Extraction, generation and analisys of alignments between ontologies on Agroportal

Elcio Abrahão

elcio.abrahao@lirmm.fr

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

This presentation:

- Introdution
- Context backgroud
- 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 **A**ccessible **I**nteroperable **R**e-usable

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

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

http://agroportal.lirmm.fr

- ▶ 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
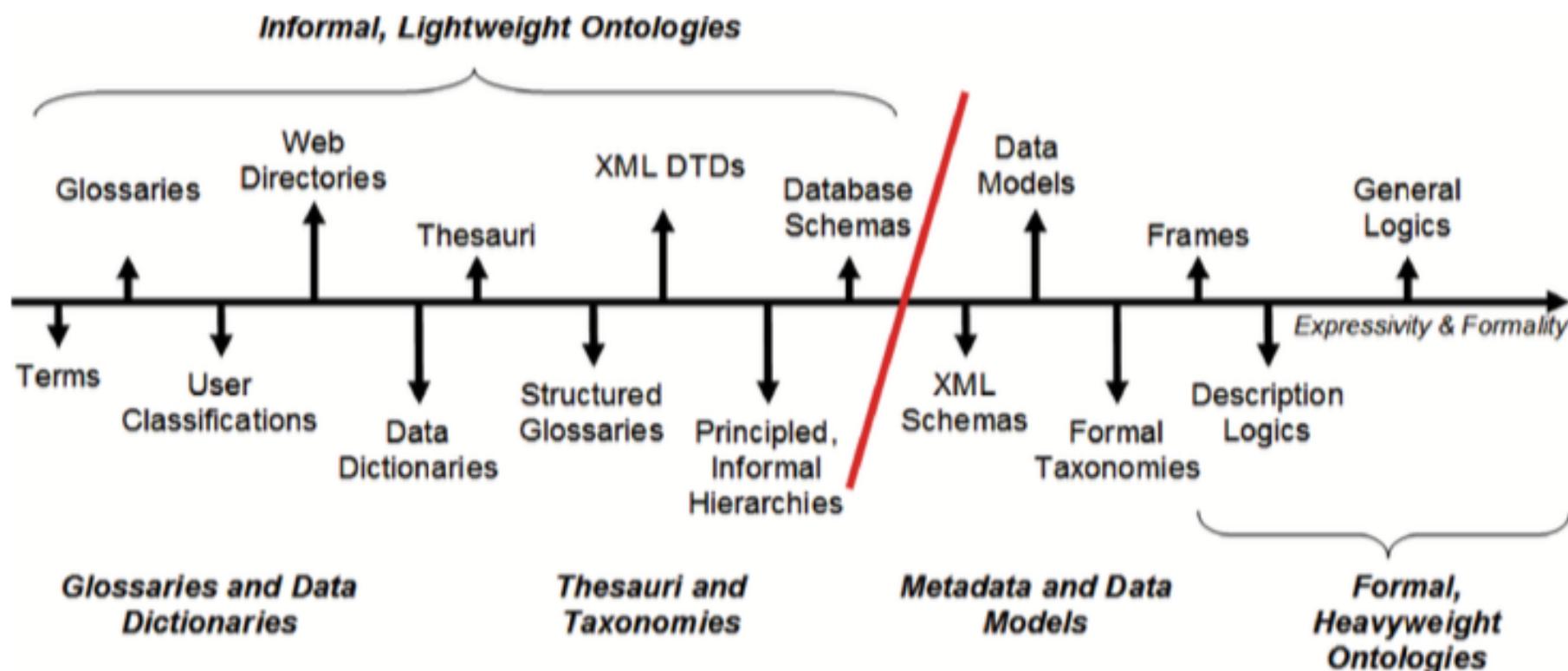
Ontology Alignment Context

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 commmom)
- **Conceptual heterogeneity / semantic heterogeneity**
    - *Difference in coverage* --> ontologies that covers different parts of a domain, Ex: plant desease 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 PRODUTION perspective;
- **Semiotic heterogeneity / pragmatic heterogeneity** --> difference on the human interpretation;

Ontology Alignment Context

Ontology matching methodology life cycle

Euzenat and Shvaikio, 2013.

