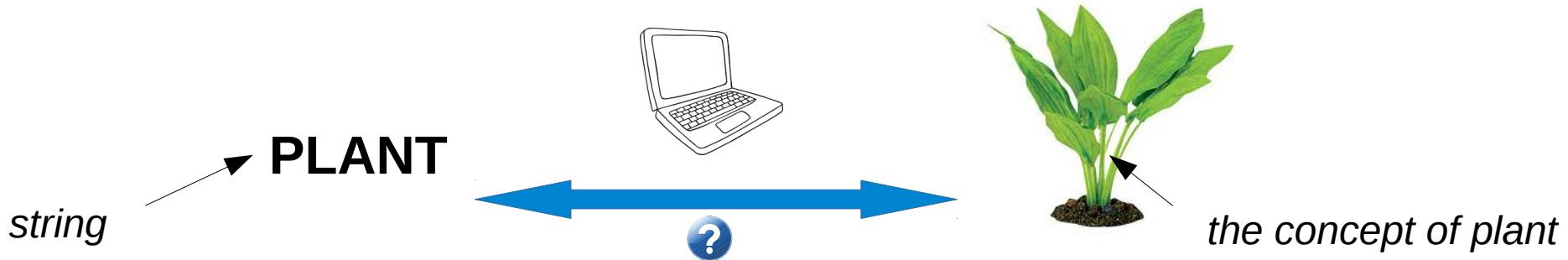


# *Gestion des connaissances dans le domaine agronomique*

Nordine EL Hassouni 31 janvier 2017

# Défis de l'intégration des données

- ▶ Augmentation du volume et de la variété des données biologiques et agronomiques
- ▶ Manque d'approches efficaces pour intégrer les données: un écart entre les données et les connaissances



- ▶ FAIR principle : data should be **F**indable **A**ccessible **I**nteroperable **R**e-usable

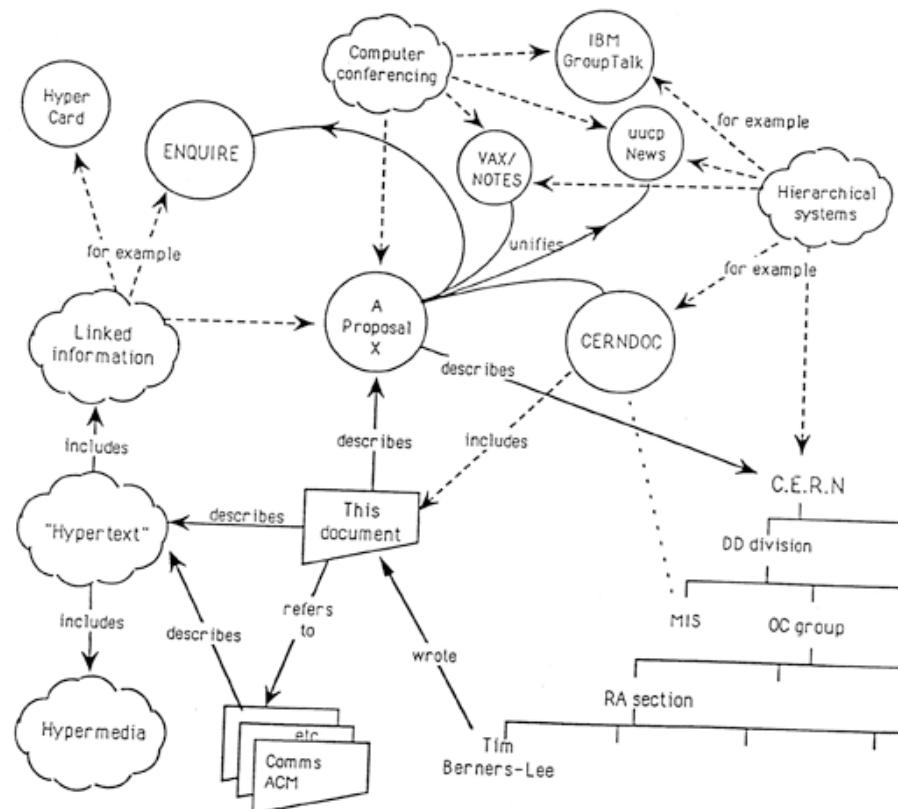
Comment gérer et extraire des informations significatives ?

## Information Management: A Proposal

### Abstract

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.

