

L'écosystème des graphes de connaissances D2KAB et son interrogation

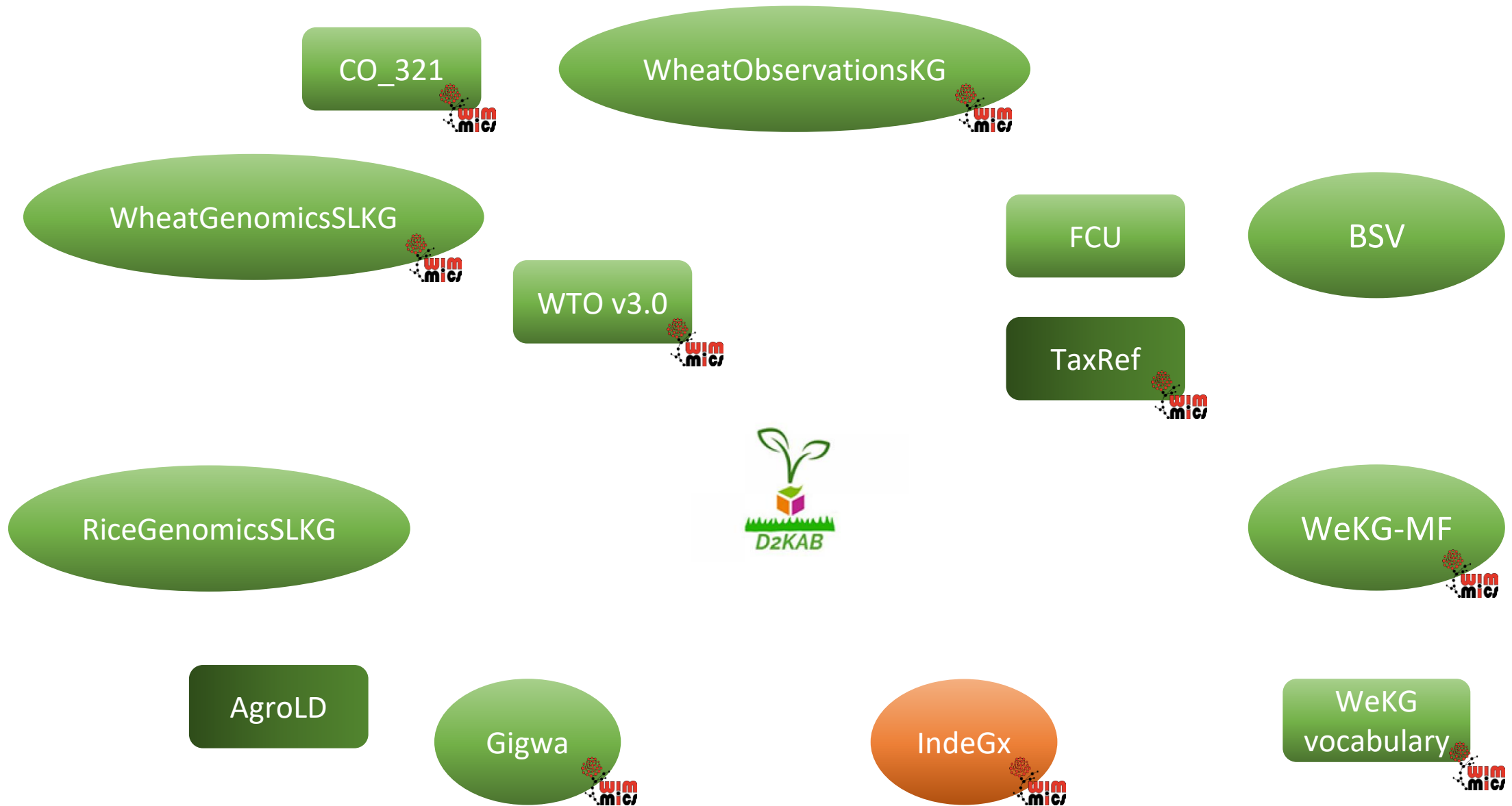
Catherine Faron
Université Côte d'Azur, I3S, CNRS, Inria