Identifying ontologies, characterising need

Retrieving existing alignments

found

not found

Selecting and composing matches

Enhancing

Matching

Evaluating

failed

passed

Storing and sharing

Rendering
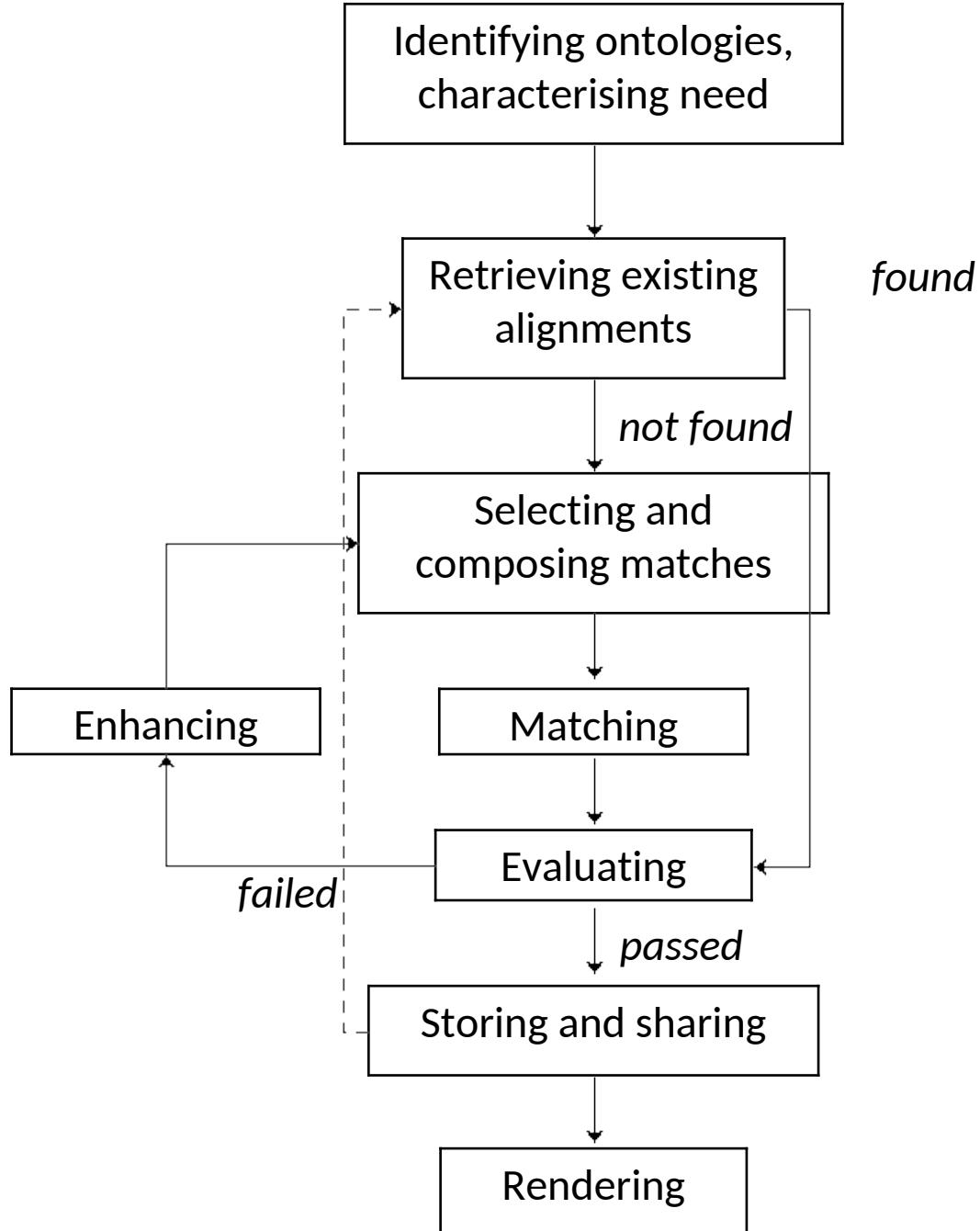
Ontology Alignment on Agroportal Project context

Phase 1 - Mappings already inside ontologies.