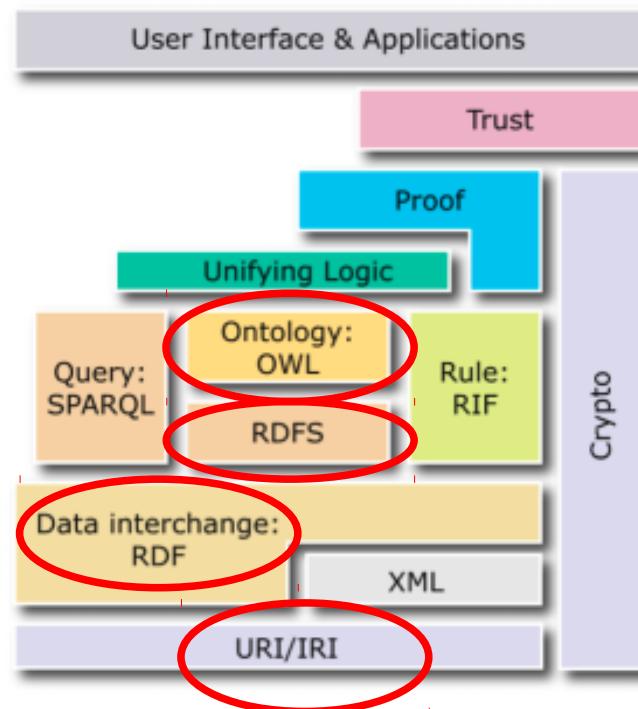
Keywords: Hypertext, Computer conferencing, Document retrieval, Information management, Project control



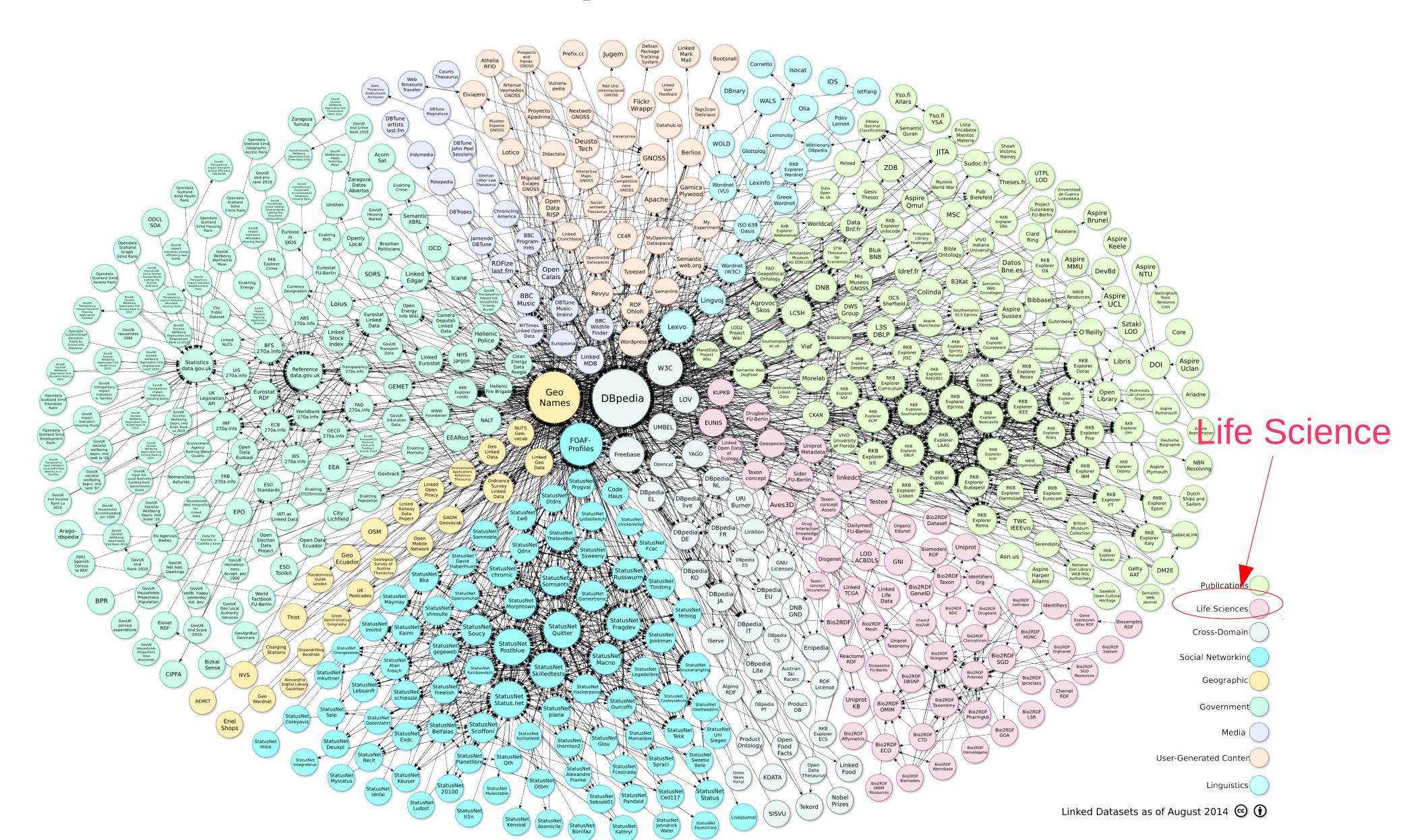


# Semantic Web Challenge

- ▶ Rendre les données intelligibles et interprétables par les machines
- ▶ Basé sur des formats courants et standards
  - Semantic web langages : rdf(s), owl, skos
  - Modélisation du savoir: ontologies, vocabulaires, terminologies