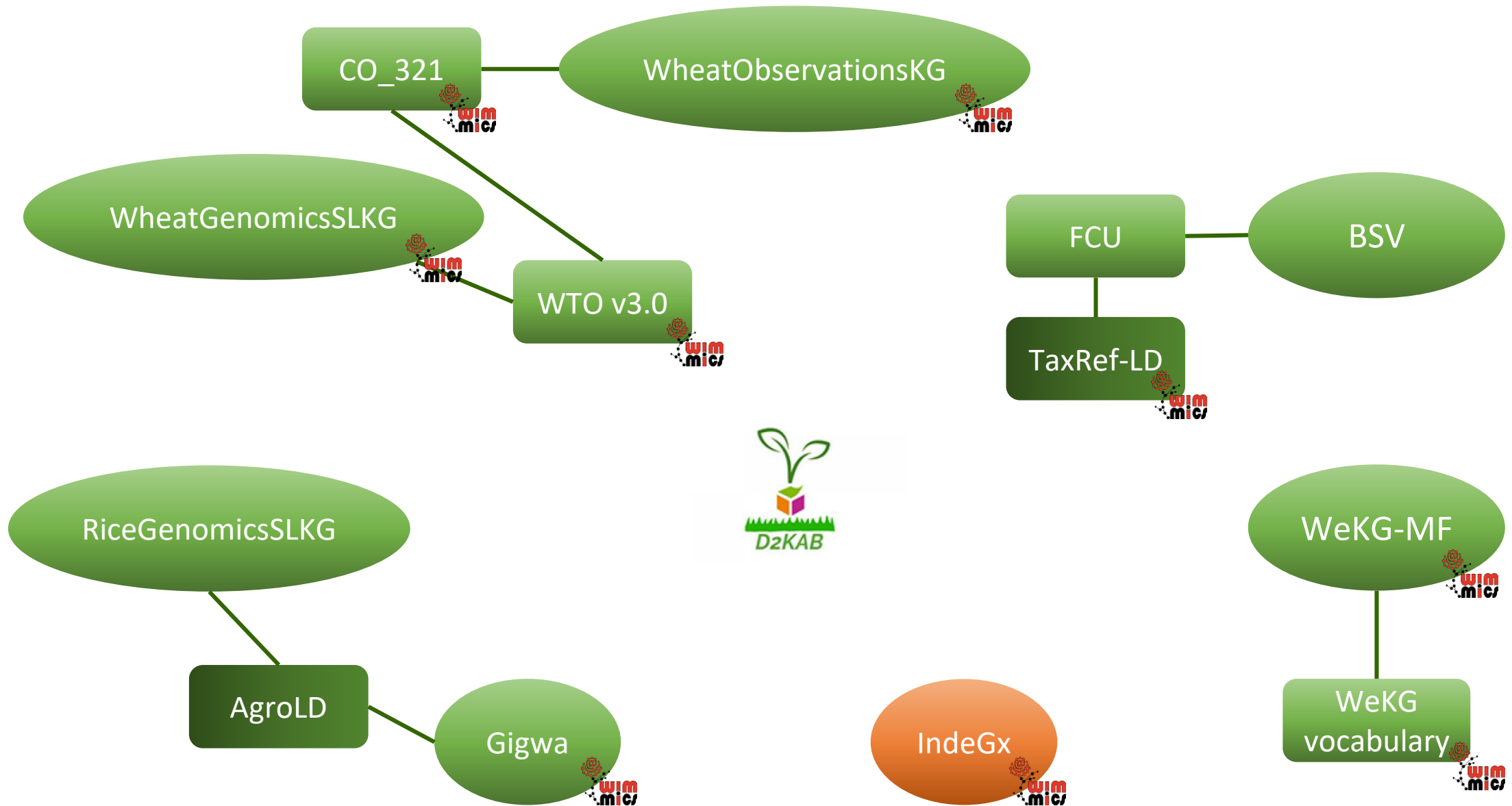
séminaire résidentiel Inrae *Semantic Linked Data* édition 2023



