

# D2KAB project taking off: Data to Knowledge in Agronomy and Biodiversity

*Submission for Journée thématique Agronomie et IA:*  
<http://pfia2020.fr/journeesthematiques/agronomie-et-i-a/>

- Clement Jonquet, LIRMM, University of Montpellier, CNRS
- Sophie Aubin, DIST, INRAE, Versailles
- Anne-Françoise Adam-Blondon, URGI, INRAE, Université Paris-Saclay
- Michael Alaux, URGI, INRAE, Université Paris-Saclay
- Florence Amardeilh, Elzeard, Bordeaux
- Mouhamadou Ba, MaIAGE, INRAE, Université Paris-Saclay
- Stephan Bernard, UR TSCF, INRAE, Université Clermont Auvergne
- Pierre Bisquert, IATE, INRAE, University of Montpellier
- Robert Bossy, MaIAGE, INRAE, Université Paris-Saclay
- François Brun, ACTA
- Patrice Buche, IATE, INRAE, University of Montpellier
- Arnaud Castellort, LIRMM, University of Montpellier, CNRS
- Olivier Corby, I3S, Université Côte d'Azur, CNRS, Inria
- Julien Cufi, IATE, INRAE, University of Montpellier
- Romain David, MISTEA, INRAE, Montpellier SupAgro, University of Montpellier
- Louise Deléger, MaIAGE, INRAE, Université Paris-Saclay
- Stephane Dervaux, MIA, INRAEE
- Esther Dzalé-Yeumo, DIST, INRAE, Versailles
- Catherine Faron Zucker, I3S, Université Côte d'Azur, CNRS, Inria
- Eric Garnier, CEFÉ, CNRS, Univ. of Montpellier, Univ. Paul Valéry, EPHE, IRD
- John Graybeal, BMIR, Stanford University, USA
- Théo-Paul Haezebrouck, API-AGRO, 149 rue de bercy 75012 PARIS
- Matthieu Hirschy, ACTA
- Liliana Ibanescu, AgroParisTech, MaIAGE, INRAEE
- Pierre Larmande, DIADE, IRD & University of Montpellier
- Luc Menut, IATE, INRAE, University of Montpellier
- Franck Michel, I3S, Université Côte d'Azur, CNRS, Inria
- Mark Musen, BMIR, Stanford University, USA
- Claire Nédellec, MaIAGE, INRAE, Université Paris-Saclay
- Christian Pichot, INRAE-URFM
- François Pinet, UR TSCF, INRAE, Université Clermont Auvergne
- Cyril Pommier, URGI, INRAE, Université Paris-Saclay
- Catherine Roumet, CEFÉ, CNRS, Univ. of Montpellier, Univ. P. Valéry, EPHE, IRD
- Catherine Roussey, UR TSCF, INRAE, Université Clermont Auvergne
- Konstantin Todorov, LIRMM, University of Montpellier, CNRS
- Anne Toulet, DGD-RS, DIST, CIRAD

*Note on the authors: C. Jonquet is the contact author, S. Aubin will represent the project during the workshop. All the authors are the researchers and staff involved at project kick-off in June 2019 or later. Except BMIR, all the author affiliations are in France.*

*Note about the submission: A equivalent short note about the project was presented as a poster at the 14th Plenary meeting of the Research Data Alliance, Helinski, October 2019:*  
<https://doi.org/10.5281/zenodo.3520300>

## Abstract

Agronomy/agriculture and biodiversity (ag & biodiv) communities face several major societal, economic, and environmental challenges that data science approaches will help address. To achieve their goals, researchers of these communities must be able to rapidly discover, aggregate, integrate, and analyse different types of data and information sources. Semantic technologies, combined to open, FAIR data and services, is one of the answers to fully knowledge-driven, and transparent science and innovation. **The D2KAB project ([www.d2kab.org](http://www.d2kab.org)) aims to create a framework to turn agronomy and biodiversity data into knowledge – semantically described, interoperable, actionable, open – and investigate the scientific methods and tools to exploit this knowledge for applications in agriculture and biodiversity sciences.** This project, funded by French ANR (2019-2023), will provide the means –ontologies and linked open data– for ag & biodiv to embrace semantic Web technologies in order to produce and exploit FAIR data and services. To do so, D2KAB will develop new original methods and algorithms in the following areas: data integration, text mining, semantic annotation, ontology alignment and linked data exploitation and visualization.