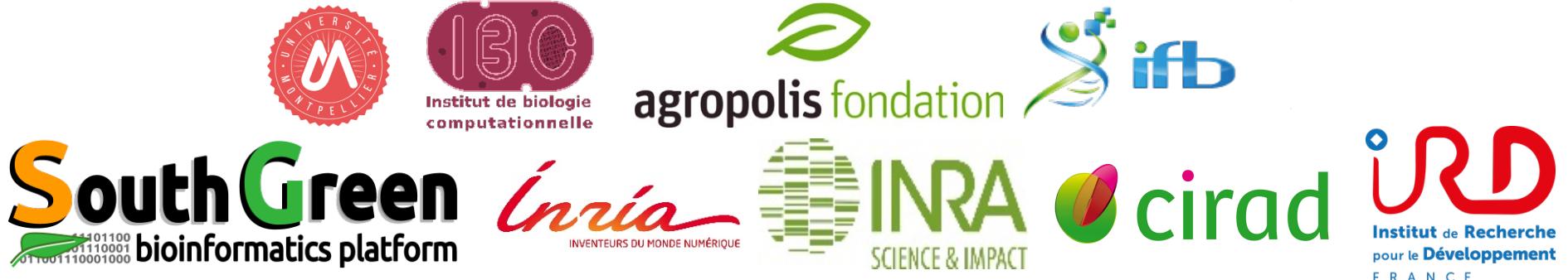


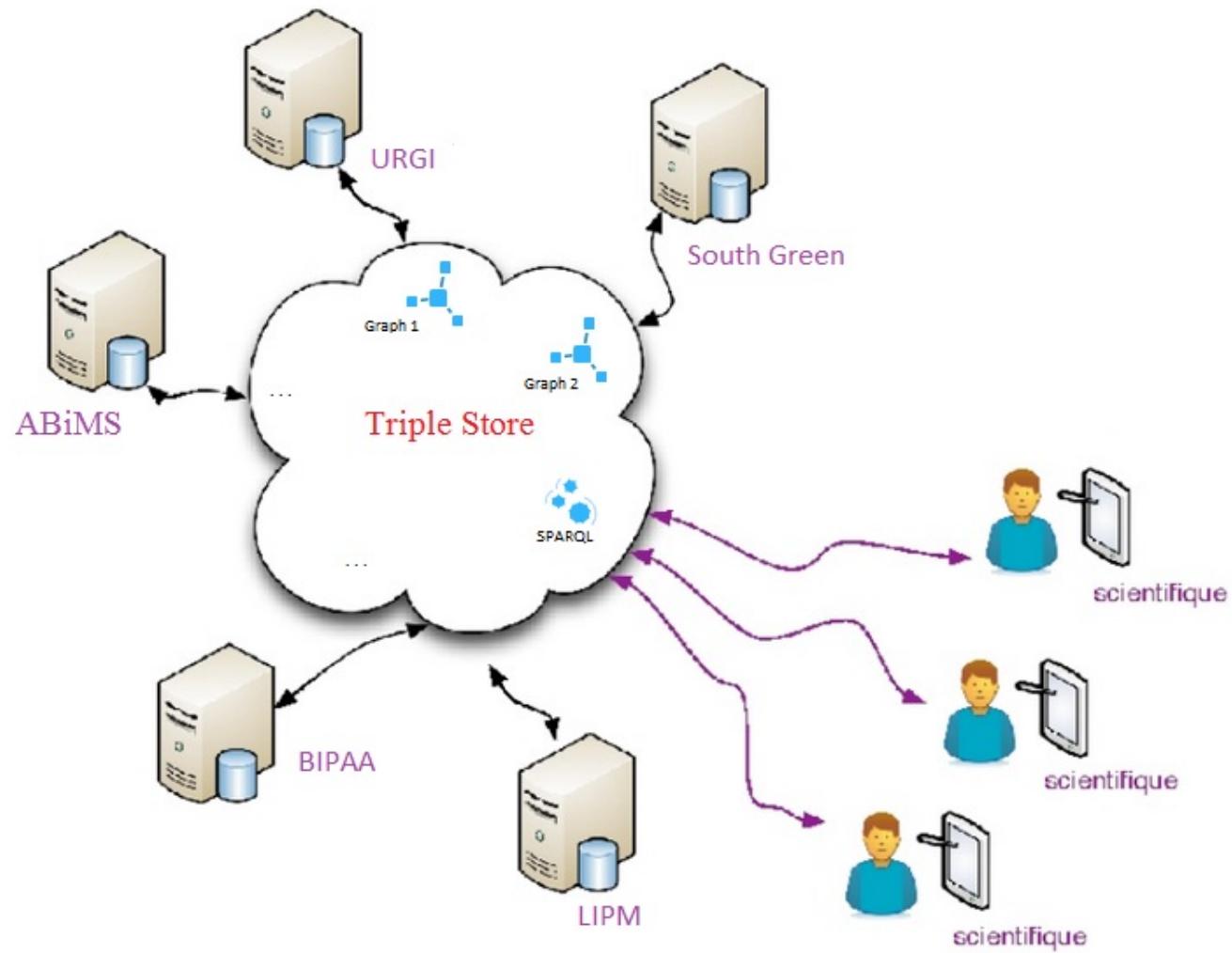
# Linked Open Data



Linked Datasets as of August 2014

# AgroLD : The Agronomic Linked Data project





Le même type de données (par exemple SNP , collections génétiques , cartes génétiques ) sont souvent représentées différemment dans différentes bases de données.

# *Knowledge management in the Agronomic domain*

## AgroLD

RDF knowledge base that integrates data about plants

### Agronomic Linked Data (AgroLD)

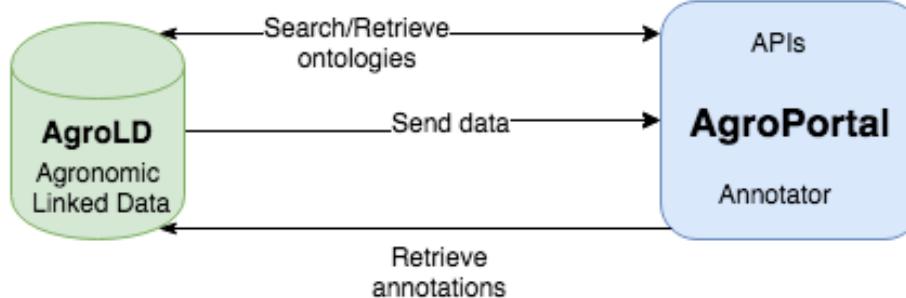
Home | Search | Help | About | Please send us your feedback! ★

#### The Agronomic Linked Data (AgroLD) Project

At the Institute of Computational Biology (IBC), we are involved in developing methods to aid data integration and knowledge management within the plant biology domain to improve information accessibility of heterogeneous data. Among others, a solution for the data integration challenges is offered by the Semantic Web technologies. The semantic web has emerged as one of the most promising solutions for high scale integration of distributed resources. This is made possible by a stack of technologies such as the Resource Description Framework (RDF), RDF Schema (RDFS), Web Ontology Language (OWL) and the SPARQL Query Language (SPARQL) proposed by the World Wide Web Consortium (W3C). RDF forms the basis of the stack allows modeling information as a directed graph composed of triples that can be queried using SPARQL.

AgroLD is a RDF knowledge base that consists of data integrated from a variety of plant resources and ontologies. The aim of the Agronomic Linked Data (AgroLD) project is to provide a portal for bioinformaticians and domain experts to exploit the homogenized data models towards efficiently generating research hypotheses.