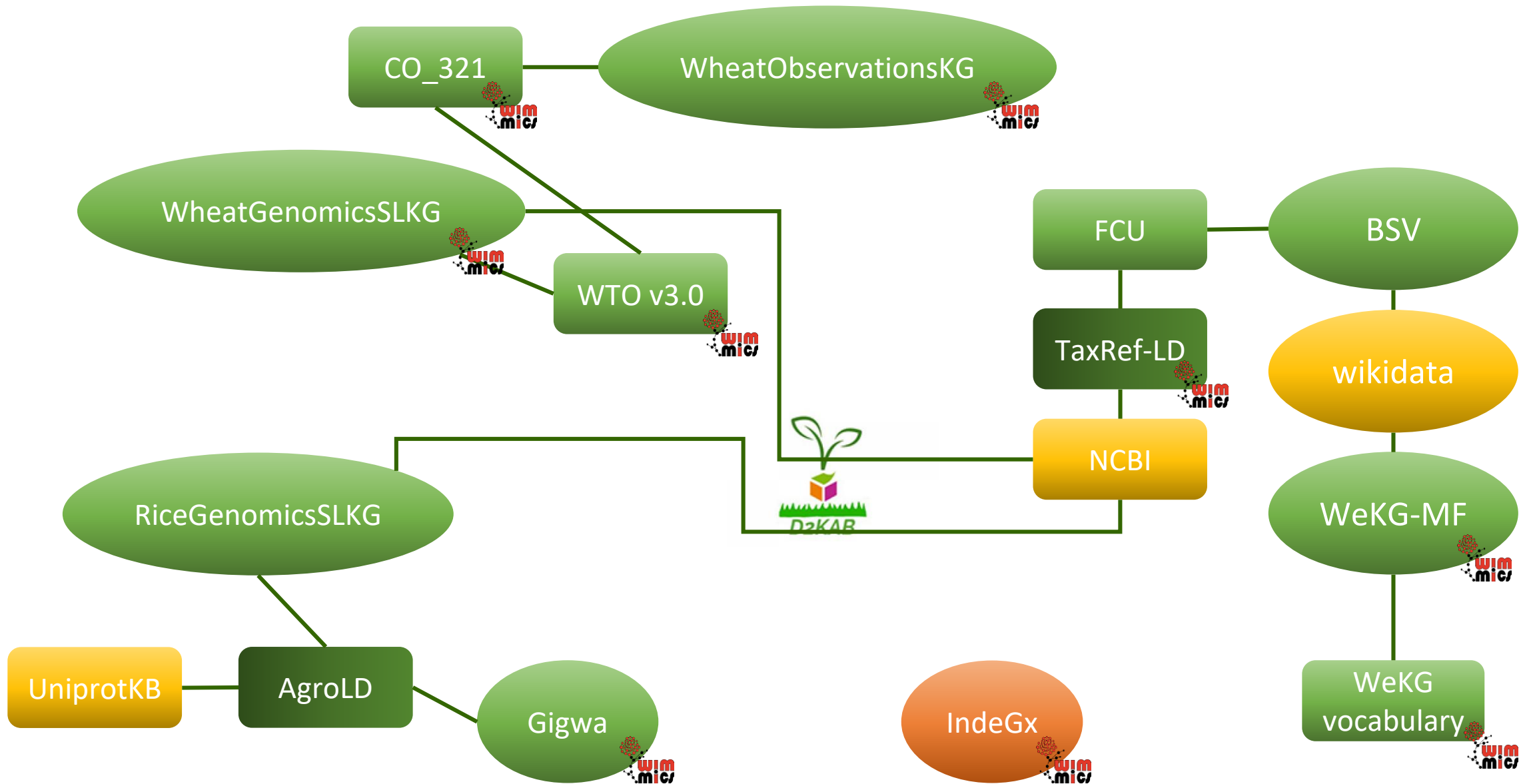
L'écosystème des graphes de connaissances D2KAB



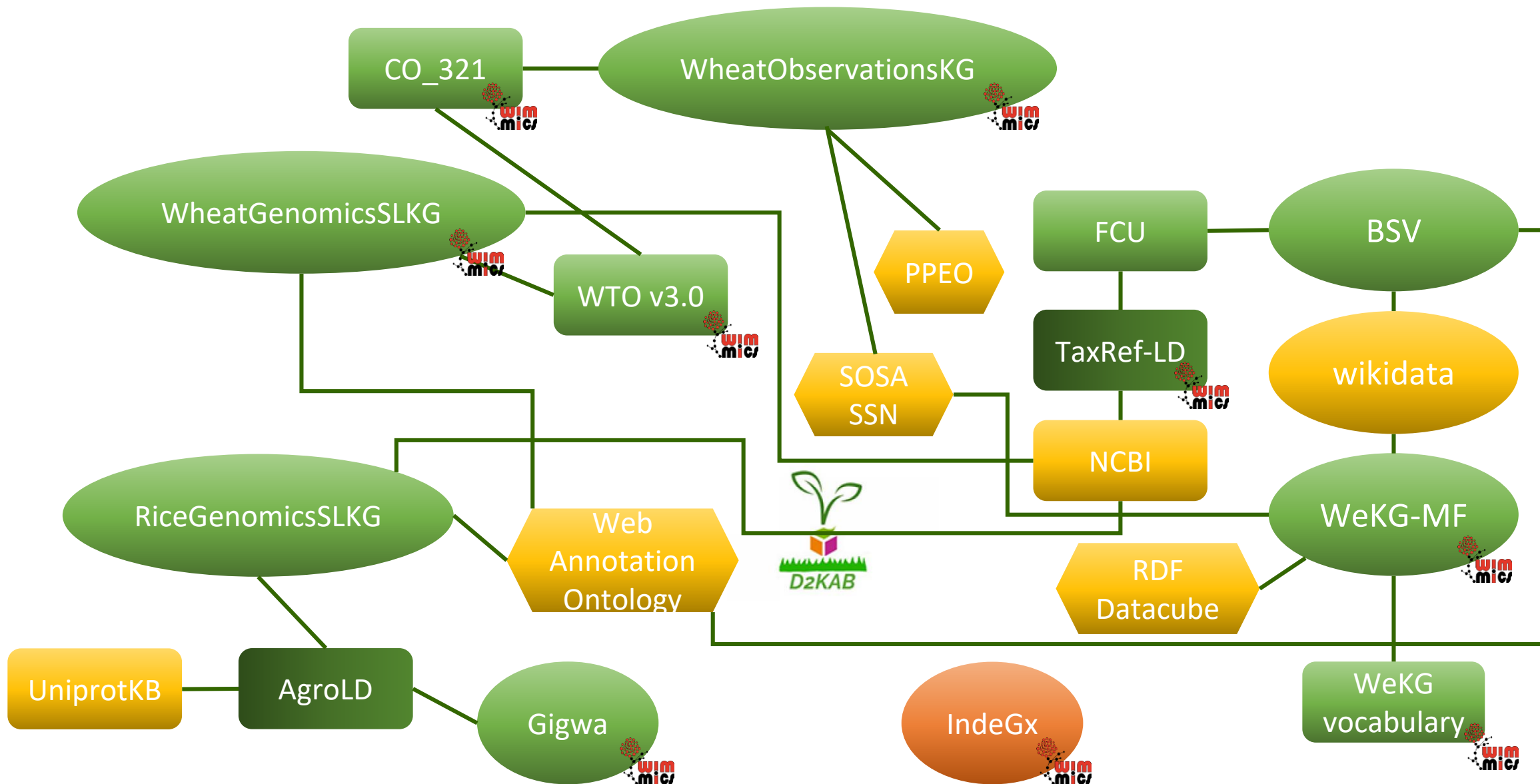
L'écosystème des graphes de connaissances D2KAB



