



# Project D2KAB

Data to Knowledge in Agronomy and Biodiversity  
2019-2023

[WWW.D2KAB.ORG](http://WWW.D2KAB.ORG)

## Journée Agronomie et IA

PFIA 2020, en ligne - 1er juillet 2020

Sophie Aubin – [sophie.aubin@inrae.fr](mailto:sophie.aubin@inrae.fr)





# Data to Knowledge in Agronomy and Biodiversity

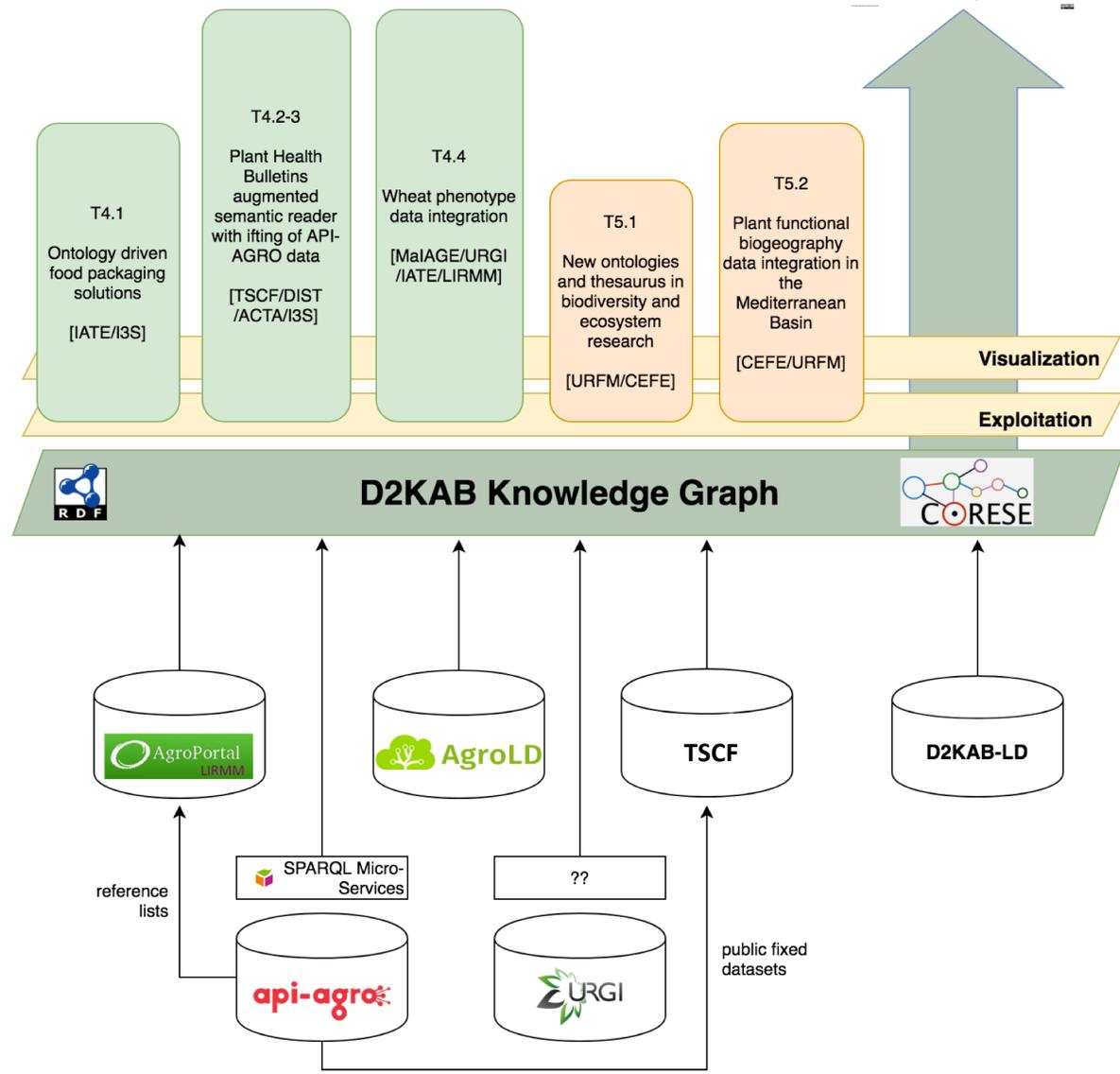
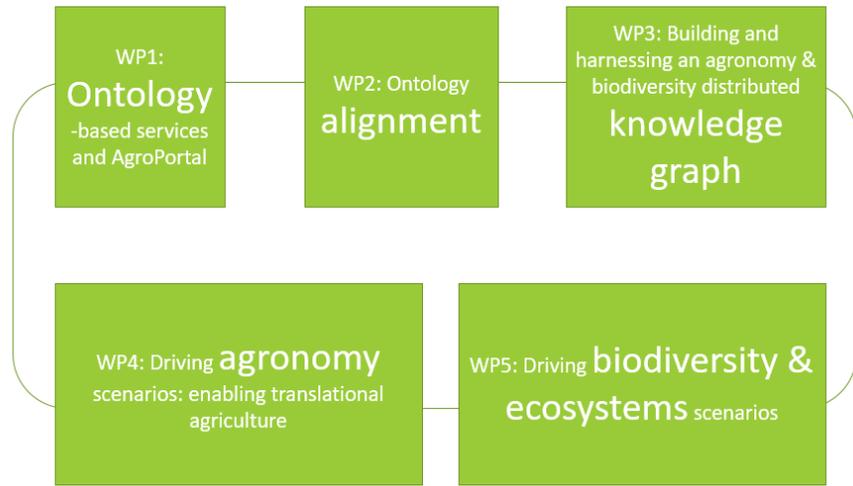