- Quick Search**  
Search with keywords and browse AgroLD Knowledge Base
- Advanced Search**  
Search with keywords, browse, and get answers to some biological questions
- Explore Relationships**  
Search easily existing relationships between entities
- SPARQL Query Editor**  
Edit and submit your SPARQL Queries to the sparql endpoint of AgroLD located at <http://volvestre.cirad.fr:8890/sparql>



## AgroPortal

An ontology repository for the agronomic domain

AgroPortal

Browse | Search | Mappings | Recommender | Annotator | Projects | Sign In | Help | Feedback | Cite Us | Download AgroPortal ontology (December 2016) | Issue tracking on GitHub

Please 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 and annotations, and edit existing ones.

Search all ontologies | Find an ontology | Search resources

Enter concept, e.g. Melonoma | Enter ontology name, e.g. NGI Thesaurus | Enter a concept, e.g. Melonoma | Search | Explore | Search | Advanced Search | Browse Ontologies | Advanced Resource Search

Ontology Visits (December 2016)

Gene Ontology (GO)	48
Sequence Types and Features Ontology (SO)	21
Cell Ontology (CL)	17
Plant Ontology (PO)	16
Wheat Trait Ontology (WHEATPHENOTYPE)	12
More	

Latest Notes

**Un peu d'histoire (Banana Anatomy)**  
about 3 years ago by junquet  
Inflorescence est un mot d'origine latine qui signifie "fleurir". Il est le même en français et ...

**Can measure by mapped to another ontology ? (Biorefinery)**  
about 1 year ago by junquet  
Such as Unit of Measurement ?

**Is spadice a kind of inflorescence for banana? (Banana Anatomy)**  
about 1 year ago by junquet  
Can we consider spadice an appropriate inflorescence for banana?