L'écosystème des graphes de connaissances D2KAB



L'écosystème des graphes de connaissances D2KAB



L'écosystème des graphes de connaissances D2KAB

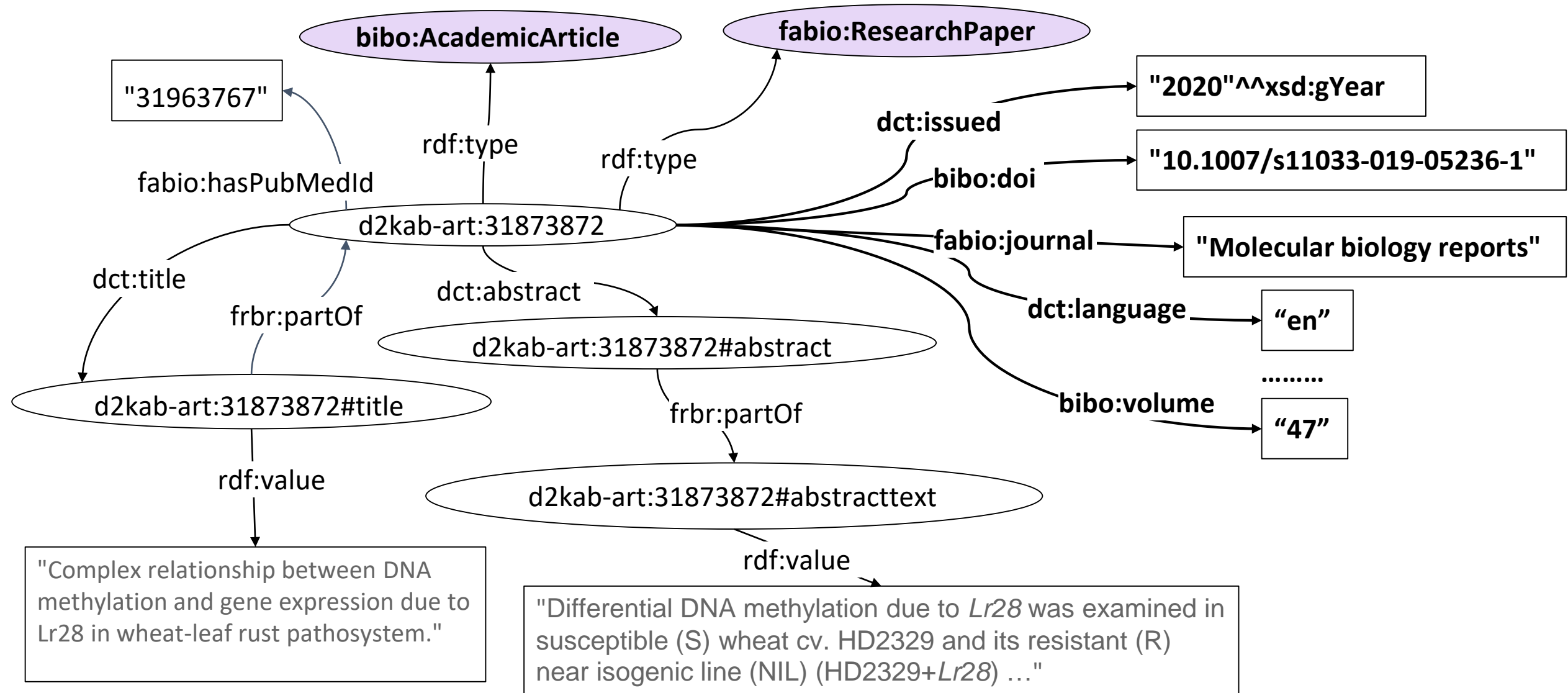
- Des KG liés entre eux car reposant sur des ontologies communes
- Des KG liés entre eux car reposant sur des thesauri communs
- Des KG liés au LOD
- Un sous-graphe du LOD