<b>Theme(s)</b>	<b>B.7 Axe 4 – CES 23</b>
<b>Type of research/project</b>	Experimental Development / Collaborative Research Project
<b>Project duration</b>	48 months
<b>Grant requested</b>	951K€ (overall budget of 3.1M€)
<b>Coordinator partner</b>	LIRMM (Dr. Clement Jonquet)
<b>Multidisciplinary Consortium</b>	<p>11 partners:</p> <ul style="list-style-type: none"> <li>- 2 informatics research units (LIRMM, I3S);</li> <li>- 5 INRAE applied informatics research units (URGI, MaIAGE, IATE, DipSO, TSCF) specialized in agronomy or agriculture;</li> <li>- 2 labs in biodiversity and ecosystem research (CEFE, URFM);</li> <li>- 1 association of agriculture stakeholders (ACTA);</li> <li>- 1 international partnership with Stanford BMIR</li> </ul>
<b>Int. cooperation</b>	<p>Formal: STANFORD-BMIR (Stanford Center for Biomedical Informatics Research)</p> <p>Exchanges (not in consortium): IRD-USTH (Hanoi), UC Davis (USA)</p>
<b>Link with other PIAs</b>	Labex NUMEV, AGRO CEMEB (Montpellier), I-Site MUSE and CAP2025 PIA projets IBC, ISTEEX, IFB, AnaEE-France, #DigitAg, Idex Paris-Saclay, UCA



# Objectives

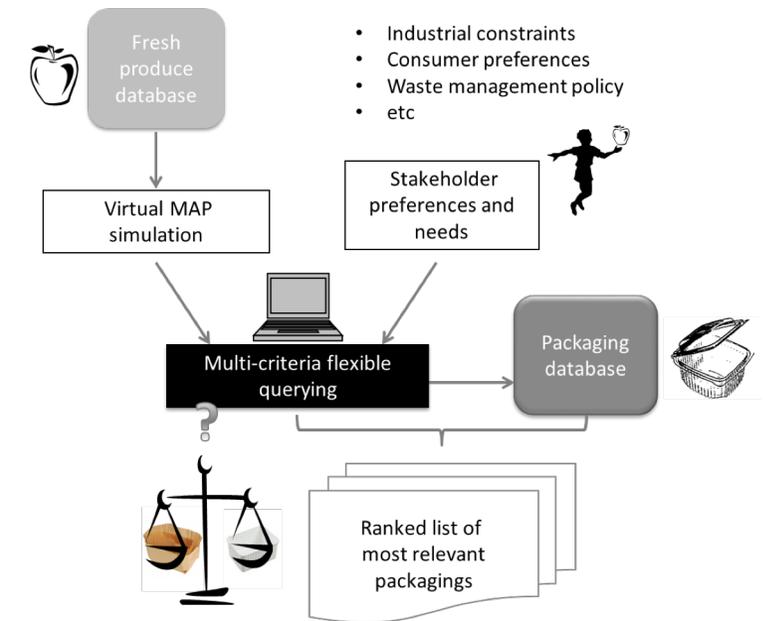
Create a framework to **turn agronomy and biodiversity data into –semantically described, interoperable, actionable, open–** and investigate the scientific methods and tools to exploit this knowledge for applications in agriculture and biodiversity sciences