Statistics

Ontologies	53
Classes	1.169.623
Projects	14
Users	51

Latest Mappings

Season (CO\_715) <-> Season (<http://dbpedia.org>)  
External Mapping 01/08/2016 by junquet

Season (CO\_711) <-> season (<http://data.biopontology.org>)  
External Mapping ncbo 01/08/2016 by junquet

plant organ (PO) <-> Plant\_organ (<http://dbpedia.org/ontology>)  
External Mapping 11/30/2015 by junquet

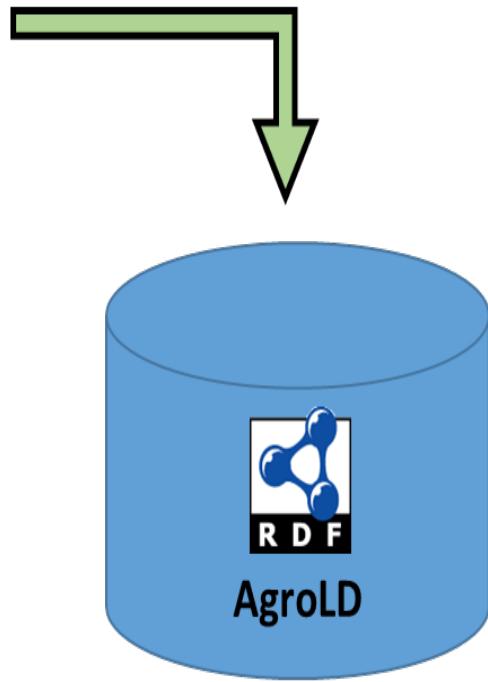
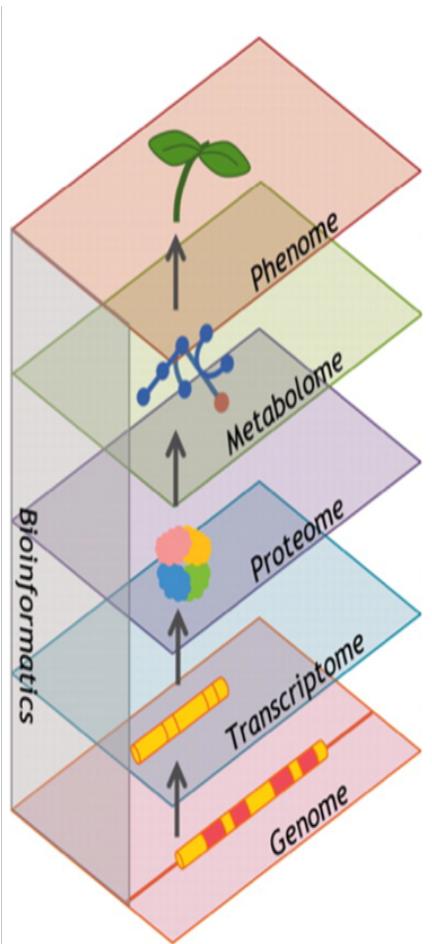
Plant (GTB) <-> Plant (<http://dbpedia.org/ontology>)  
External Mapping 11/06/2015 by junquet

Tissue (BT) <-> Tissue (CL)  
REST Mapping 09/24/2013 by junquet

Slices

Crop Ontology Curation Tool (crop)  
INRA Linked Open Vocabularies (livrira)  
The Agronomic Linked Data (AgroLD) (agrold)  
RDA Wheat Data Interoperability working group (wheat)  
Exclusive AgroPortal ontologies (exclu)

# AgroLD : Agronomic Linked Data



- ▶ Triple Store de connaissances qui intègre des données à partir d'une variété de ressources végétales
- ▶ Intégrer l'information à différents niveaux
- ▶ Extensibilité facile