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

Phase 1

Phase 2

```
┌─────────────────────────┐
│ Identifying ontologies, │
│   characterising need   │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│   Retrieving existing   │      found
│       alignments        │
└─────────────────────────┘
            │
        not found
            │
            ▼
┌─────────────────────────┐
│     Selecting and       │
│   composing matches     │
└─────────────────────────┘
            │
            ▼
┌───────────┐      ┌───────────┐
│ Enhancing │─────▶│ Matching  │
└───────────┘      └───────────┘
            │
            ▼
     ┌─────────────┐
     │ Evaluating  │
     └─────────────┘
        failed │ passed
            ▼
┌─────────────────────────┐
│   Storing and sharing   │
└─────────────────────────┘
            │
            ▼
     ┌─────────────┐
     │  Rendering  │
     └─────────────┘
```

Ontology Alignment on Agroportal

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

How ?
- Acessing 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 narowMatch;
        - 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:

```json
{
  "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:

ORIGIN ONTOLOGY

```
"classes":

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

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

ORIGIN CONCEPT

TARGET CONCEPT
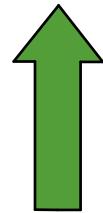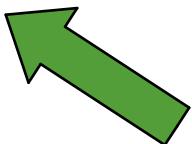
TARGET ONTOLOGY

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
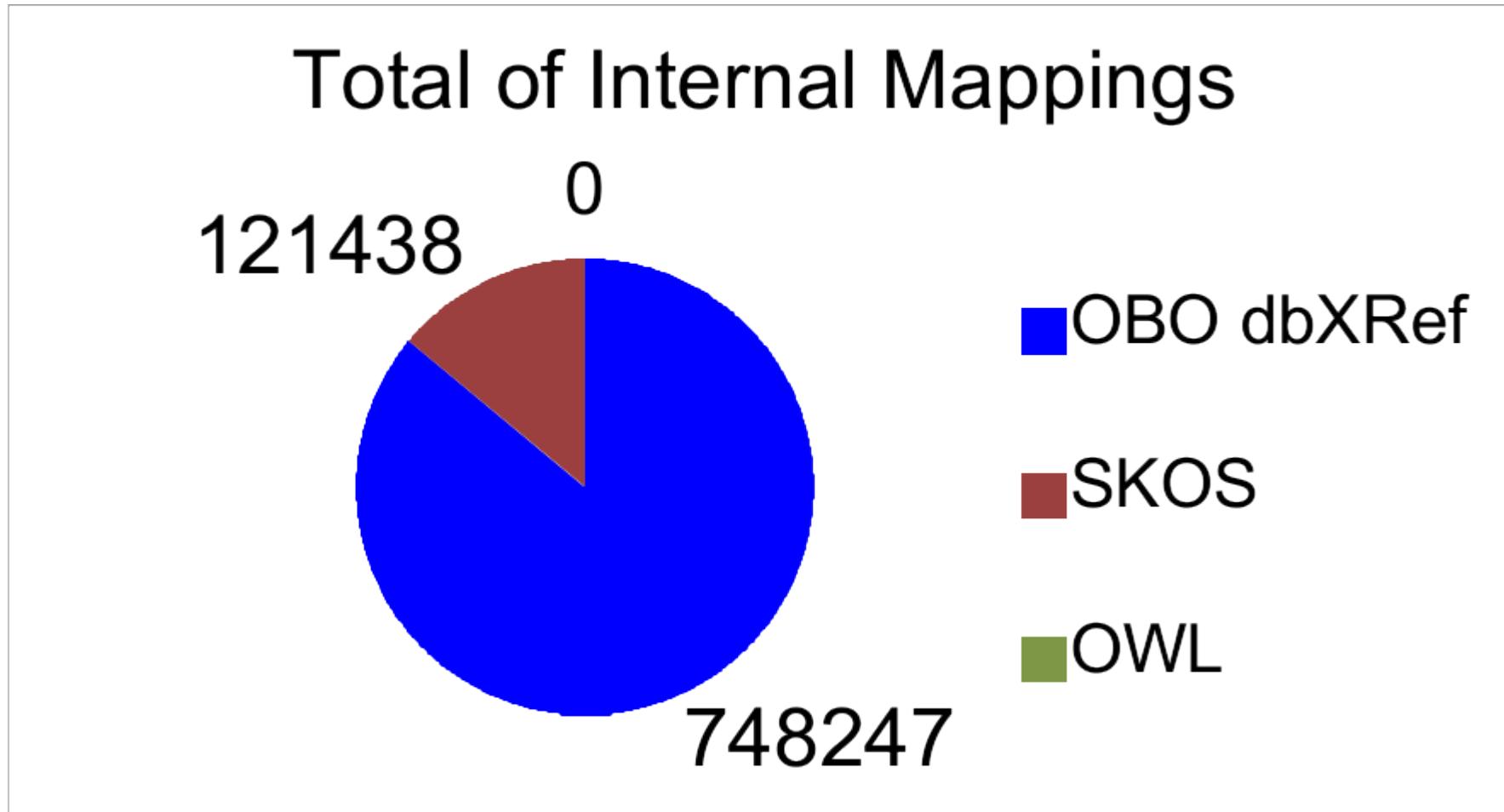
Ontology Alignment on Agroportal

