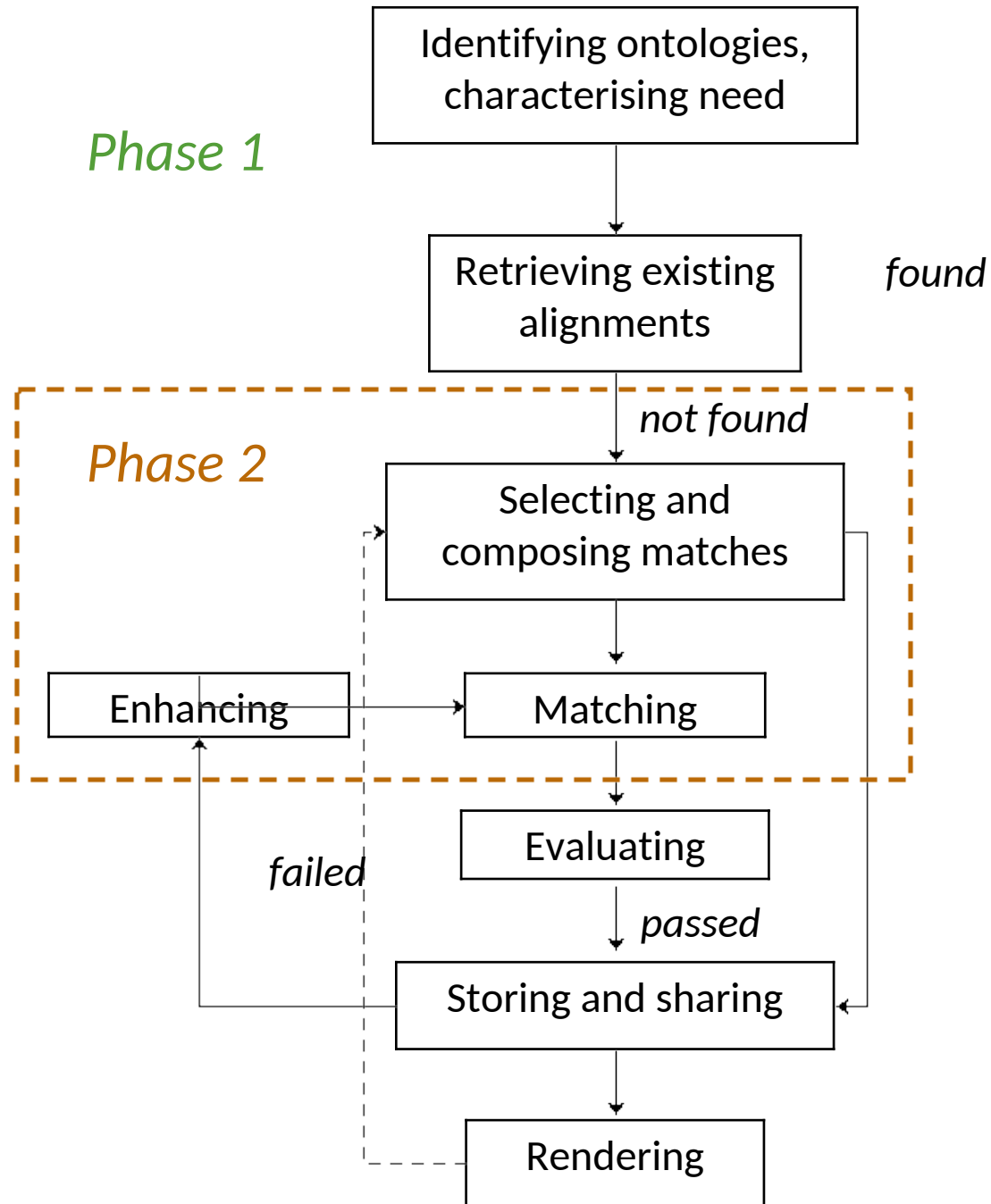
Euzenat and Shvaikio, 2013.



Ontology Alignment on Agroportal Project context

Phase 1 - Mappings already inside ontologies.

Phase 2 - Software tools to find matches on entity relations.



Ontology Alignment on Agroportal

Atual project stage **PHASE 1: Extracting, analysing and processing internal ontology mappings;**

How ?