# RDF

Formalisme pour échanger des descriptions de ressources

- numériques
- physiques
- conceptuelles

Ressources identifiées par un URI

- Uniform Resource Identifier

URI peut être un URL

- Uniform Resource Locator

Un URL peut être déréférencé via le protocole HTTP

- Retourne une description de la ressource en RDF

# Triplet RDF

Resource Property Value

Nordine, travaille à, Montpellier

Subject Predicate Object

Montpellier, est une, Technopole

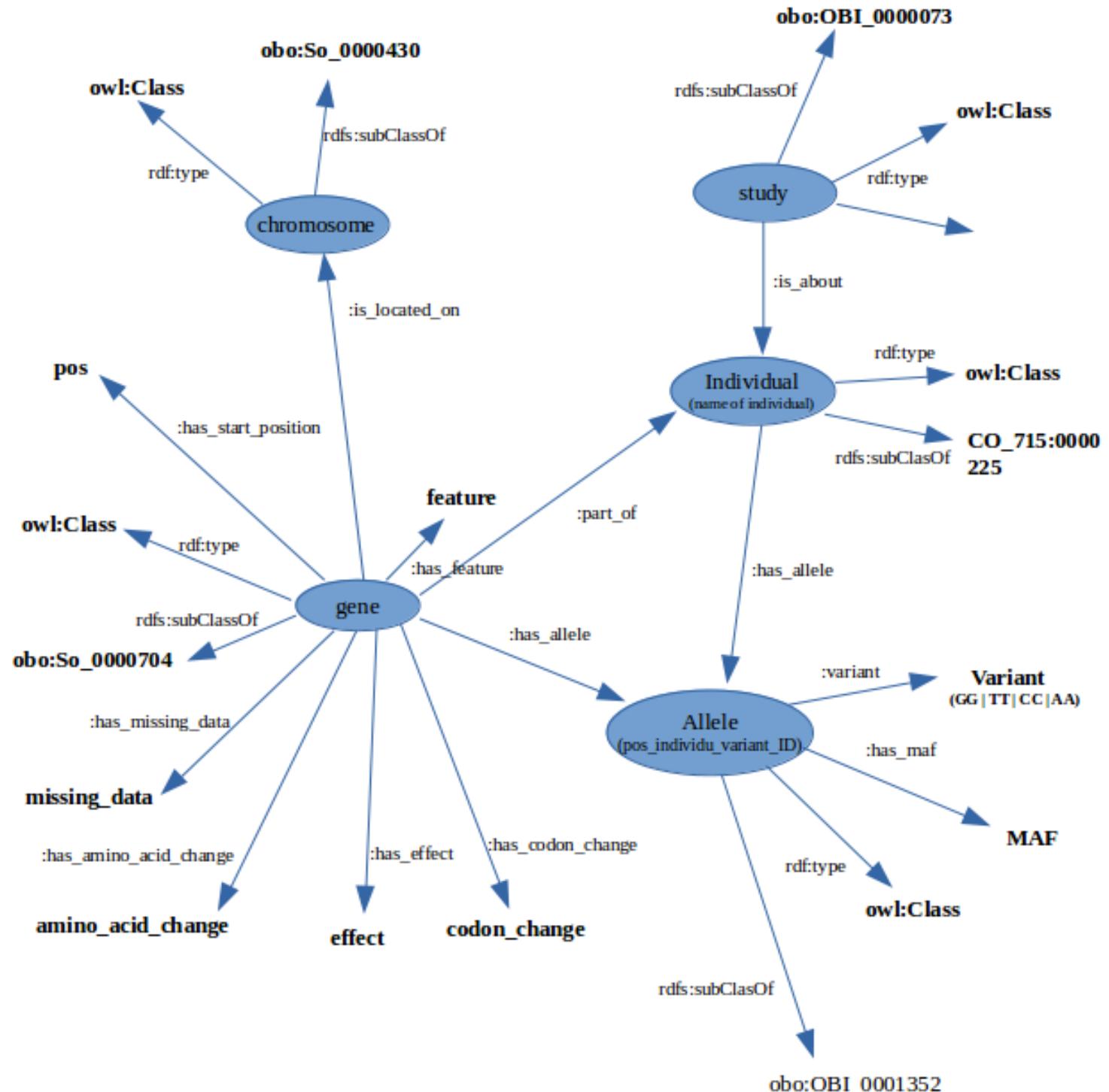
# Graph RDF



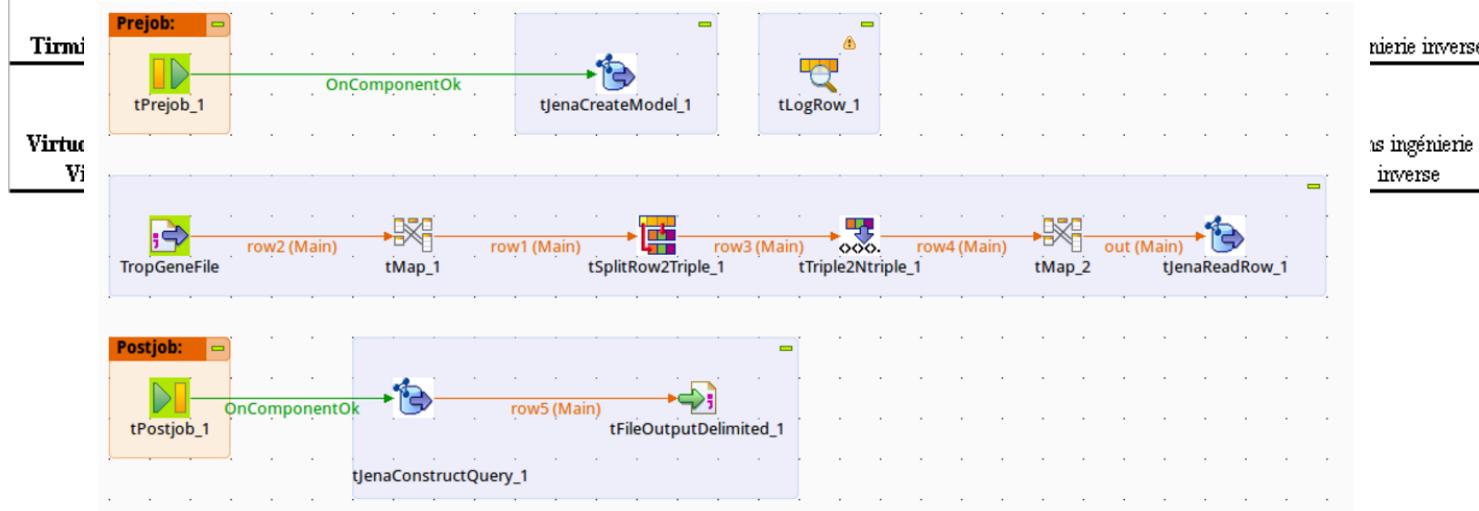
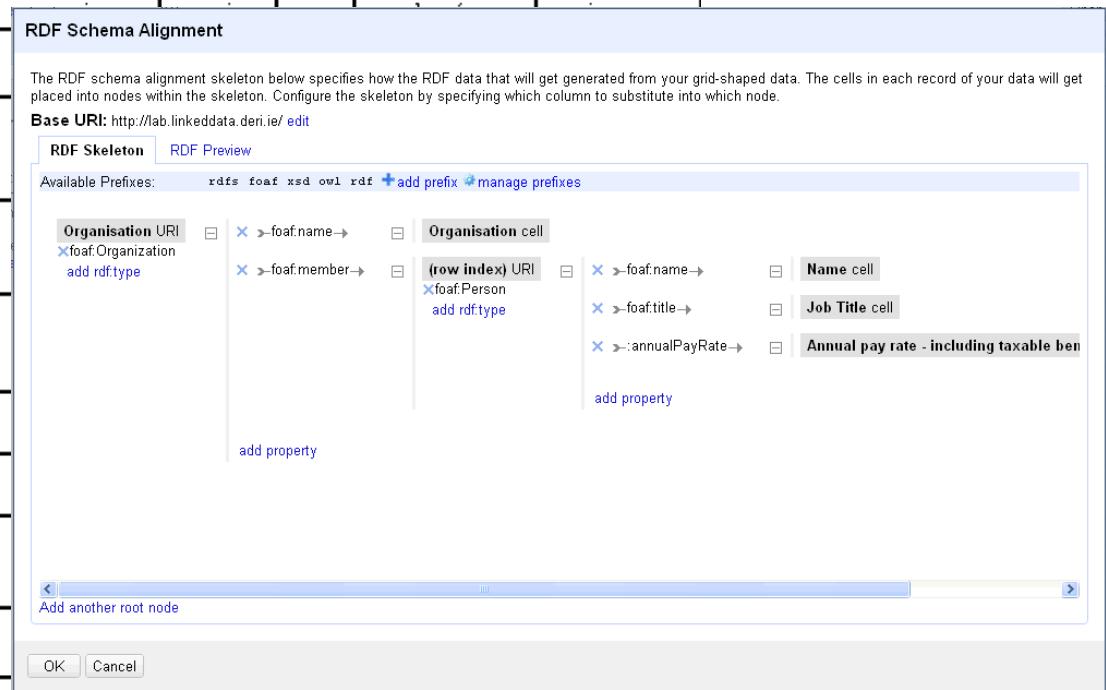
# Démarche

HapMap format

rs#	alleles	chrom	pos	gene	feature	effect	codon_change	amino_acid_change	MAF	missing_data	AZUCENA	BULUPANDAK	GIZA171	IAC165	KHAODAM	M202	MOROBEREKAN
Chr1:1492913	T/C	Chr1	1492913	intergenic:	#	#	#	#	4.2%	0.0%	TT	TT	TT	TT	TT	TT	TT
Chr1:1496524	T/A	Chr1	1496524	intergenic:	#	#	#	#	36.5%	0.0%	TT	TT	TT	AA	TT	TT	AA
Chr1:1509728	A/G	Chr1	1509728	LOC_Os01g03660:	exon	SYNONYMOUS_CODING	ccA/ccG	P/P	2.4%	0.0%	AA	AA	AA	AA	AA	AA	AA
Chr1:1515087	A/T	Chr1	1515087	intergenic:	#	#	#	#	42.5%	0.0%	AA	AA	AA	TT	AA	AA	TT
Chr1:1546579	C/T	Chr1	1546579	intergenic:	#	#	#	#	48.5%	0.0%	TT	TT	CC	TT	CC	TT	CC
