

**Title:**

Postdoc position: Ontology alignment and engineering in agriculture and biodiversity

**Information:**

**Context:** Lingua project, in the context of AgroPortal project (<http://agroportal.lirmm.fr>), supported by [NUMEV](#), [AGRO](#) and [CEMEB](#) Labex. Collaboration between IRD, INRA, CNRS (CEFE), as well as with the [NCBO](#) (Stanford University).

**When:** January 2018 – June 2019 (18 months) - extensions might be possible on other supports. Application will be accepted until November 28<sup>th</sup>

**Where:** Montpellier. [LIRMM](#) mainly, with assignments at INRA and/or CEFE. Missions at Stanford.

**Net salary:** Between 2100 and 2500€ per month depending on qualifications. Includes benefits.

**Keywords:**

(agronomical) ontologies & vocabularies, semantic web, ontology management, ontology alignment, semantic interoperability, linked data, semantic annotation, application to agronomy & biodiversity.

**Technologies:**

Web development, Ruby/Rails, Java/JEE, RESTful web services, XML/JSON, Semantic Web technologies (OWL, RDF, SPARQL, triplestore (4store), Linked data), NCBO technology (AgroPortal/BioPortal).

**Context:**

Standards vocabularies and ontologies are key elements to achieve data interoperability. The **AgroPortal project** (<http://agroportal.lirmm.fr>) develops and supports a reference ontology repository for agronomy, plant sciences, biodiversity and nutrition. We have already designed and implemented an advanced prototype offering ontology-based services that hosts 64 ontologies or vocabularies including some reference resources in the domain: Agrovoc, NAL thesaurus, Crop Ontology, etc. **One of the challenges when dealing with multiple ontologies is to determine their overlap and align them.**

We are offering a postdoc position to investigate and develop ontology mapping capabilities for AgroPortal ontologies and participate into the **international Global Agricultural Concept Scheme (GACS) project**. With the experience and technology developed with the YAM++ application (LIRMM's ontology alignment matcher - <http://yamplusplus.lirmm.fr>), we will make AgroPortal a state-of-the-art platform for mapping extraction, generation, validation, evaluation, storage and retrieval by adopting a complete semantic web and linked open data approach and engaging the community for curation. We will first focus on the ontologies of the Montpellier community (agronomy, food, biodiversity) and then **join the GACS project** (integration of Agrovoc, NAL Thesaurus & CAB Thesaurus) of the RDA AgriSemantics working group (<http://agrisemantics.org>).

**Detailed description:**

A key aspect in **addressing semantic interoperability in agronomy, plant sciences, nutrition and biodiversity is the use of ontologies as a common denominator to describe data, make them** interoperable and turn them into structured and formalized knowledge. Biomedicine has always been a leading domain for semantic interoperability pioneering the development of reference ontologies such as the Gene Ontology. This has served as model for the agronomic, environmental and plant sciences e.g., Plant Ontology [1], Crop Ontology [2], opening the space to various types of semantic applications [3], to data integration or decision support. Semantic interoperability has been identified as a key issue for agronomy and biodiversity sciences, and the use of ontologies a way to address it [4], [5]. **The more ontologies and vocabularies are being produced in the domain, the more the need to create, store and retrieve alignments between those ontologies becomes important.**

By reusing the NCBO BioPortal technology, **we have designed AgroPortal, an ontology repository for the agronomy domain** (<http://agroportal.lirmm.fr>) [7]. The main objective of the AgroPortal project is to develop

and support a reference ontology repository for agronomy, plant sciences, nutrition, and biodiversity. It offers a robust and reliable service to the community that features ontology hosting, search, versioning, visualization, comment, services for semantically annotating data with the ontologies, as well as storing and exploiting ontology alignments, all of these in a semantic web compliant infrastructure. Ontologies in the portal are being developed within multiple **agronomic use cases**, including the *Agronomic Linked Data* (<http://agroid.org>), *INRA Linked Open Vocabularies* (<http://lovinra.inra.fr>) which is an effort to publish vocabularies produced or co-produced by INRA.

YAM++ is a state-of-the-art ontology alignment system being developed at LIRMM [8]. YAM++ uses machine-learning techniques to combine different similarity measures, exploiting the intrinsic textual features of ontologies to provide similarity scores based on information retrieval techniques. YAM++ obtained excellent results during the OAEI 2013 campaign. Since 2016, YAM++ exists also in the form of a multifunctional web service application (<http://yamplusplus.lirmm.fr>) allowing manual mapping validation and enrichment.

The postdoc's mission will be to:

- Work with partners on the **design** (with use of semantic web standards) of their ontologies/vocabularies and the integration (when not done yet) within AgroPortal.
- **Align** the ontologies within AgroPortal to one another and to the GACS vocabulary (cf. below), focusing on ontologies developed by the Montpellier partners first. Release mappings as linked open data.
- **Design and develop** a state-of-the-art ontology alignment framework based on YAM++/AgroPortal to extract, generate, validate, evaluate, store and retrieve ontology alignments. **Work** with partners on generating and curating mappings thanks to the framework developed.
- Contribute to the **GACS** project with the AgroPortal alignment framework and customize AgroPortal to use GACS as a pivot vocabulary.
- **Demonstrate** to/with each partner (mainly INRA & CEFE) the outcomes of the use of ontologies, mappings and annotations.