# Ontology driven food packaging solutions

**Decision Support System help select the most appropriate food package for a given food considering multiple variables** (food respiration, temperature, material to use, etc.)

- Roadblock: get appropriate, clean data
  - Automatic data acquisition from tables in publications (@Web platform)
  - Based on ontology annotation (MAP-OPT ontology) and semantic web technologies
- What's done in D2KAB?
  - New methods and tools to manage constraints to automatically analyze the quality of annotated data
  - Apply SPARQL Template Transformation Language (STTL)



# Plant Health Bulletins augmented semantic reader (wine, cereals, market garden)

- Challenges: provide farmers and advisers with **contextualized information on pest attacks and direct access to connected data and informations**
- What: Bulletins de santé du végétal ← ontology → related datasets
- How:
  - Automated text annotation of growth stage and pest attacks: **Alvis Platform** (INRAE)
  - Semantic resources : **French Crop Usage, GECO, Taxref-Id, Serre des savoirs...**
  - Interconnect data sources through ontology alignment
- Roadblocks:
  - Valuable resources may not be available as RDF
  - Differences in granularity and points of views
- What's done in D2KAB:
  - Develop text annotation models better adapted to data integration
  - Make it easier by providing alignment services (**Agroportal**) and advices
  - Make more reference vocabularies available as RDF resources

**BULLETIN DE SANTÉ DU VÉGÉTAL BOURGOGNE**  
Grandes cultures n° 20 du 5 avril 2011

**A RETENIR CETTE SEMAINE**

**Colza**

- Grande majorité de parcelle avec des fleurs
- Avec le développement rapide des colzas le risque méligèthes s'estompe
- Signalement de quelques dégâts de charançons de la tige du colza
- Premiers charançons des sillques piégés mais sans incidence car stade de sensibilité des colzas pas atteint (à partir du stade G2)

**Réseau 2010-2011**

Cette semaine, 53 parcelles ont fait l'objet d'au moins une observation.

**Stade des colzas**

Rappel : un stade est atteint lorsque 50% des plantes sont à ce stade.

Les conditions climatiques estivales de la semaine dernière ont permis une accélération des stades et a fortiori dans les secteurs qui ont eu la chance de bénéficier des pluies.

- D1 boutons accolés encore cachés par les feuilles terminales	: 2%
- D2 inflorescence principale dégagée, boutons accolés, inflorescences secondaires visibles	: 8%
- E boutons séparés, les pédoncules floraux s'allongent en commençant par ceux de la périphérie	: 60%
- F1 premières fleurs ouvertes	: 20%
- F2 allongement de la hampe florale, nombreuses fleurs ouvertes	: 4%