Chr1:1551208	A/T	Chr1	1551208	intergenic:	#	#	#	#	47.9%	0.0%	AA	AA	TT	AA	TT	TT	AA
Chr1:1563026	T/A	Chr1	1563026	LOC_Os01g03740:	exon	SYNONYMOUS_CODING	gcA/gcT	A/A	46.7%	0.0%	TT	TT	AA	TT	AA	TT	AA
Chr1:1563029	A/T	Chr1	1563029	LOC_Os01g03740:	exon	SYNONYMOUS_CODING	gcT/gcA	A/A	46.7%	0.0%	AA	AA	TT	AA	TT	AA	TT
Chr1:1565548	T/A	Chr1	1565548	LOC_Os01g03740:	UTR_5_PRIME	#	#	#	3.0%	0.0%	AA	AA	AA	AA	AA	AA	AA
Chr1:1598133	T/G	Chr1	1598133	LOC_Os01g03820:	UTR_5_PRIME	NON_SYNONYMOUS_CODING	#	#	43.7%	0.0%	GG	GG	TT	GG	TT	TT	GG
Chr2:21833512	G/A	Chr2	21833512	intergenic:	#	#	#	#	4.8%	0.0%	AA	AA	AA	AA	GG	AA	AA
Chr2:21833524	G/A	Chr2	21833524	intergenic:	#	#	#	#	4.8%	0.0%	AA	GG	GG	GG	GG	GG	GG
Chr2:21860465	T/A	Chr2	21860465	LOC_Os02g36220:	exon	NON_SYNONYMOUS_CODING	cAg/cTg	Q/L	6.0%	0.0%	AA	AA	AA	AA	AA	AA	AA
Chr2:21887225	A/T	Chr2	21887225	LOC_Os02g36264:	exon	NON_SYNONYMOUS_CODING	cTg/cAg	L/Q	7.2%	0.0%	AA	TT	TT	TT	TT	TT	TT
Chr2:21914519	A/T	Chr2	21914519	intergenic:	#	#	#	#	3.0%	0.0%	TT	TT	TT	TT	AA	TT	AA
Chr2:21942006	T/A	Chr2	21942006	intergenic:	#	#	#	#	30.5%	0.0%	AA	AA	TT	AA	AA	TT	AA
Chr2:21966132	T/A	Chr2	21966132	intergenic:	#	#	#	#	9.0%	0.0%	AA	TT	TT	TT	TT	TT	TT