The project will have four uses cases:

- **AgroLD:** AgroLD uses the OWL versions of multiple AgroPortal ontologies and rely on the AgroPortal Annotator web service to annotate more than 50 datasets. We will build a resource that bridges the gap between these references ontologies and formalize their alignments to AgroLD data.
- **LovInra:** LovInra ontologies are not always interconnected one another (when relevant) therefore, we will especially focus on producing alignments between LovInra ontologies. Ontologies will be mapped to GACS to implement a larger agricultural interoperability strategy (cf. GACS here after).
- **Biodiversity:** In partnership with CEFE, we will work on integrating the Thesaurus Of Plant characteristics (TOP) [10], within AgroPortal and work on the alignments (existing and to be created) to other ontologies.
- **GACS:** In collaboration with RDA Agrisemantics working group (<http://agrisemantics.org>) we will work on the development of Global Agricultural Concept Scheme (GACS) which is an important international initiative to integrate the Agrovoc, CAB Thesaurus, and NAL Thesaurus ([www.agrisemantics.org/gacs](http://www.agrisemantics.org/gacs))[6]. Because of this size and endorsements by major organizations, the GACS will certainly become the future pivot vocabulary in the lingua franca for agriculture (and related domains) and AgroPortal has been proposed to the Agrisemantics WG as the platform for accessing each of the three original thesaurus as well as the GACS itself. We will produce alignments to build GACS and to interconnect it to other ontologies in AgroPortal.

#### Expected profile:

- Researcher with a recent (less than 5 years) PhD in Informatics / Computer science.
- Experience abroad (PhD or previous postdoc done outside of France), strongly recommended.
- Good Web developer experience with knowledge of Web, JEE technologies and Ruby/Ruby On rails.
- Experience with semantic Web technologies.
- Background knowledge and/or experience in the biological / agronomical context is preferred.
- Excellent research skills to gather both the local and international community on AgroPortal.
- Perfect English oral and writing skills.
- Basic knowledge of French with objective to learn the language during the contract.
- Excellent writing skills and publication motivation.

- International trips accepted (collaboration with Stanford) and possibility to get a visa for the USA.
- Autonomy and initiative, take on technical decisions within the project and justification of choices.
- Friendly person to join a small research team in Montpellier.

#### Application:

For more information about this position, please contact Clement Jonquet ([jonquet@lirmm.fr](mailto:jonquet@lirmm.fr)) and Konstantin Todorov ([konstantin.todorov@lirmm.fr](mailto:konstantin.todorov@lirmm.fr)). To apply, please send an email including links to (PLEASE, NO ATTACHED DOCUMENTS) the following:

- a motivation letter describing an explanation of your interest for the position;
- a curriculum vitae describing your experience and the matches with the expected profile;
- copies of diplomas and other relevant certificates;
- names and contact details of referees.

#### References:

- [1] L. Cooper *et al.*, "The Plant Ontology as a Tool for Comparative Plant Anatomy and Genomic Analyses," *Plant Cell Physiol.*, 54, 2, 2012.
- [2] R. Shrestha *et al.*, "Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature.," *AoB Plants*, vol. 2010, p. plq008, Jan. 2010.
- [3] X. Meng, "Special Issue – Agriculture Ontology," *Journal of Integrative Agriculture*, vol. 11, no. 5. Elsevier, p. i, 2012.
- [4] J. S. Madin, S. Bowers, et al. "Advancing ecological research with ontologies.," *Trends Ecol. Evol.*, 23, no. 3, pp. 159–68, Mar. 2008.
- [5] R. L. Walls *et al.*, "Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies," *PLoS One*, vol. 9, no. 3, p. e89606, Mar. 2014.
- [6] T. Baker, C. Caracciolo, and O. Suominen, "GACS Core: Creation of a Global Agricultural Concept Scheme," 2016, pp. 311–316.
- [7] C. Jonquet *et al.*, "Reusing the NCBO BioPortal technology for agronomy to build AgroPortal," in *7th International Conference on Biomedical Ontologies, ICBO'16, Demo Session*, 2016, no. D203, p. 3. EXTENDED VERSION UNDER REVIEW COMPAG JOURNAL
- [8] D. Ngo and Z. Bellahsene, "YAM++ : A Multi-strategy Based Approach for Ontology Matching Task," in *18th International Conference on Knowledge Engineering and Knowledge Management, EKAW'12*, 2012, vol. 7603, pp. 421–425.
- [9] D. Ngo and Z. Bellahsene, "YAM++ results for OAEI 2013," in *8th Int. Work. on Ontology Matching*, 2013, vol. 1111, pp. 211–218.
- [10] E. Garnier *et al.*, "Towards a thesaurus of plant characteristics: an ecological contribution," *Ecology*, 105, 2, pp. 298-309, Mar. 2016.