**Méligèthes**

Risque table Risque fleur

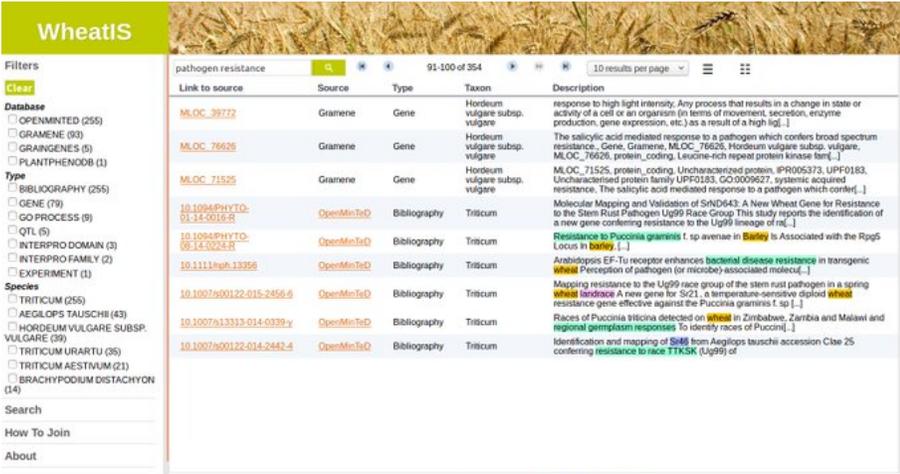
# Wheat phenotype data integration

## Phenotypes?

observable characteristics of an organism as a multifactorial consequence of genetic traits and environmental influences.

<https://www.biologyonline.com/dictionary/phenotype>

- Societal challenges: improved product quality (e.g. low gluten), plant health (with reduction of chemicals), crop yield, innovation (e.g. functional food)
- Researchers need a **federated access to multiple sources of experimental and observational data, publications**
- How: data integration based on ontologies alignment
  - **Wheat Crop Ontology-CO\_321** for data
  - **Wheat Trait Ontology** for publications
- Roadblocks: mapping of low level observation measures (e.g., weight of 1000 grains) and abstract qualitative properties (e.g., high yield)
- What's done in D2KAB:
  - Representation of alignments as formal identified objects relating 2 to N ontology classes, with properties (e.g. threshold or condition)
  - Publication of high value alignments made reusable



The screenshot shows the WheatIS database search results for the query 'pathogen resistance'. The interface includes a search bar, filters, and a table of results. The table has columns for Link to source, Source, Type, Taxon, and Description. The results list various genes and publications related to wheat pathogen resistance.

Link to source	Source	Type	Taxon	Description
<a href="#">MLOC_39772</a>	Gramene	Gene	Hordeum vulgare subsp. vulgare	response to high light intensity. Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a high light. [ ]
<a href="#">MLOC_76626</a>	Gramene	Gene	Hordeum vulgare subsp. vulgare	The salicylic acid mediated response to a pathogen which confers broad spectrum resistance. Gene, Gramene, MLOC_76626, Hordeum vulgare subsp. vulgare, MLOC_76626, protein, coding, L-leucine-rich repeat protein kinase fam. [ ]
<a href="#">MLOC_71525</a>	Gramene	Gene	Hordeum vulgare subsp. vulgare	MLOC_71525, protein, coding, Uncharacterised protein, [P0005373, UPF0183, Uncharacterised protein family UPF0183, GO:0009627, systemic acquired resistance. The salicylic acid mediated response to a pathogen which confer[ ]
<a href="#">10.1094/PHYTO-03-14-0224-F</a>	OpenMnTD	Bibliography	Triticum	Molecular Mapping and Validation of SNR643: A New Wheat Gene for Resistance to the Stem Rust Pathogen Ug99 Race Group This study reports the identification of a new gene conferring resistance to the Ug99 lineage of ra[ ]
<a href="#">10.1111/nph.13356</a>	OpenMnTD	Bibliography	Triticum	Arabidopsis EF-Tu receptor enhances bacterial disease resistance in transgenic wheat Perception of pathogen (or microbe)-associated molecu[ ]
<a href="#">10.1007/s00122-015-2456-6</a>	OpenMnTD	Bibliography	Triticum	Mapping resistance to the Ug99 race group of the stem rust pathogen in a spring wheat landrace A new gene for Sr21, a temperature-sensitive duplicated wheat resistance gene effective against the Puccinia graminis f. sp [ ]
<a href="#">10.1007/s13313-014-0339-y</a>	OpenMnTD	Bibliography	Triticum	Races of Puccinia triticina detected on wheat in Zimbabwe, Zambia and Malawi and regional genepattern responses to identity races of Puccin[ ]
<a href="#">10.1007/s00122-014-2442-4</a>	OpenMnTD	Bibliography	Triticum	Identification and mapping of SRR6 from Aegilops tauschii accession Clae 25 conferring resistance to race TTKSK (Ug99) of



# Pour en savoir plus sur les activités IA