nom de l'approche	niveau d'automatisation	accessibilité des données	langage de mapping	langage de l'ontologie	réutilisation du vocabulaire	logiciel disponible	GUI	but principal	type d'approche
D2R MAP	Automatique / manuel	ETL	basé sur du XML	RDF				intégration de	Sans ingénierie
D2RQ / D2R Serveur	automatique / manuel	SPARQL / Web des	langage D2RQ basé sur RDF	RDFS					
MAPONTO	semi automatique	-	basé sur XML	OWL DL					
MASTRO / ODBA	manuel	SPARQL	représentation spécifique à l'outil	DL-Lite JOWL					
MOMIS	semi automatique	ETL / langage de requête OQL <sub>DB</sub>	-	ODL <sub>DB</sub>					
Ontegrate	Semi-automatique	Web-PDDL	-	Web-PDDL					
R2O	semi automatique	ETL	Langage R2O	OWL DL					
R2O + ODEMMaster	manuel	ETL / langage de requête ODEMLQ	langage R2O	OWL					
Relational.OWL	automatique	ETL / SPARQL	-	OWL Full					
ROSEX	automatique	SPARQL	ontologie spécifique	OWL DL	non	non	non	accès aux données basé sur l'ontologie	+ ingénierie inverse



# Python

The screenshot shows the PyCharm IDE interface with the following details:

- Project Structure:** On the left, the project tree shows a hierarchy of files and folders under "riceKB". Key files include "phenobrapiToRDF.py", "TropgeneParser.py", and "os\_japonicaModel.py".
- Code Editor:** The main window displays the content of "phenobrapiToRDF.py". The code is a Python script that uses RDFWriter to generate RDF triples. It includes logic for handling study data, observation data, and phenotype data.
- Run Tab:** At the bottom, the "Run" tab is active, showing the command used to run the script: "/home/elhassouni/Bureau/virtual\_env\_test/gff2rdf\_env/bin/python2.7 /media/elhassouni/donnees/Noeud-plante-projet/workspace/AgroLD/AgroLD\_ETL/riceKB/phenobrapiToRDF.py".
- Output Tab:** The "Output" tab shows the generated RDF triples. The first few lines of output are:

```
observation_data:0774de21-b3dd-4288-bfc5-e03b9816bf35
  rdf:type owl:Class ;
  agrold_vocabulary:has_season "2014" ;
  agrold_vocabulary:has_observation_value "3,5" ;
  agrold_vocabulary:has_observation_time_stamp "None" ;
  agrold_vocabulary:has_observation_variable_id "observation_data::SCNO:00000011" .
```

```
observation_data:9802fddd-bb4a-4679-9692-6af14526387d
  rdf:type owl:Class ;
  agrold_vocabulary:has_season "2014" ;
  agrold_vocabulary:has_observation_value "2" ;
  agrold_vocabulary:has_observation_time_stamp "None" ;
  agrold_vocabulary:has_observation_variable_id "observation_data::SCNO:00000008" .
```

```
observation_data:cb094398-a646-4453-b900-350936ee6107
  rdf:type owl:Class ;
  agrold_vocabulary:has_season "2014" ;
  agrold_vocabulary:has_observation_value "None" ;
```

# Linking Information