Possibilité d'interroger conjointement plusieurs KG

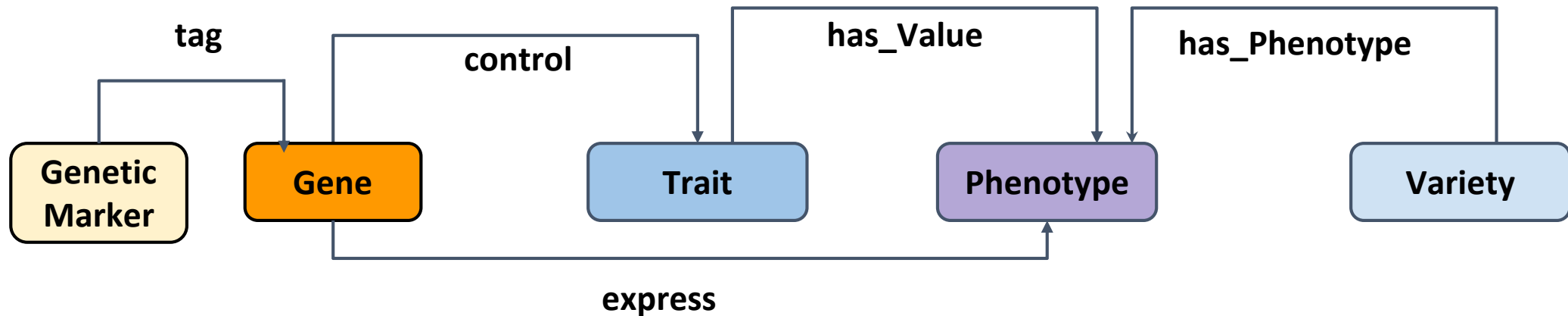
WheatGenomicsSLKG

WheatGenomicsSLKG



WheatGenomicsSLKG: questions de compétence

- Retrieve the **genes** that are the most frequently mentioned proximal to a given **trait**, e.g., resistance to leaf rust
- Retrieve publications in which **genes** are mentioned proximal to a specific **taxon**, e.g., *Puccinia triticina*
- Retrieve publications in which **genetic markers** are mentioned related to different **genes** which are mentioned proximal to a given **trait**, e.g., *resistance to Stripe rust*.
- Retrieve publications in which **genes** are mentioned proximal to wheat **varieties** and **traits** from a specific class, e.g., all wheat traits related to resistance to fungal pathogens



WheatGenomicsSLKG: requêtes SPARQL

Retrieve the **genes** that are the most frequently mentioned proximal to a given **trait**, e.g., resistance to leaf rust

```
SELECT ?GeneName (count(distinct ?paper) as ?NbOcc)
WHERE {
  GRAPH <http://ns.inria.fr/d2kab/graph/wheatgenomicsslkg> {
    ?a1 a oa:Annotation; oa:hasTarget [ oa:hasSource ?source1 ] ;
      oa:hasBody ?WTOtraitURI .
    ?source1 frbr:partOf+ ?paper .
    ?a2 a oa:Annotation ;
      oa:hasTarget [ oa:hasSource ?source2 ] ;
      oa:hasBody [ a d2kab:Gene; skos:prefLabel ?GeneName ] .
    ?source2 frbr:partOf+ ?paper . ?paper a fabio:ResearchPaper . }
  GRAPH <http://ns.inria.fr/d2kab/ontology/wto/v3> {
    ?WTOtraitURI skos:prefLabel "resistance to Leaf rust" . }
}
GROUP BY ?GeneName
HAVING (count(distinct ?paper) > 1)
ORDER BY DESC(?NbOcc)
```