- Accessing portal webAPI thru a script program developed in Java language that:
 - **Parse** the ontology on the XRDF format (could process other formats to);
 - **Trasnform** classes, individuals and relations in java objects;
 - **Identify** mappings of the types:
 - OWL sameAs and relatedTo;
 - SKOS exactMatch, relatedMatch, broaderMatch and narrowMatch;
 - OBO dbXref
 - **Classify** mappings according targets: Internal (Agroportal and Bioportal) and External (all ontologies outside Agroportal and Bioportal)
- Automaticly **lookup for valid URI** on:
 1. Agroportal
 2. Bioportal
 3. OBO Foundry
 4. Identifier.org
 5. Manual curation
- **Build JSON** file on the Agroportal Mapping format--> Share and Query porpose

Ontology Alignment on Agroportal

Challenges encountered on Phase 1 (internal mappings):

1. Different **ontology formats** (OWL, OBO, SKOS...)
 - SOLUTION: use OWL API to read all formats and transform then in Java objects.
2. Different **mappings location** (classes, individuals, axioms)
 - SOLUTION: parse all types of entities with 8 different approaches to lookup mappings.
3. Lack of formalism to **identifiers (URI)** on OBO dbXref (OWL and SKOS are better but still have some problems)
 - SOLUTION: Lookup external resources to find URI for concepts.
4. Too much **bad external references** (syntax errors, broken URLs, proprietary formats)
 - SOLUTION: clean up methods and manual curation of the most numerous maps

Ontology Alignment on Agroportal

Mappings Representation JSON format:

```
{
  "creator": "http://data.agroportal.lirmm.fr/users/admin",
  "source_contact_info": "",
  "relation": [
    "http://www.w3.org/2004/02/skos/core#exactmatch"
  ],
  "source": "http://data.agroportal.lirmm.fr/ontologies/GACS",
  "source_name": "GACS",
  "comment": "Generated with the Ontology Mapping Harvest Tool - v.1.0 - Agroportal
Project - LIRMM - 22/06/2018 15:22 - FR",
  "classes": {
    "http://id.agrisemantics.org/gacs/c6038": "GACS",
    "http://lod.nal.usda.gov/nalt/13592": "agroportal:http://lod.nal.usda.gov/nalt"
  }
}
```

Ontology Alignment on Agroportal

Mappings Representation JSON format:

```
"classes":
```

```
"http://id.agrisemantics.org/gacs/c6038": "GACS",
```

```
"http://lod.nal.usda.gov/nalt/13592": "agroportal:http://lod.nal.usda.gov/nalt"
```

ORIGIN ONTOLOGY



ORIGIN CONCEPT



TARGET CONCEPT



TARGET ONTOLOGY



Ontology Alignment on Agroportal

TARGET ONTOLOGIES/DATABASES URI LOOKUP

Methodology:

OWL and SKOS: Suposed to have valid URIs. (*)

OBO Xref:

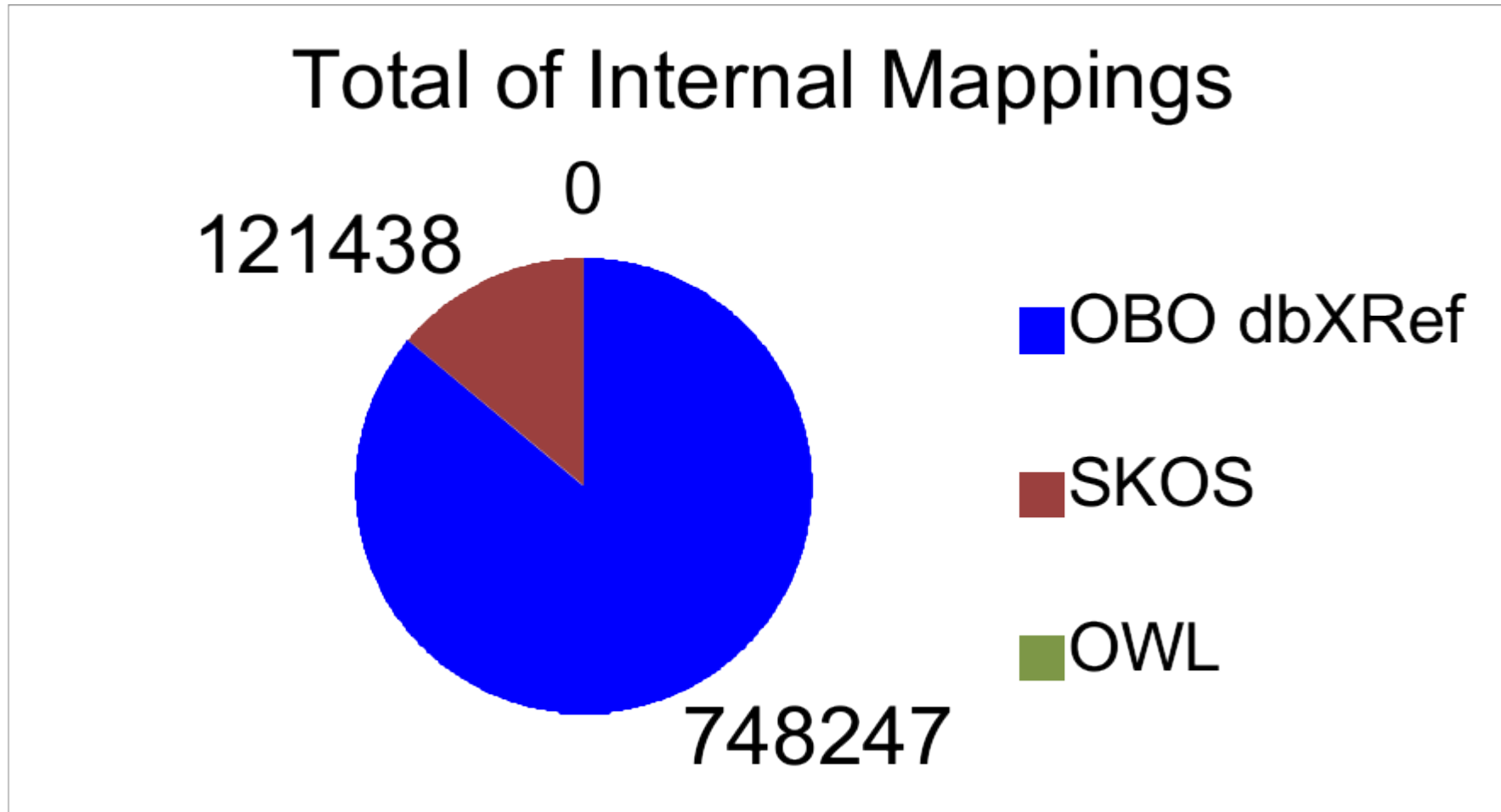
1. Find TARGET URI on Agroportal
2. Find TARGET URI on Bioportal
3. Find TARGET URI on OBO Foundry
4. Find TARGET URI on Identifiers.org

OWL, SKOS and OBO Xref:

5. Manual curation

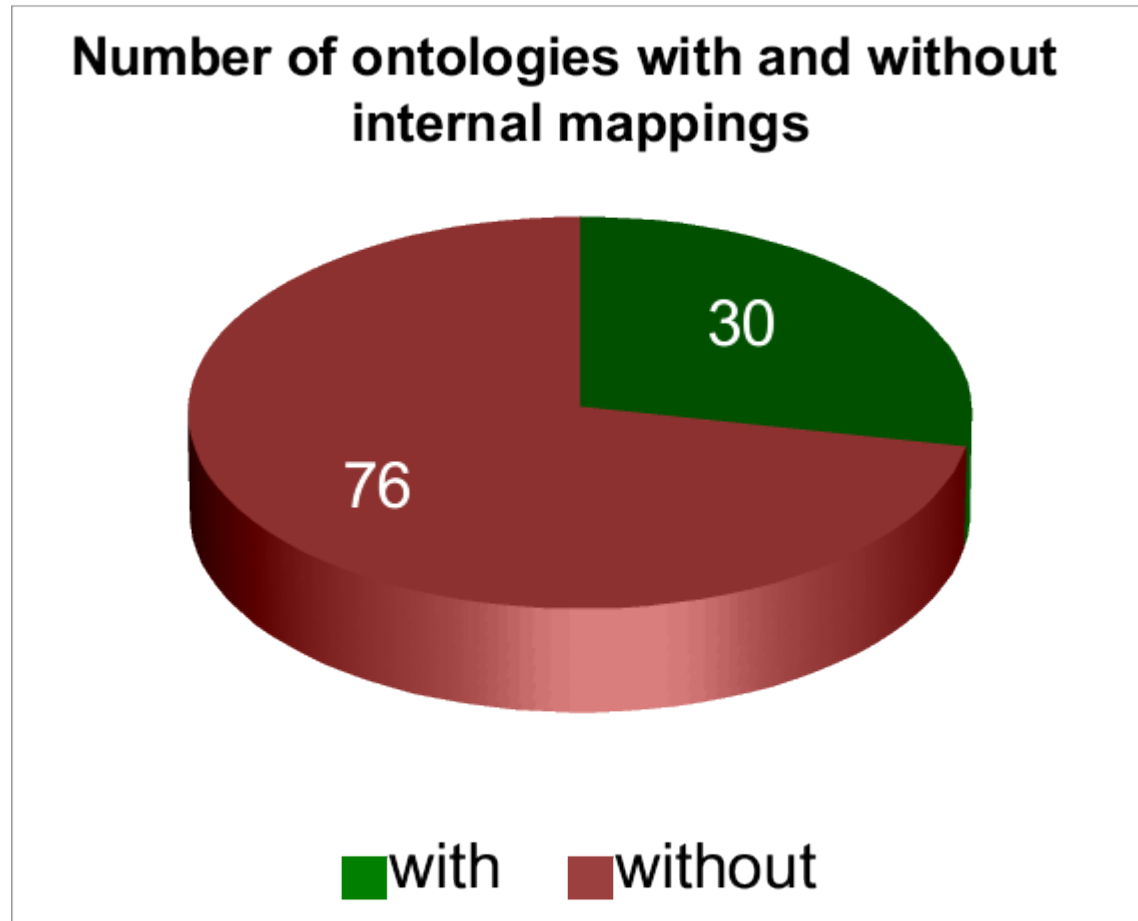
Ontology Alignment on Agroportal

Phase 1 parcial results: 869.685 mappings



Ontology Alignment on Agroportal

Phase 1 parcial results: 106 ontologies on Agroportal



N#	ONTOLOGY	MAPPINGS
1	PR	583.129
2	GO	97.599
3	AGROVOC	65.210
4	GACS	38.681
5	FOODON	13.287
6	TO	10.583
7	GEMET	9.658
8	FLOPO	7.757
9	PO	6.168
10	VT	4.816

Ontology Alignment on Agroportal

Phase 1 parcial results:

TARGET ONTOLOGIES/DATABASES URI LOOKUP

on Agroportal: 28

on Bioportal: 48

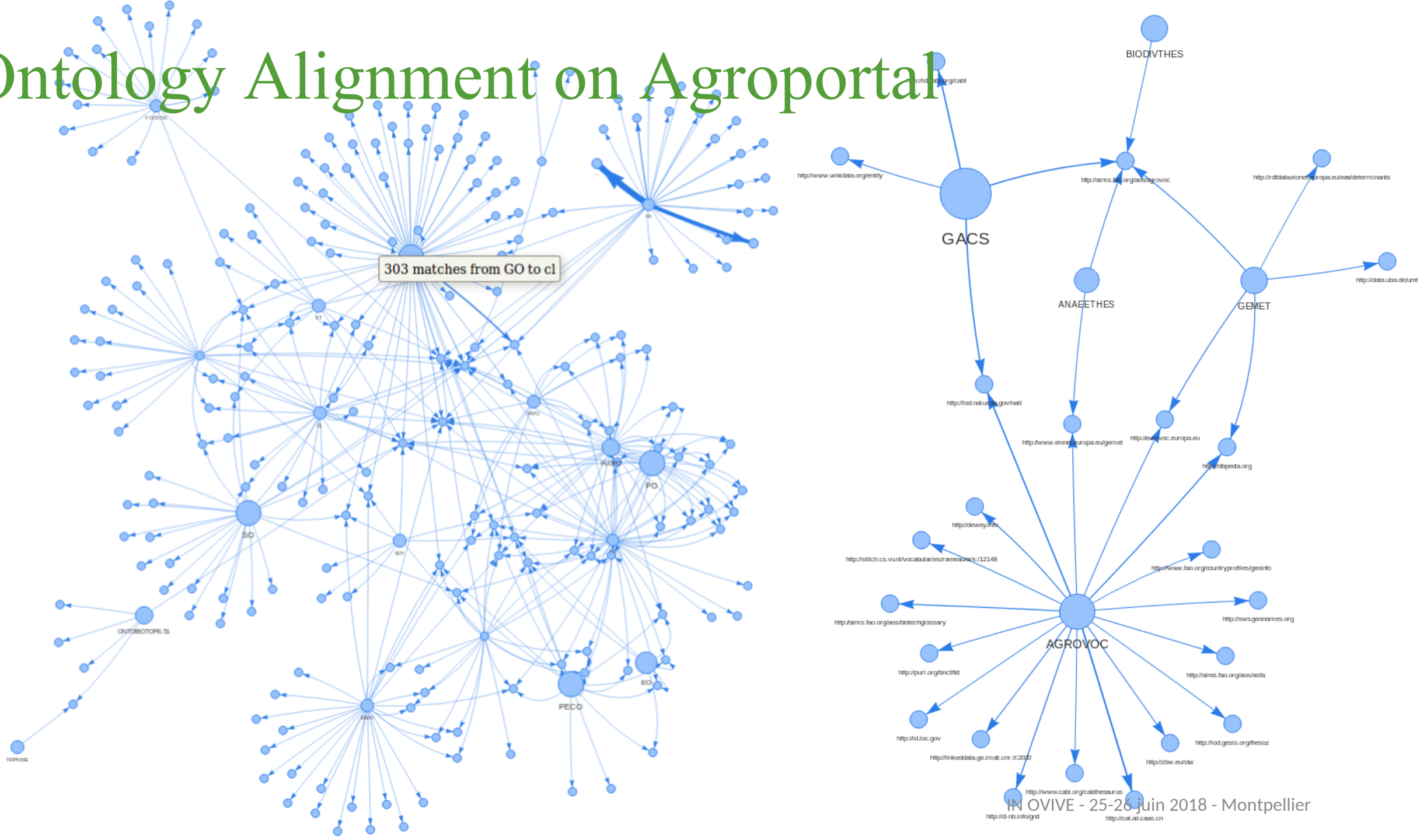
external reference: 128

not classified (on manual curation process): 574

already manually curated: 124

Total: 778 different TARGET ontologies/databases.

Ontology Alignment on Agroportal



Ontology Alignment on Agroportal

Phase 1 - on going next steps:

1. UPLOAD internal mappings to AGROPORTAL (Share thru Agroportal API);
2. Provide feedback to Agroportal community about internal mappings;

Phase 2 - next steps:

1. Find matches on ontologies with anthropomorphized tools (first is YAM++);
2. Connect YAM++ OnLine tool to Agroportal and make than extract mappings in an automatized way;
3. Provide a UI for the community in order to curate and evaluate mappings;

Ontology Alignment Roadmap Summary

1. **Extract** from all AgroPortal ontologies already existing **mappings defined inside the ontologies/vocabularies**.
2. **Upload** these mapping to AgroPortal's mappings repository with relevant provenance
3. Run **automatic ontology alignment** software (starting with YAMM++) on certain couple of ontologies in AgroPortal (if interested, please tell us).
4. **Align** all the ontologies and vocabularies in **AgroPortal to GACS**.
5. **Validate** previously generated mappings with ontology developers (YOU case by case).
6. **Implement** new visualization for community **mapping evaluation** (YOU as a community) **and curation** inside AgroPortal.
7. **Release and maintain** in AgroPortal a Linked **Open Data resource of all the mappings** generated and curated.

Thank you.

- Agroportal researchers
 - Clement Jonquet
 - Anne Toulet
 - Elcio Abrahão
- Contact us
 - agroportal-support@lirmm.fr

REFERENCES

Euzenat, J., Shvaikio, p. *Ontology Matching*, 2nd Ed. Springer Publishing Company, Incorporated ©2013 ISBN:3642387209 9783642387203

Chauhan, A. *Ontology Matching - An Introduction*, SCSE, VIT Chennai, 2015, Available at: <https://www.youtube.com/watch?v=HmhRxkAZVFA>, accessed at: 20/06/2018.

Technoist website. Notes: Ontology creation for cognitive computing. Available at: <https://technoist.com/2016/02/26/notes-ontology-creation-for-cognitive-computing/> Accessed: 20/06/18.

C. Jonquet, A. Toulet, E. Arnaud, S. Aubin, E. Dzalé Yeumo, V. Emonet, J. Graybeal, M-A. Laporte, M. A. Musen, V. Pesce & P. Larmande. *AgroPortal: an ontology repository for agronomy*, *Computers and Electronics in Agriculture*. 2017. pp. 30. Elsevier.