Phase 1 parcial results: 106 ontologies on Agroportal

Number of ontologies with and without internal mappings

with: 30
without: 76
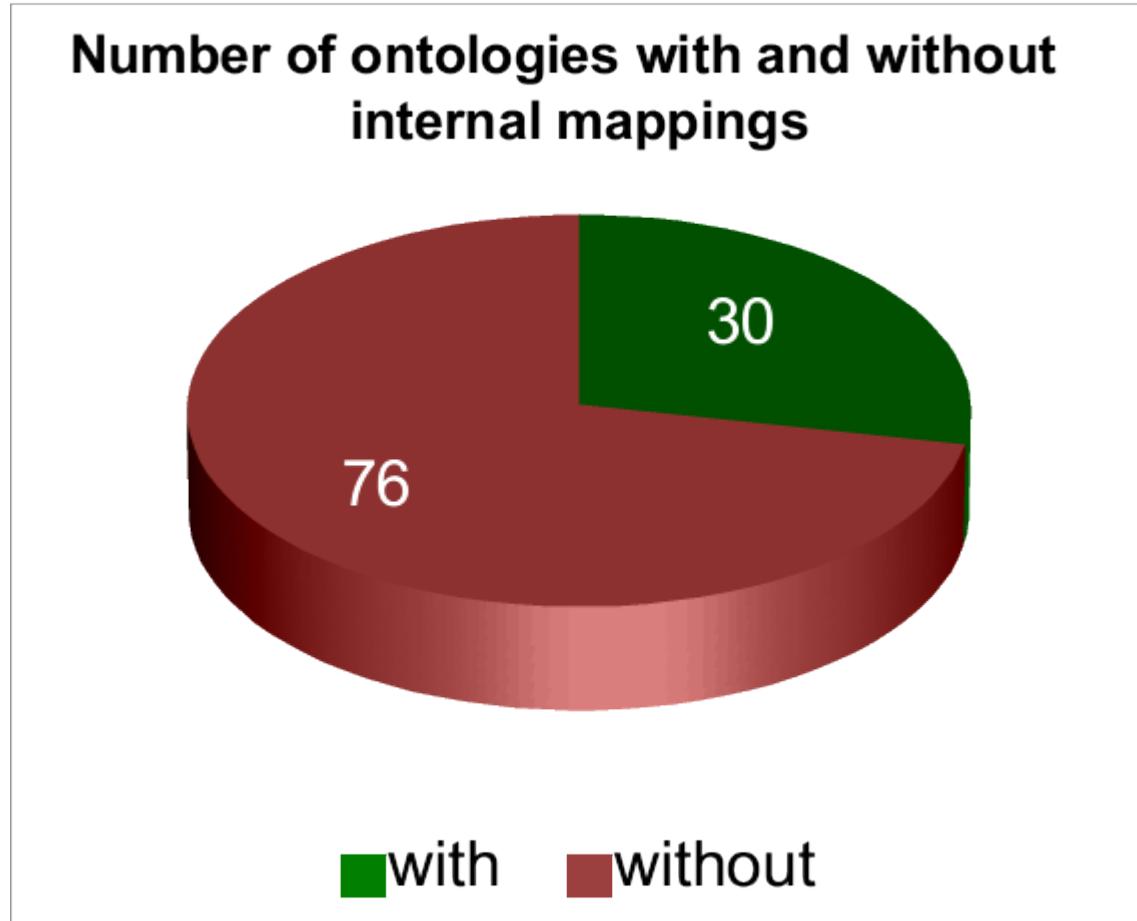
N#	ONTOLOGY	MAPPINGS
1	PR	583.129
2	GO	97.599
3	AGROVOC	65.210
4	GACS	38.681
5	FOODON	13.287
6	TO	10.583
7	GEMET	9.658
8	FLOPO	7.757
9	PO	6.168
10	VT	4.816