GeneName	NbOcc
Lr34	34
Lr10	33
Lr1	33
Lr	24
Lr26	22
Lr24	20
Lr9	19
Lr28	19
Lr21	19
Lr16	18
Lr18	15
Lr13	15
Lr19	14
Lr3	12
Lr46	12
Lr37	11

WheatGenomicsSLKG: statistiques

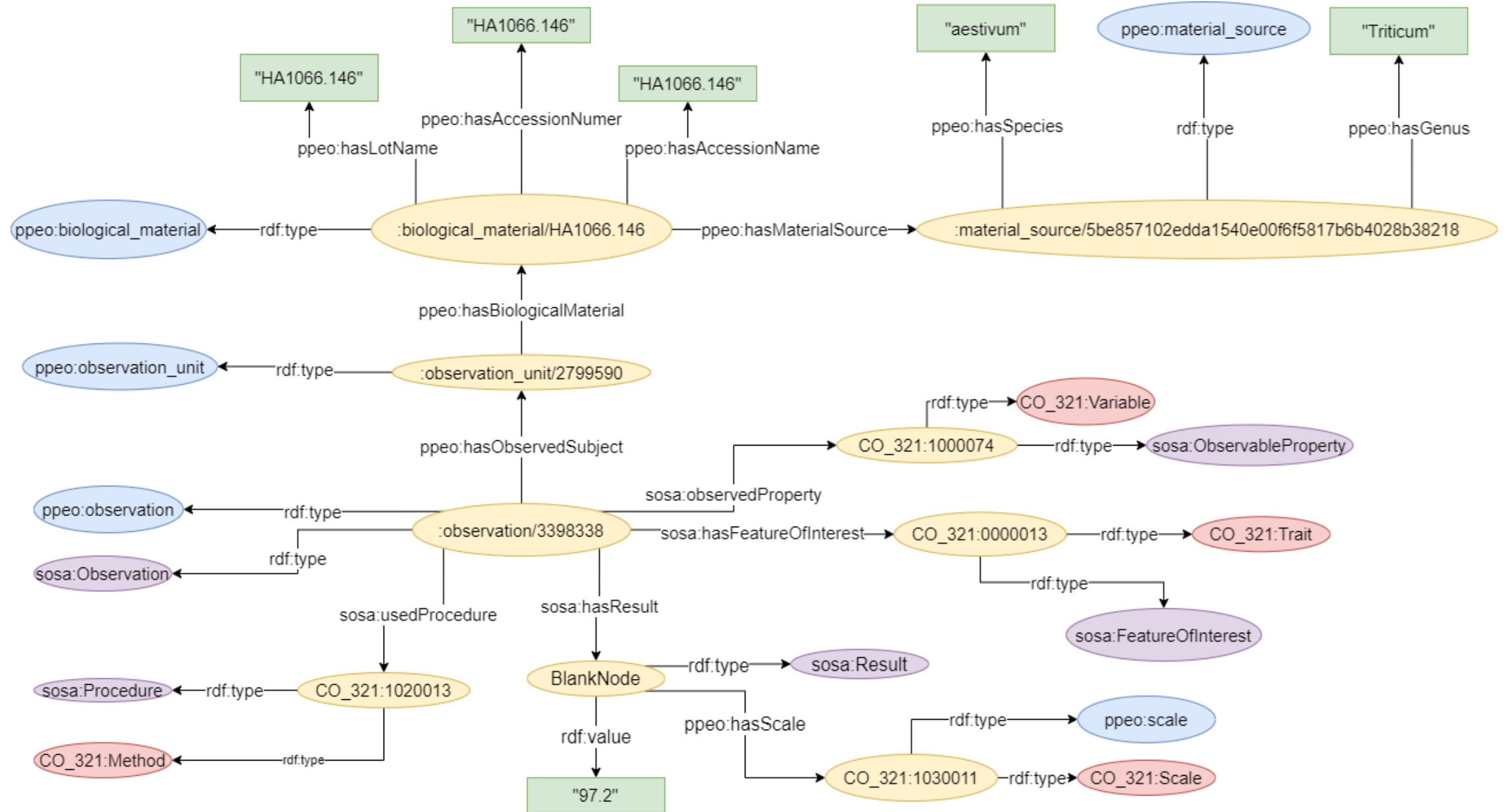
- 8 496 publications scientifiques
- NER&L basé sur AlvisNLP (BiblioMe)
 - 88 880 annotations
 - 4 318 entités nommées (5 classes)
 - 162 relations de type 'variety_has_phenotype'

WheatGenomicsSLKG et RiceGenomicsSLKG

- SPARQL endpoint:
<http://d2kab.i3s.unice.fr/sparql>
- Github WheatGenomicsSLKG :
<https://github.com/Wimmics/WheatGenomicsSLKG>
- Github RiceGenomicsSLKG: <https://github.com/ANR-DIG-AI/RiceGenomicsSLKG>
- PubMed SPARQL microservice:
https://sparql-micro-services.org/service/pubmed/getArticleByPMId_sd/
- Publication:
IC 2022: [hal-03889968](https://hal.archives-ouvertes.fr/hal-03889968)