D2KAB project brings together a unique **multidisciplinary consortium of 12 partners** to achieve this objective: 2 informatics research units (LIRMM, I3S); 6 INRAE/IRD research units at the interface of computer science and ag & biodiv (URGI, DipSO, MaIAGE, IATE, TSCF, DIADE) specialized in agronomy or agriculture; 2 labs in biodiversity and ecosystem research (CEFE, URFM); 1 association of agriculture stakeholders (ACTA); and 1 partnership with Stanford BMIR department.

### Three main goals drive D2KAB's roadmap:

1. To develop state-of-the-art methods and technologies for ontology lifecycle and alignment.
2. To build the agronomy, agriculture and biodiversity Linked Open Data cloud.
3. To enable new semantically driven agronomy and biodiversity science.

Some of the **key technological building blocks** of D2KAB are AgroPortal, a reference repository for ontologies and vocabularies in agronomy; AgroLD, a semantic Web knowledge base that integrates agronomic data from public databases including Gene Ontology associations, Gramene, UniprotKB, and OryGenesDB ; Corese, a semantic Web factory that implements the W3C standards RDF, RDFS, OWL-RL and SPARQL, and LDScript, a Linked Data Script Language, and STTL, the SPARQL Template Transformation Language for RDF; and Alvis, a text mining for semantic normalisation of free text by ontologies. Part of the consortium is involved in several **national or international working groups** such as the Research Data Alliance (e.g. Agrisemantics WG, Vocabulary Services IG, Wheat and Rice Data Interoperability WGs, Agricultural Data IG, SHARC IG), the GO FAIR Food System Implementation Network or the CNRS-GDR SemanDiv.

D2KAB will allow the valorization of ag & biodiv data into real world applications leading to economic impact, smart agriculture and ecological preservation. **Five driving scenarios** are planned:

- *Development of an ontology-based expert system to select food packaging solutions.* IATE develops an ontology for food material processing and food packaging characteristics (Matter Transfer Ontology – TRANSMAT in AgroPortal). It drives concrete solutions for farmers, packaging solution suppliers and packed food suppliers such as an application to automatically select the most appropriate food package for a given food considering multiple variables (food respiration, temperature, material to use, etc.).
- *Creation of an augmented semantic reader for Plant Health Bulletins.* We address agricultural advisers and stakeholders of biovigilance in agriculture by building a semantically augmented reading interface for the official Plant Health Bulletins (Bulletins de Santé du Végétal).
- *Advanced integration of textual and experimental data on wheat phenotypes.* Wheat phenotyping data integration (from phenotypic experimental data to bibliographic phenotype information) will enable research, economical and societal outcomes. The main challenge is the mapping of low level observation measures (e.g., weight of 1000 grains) and abstract qualitative properties (e.g., high yield). We use MalAGE's wheat text-mining application and data will be integrated in GnpIS.
- *Development of new semantic resources* such as: (i) the extension of the ANAEE Thesaurus which provides a controlled vocabulary for the semantic description of the study of continental ecosystems and their biodiversity; (ii) the extension of the Thesaurus Of Plant Characteristics to root traits;
- *Integration of plant functional biogeography data related to the Mediterranean Basin.* Here, CEFE focus on ontology-driven data integration for functional biogeography, with the aim to understand and generalize trait-environment-relationships across the Mediterranean Basin. We are currently assembling massive datasets about vegetation (botanical censuses, plant traits) and environmental (climate, soils, etc.) variables available across this region.

Each of the project scenarios will have a significant impact and produce concrete outcomes for ag & biodiv scientific communities and socio-economic stakeholders in agriculture.

## Keywords

Semantic Web, Ontologies, Ontology repository, Ontology alignment, Linked open data, Knowledge graphs, Data integration, Text mining, Agronomy, Agriculture, Biodiversity, Ecosystem