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
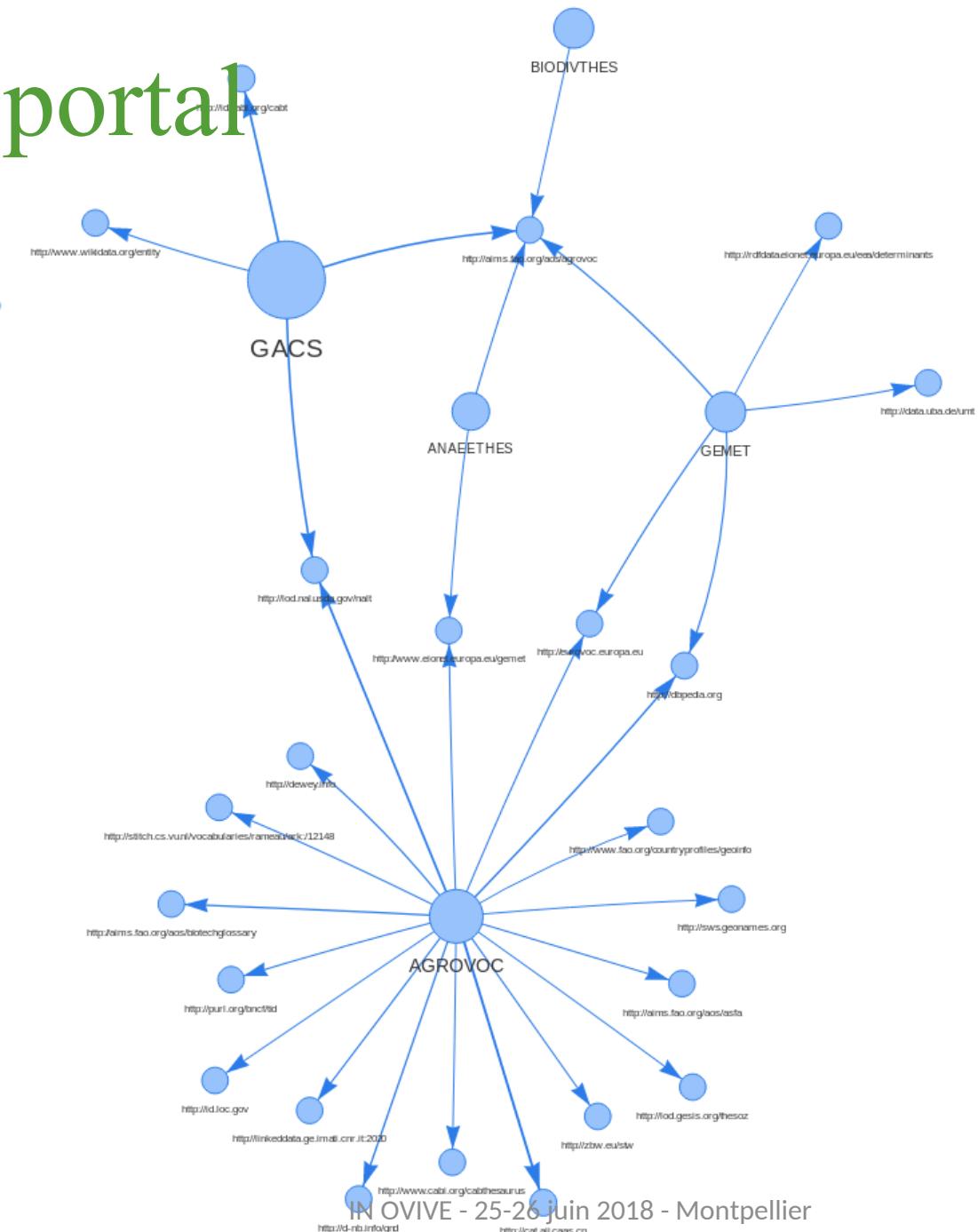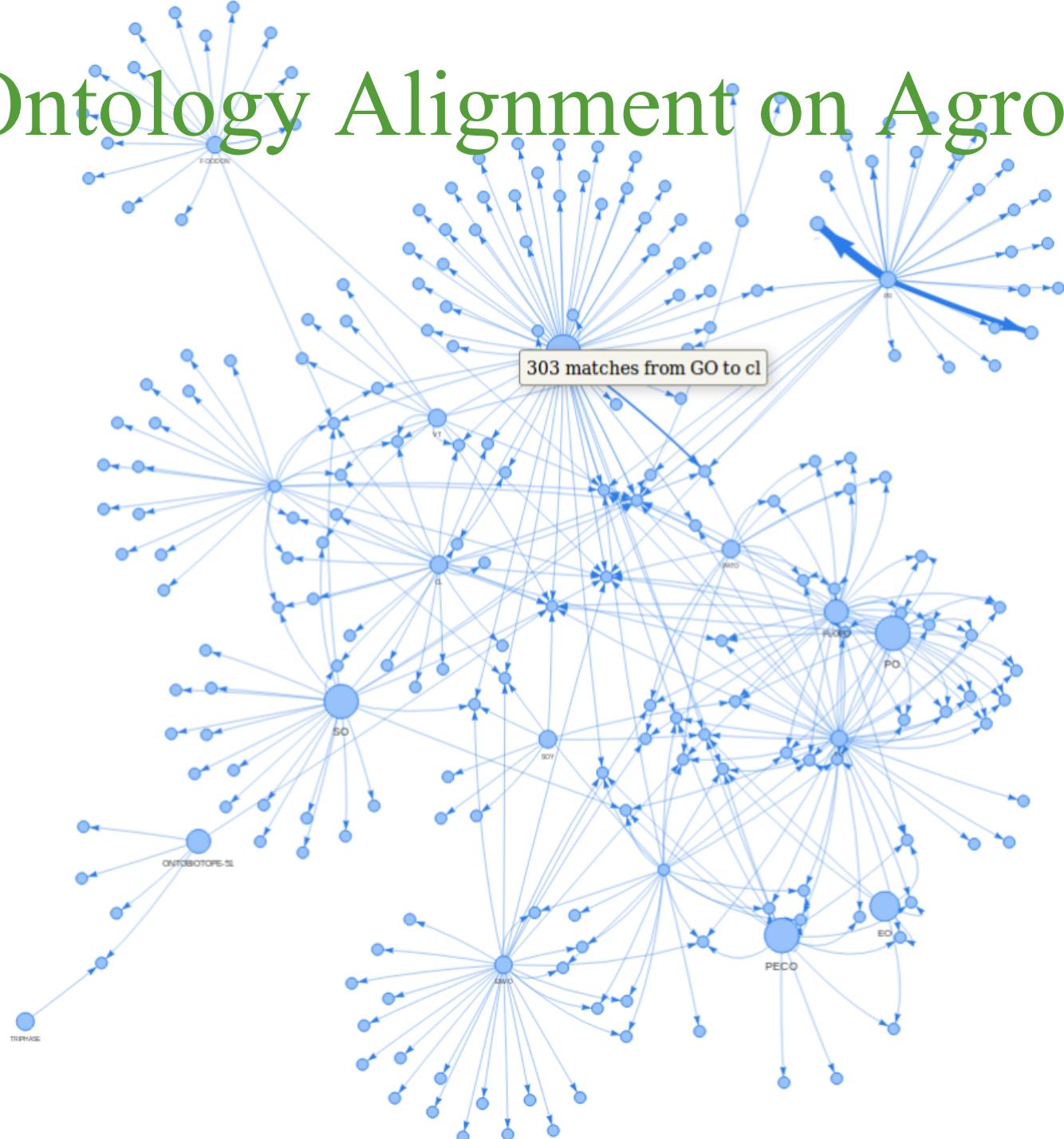
Ontology Alignment on Agroportal

Phase 1 - <u>on going</u> next steps:

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

Phase 2 - next steps:

1. Find matches on ontologies with anthomatized 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, Avaiable at: https://www.youtube.com/watch?v=HmhRxkAZVFA, accesed at: 20/06/2018.

Technoist website. Notes: Ontology creation for cognitive computing. Avaiable at: https://technoist.com/2016/02/26/notes-ontology-creation-for-cognitive-computing/ Accesed: 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.