WheatObservationsKG

WheatObservationsKG



WheatObservationsKG: questions de compétence

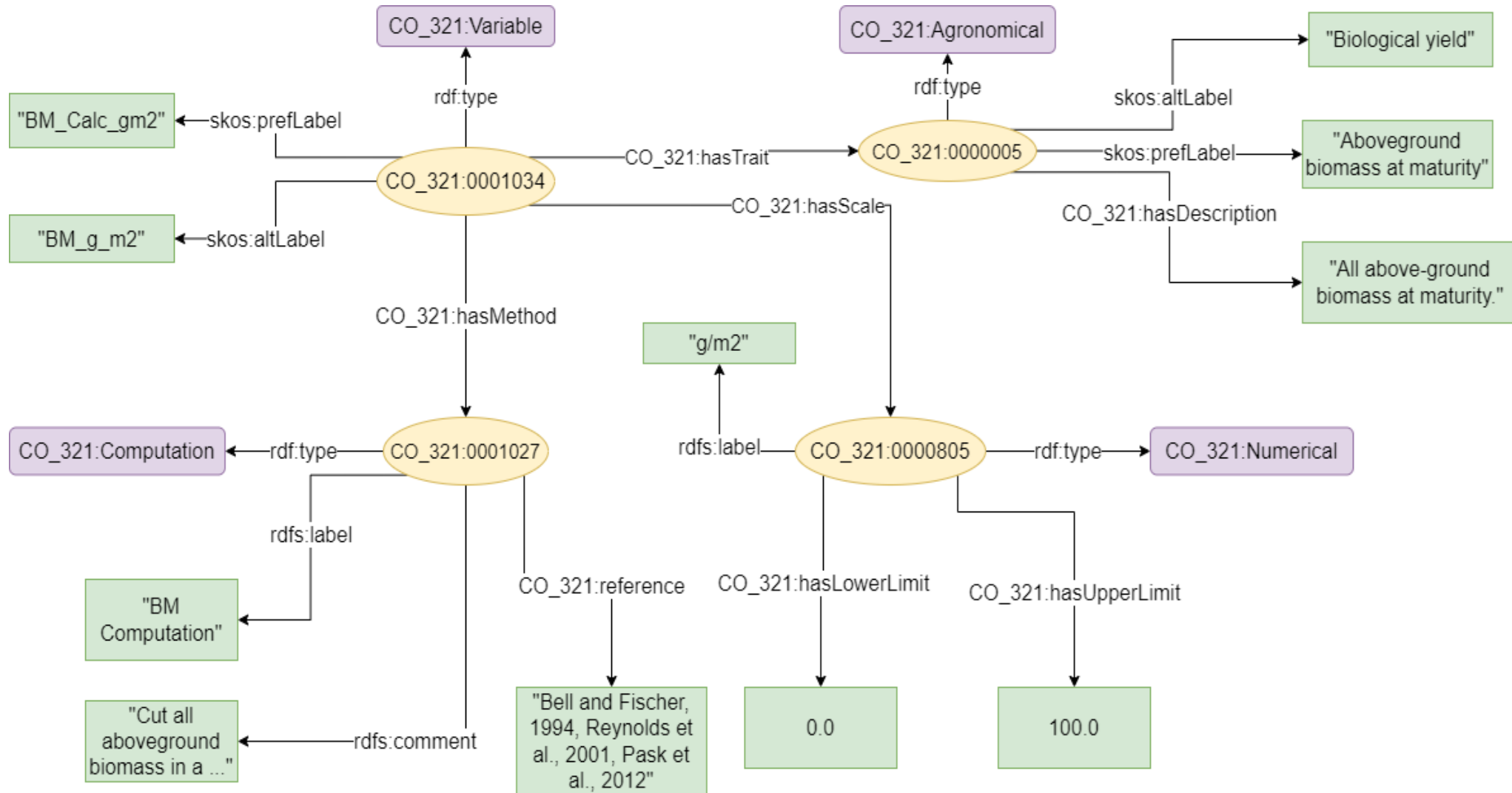
- Quelles sont les variables observées dont le trait est une maladie (eg. BioticStress), par année et lieu pour les accessions (plant material).
- Quelles sont les studies pour lesquelles on observe des mesures pour les variables “frost” et “lodging”.
- Quelles sont les valeurs observées des traits de qualité (eg. dureté du grain CO_321:0000072 et qualité boulangère CO_321:0500001) sur l’ensemble du dataset.

WheatObservationsKG: requêtes SPARQL

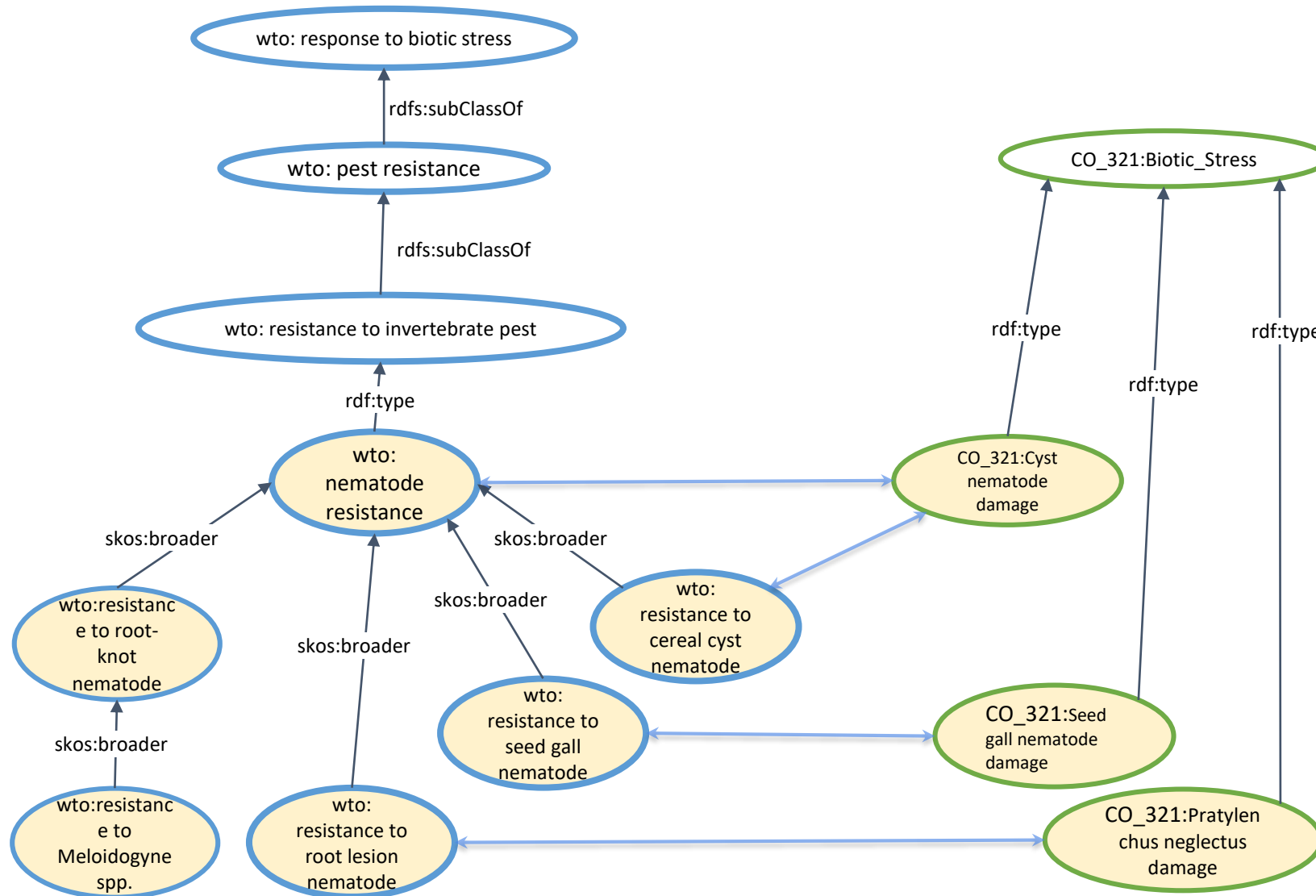
Quelles sont les **studies** pour lesquelles on observe des mesures pour la **variable** “Frost susceptibility”

```
SELECT DISTINCT ?study ?variableName WHERE {  
  ?study a ppeo:study; ppeo:hasEndDateTime ?date.  
  ?observationUnit a ppeo:observation_unit;  
  ppeo:hasBiologicalMaterial ?biologicalMaterial; ppeo:partOf ?study.  
  
  ?biologicalMaterial a ppeo:biological_material;  
    ppeo:hasAccessionName ?accession.  
  
  ?observation a sosa:Observation;  
    sosa:observedProperty ?variable;  
    ppeo:hasObservedSubject ?observationUnit.  
  
  ?variable a CO_321:Variable; skos:altLabel "Frost susceptibility".  
}
```

CO_321: thesaurus SKOS



WTO et CO_321



WTO v3.0: requêtes SPARQL

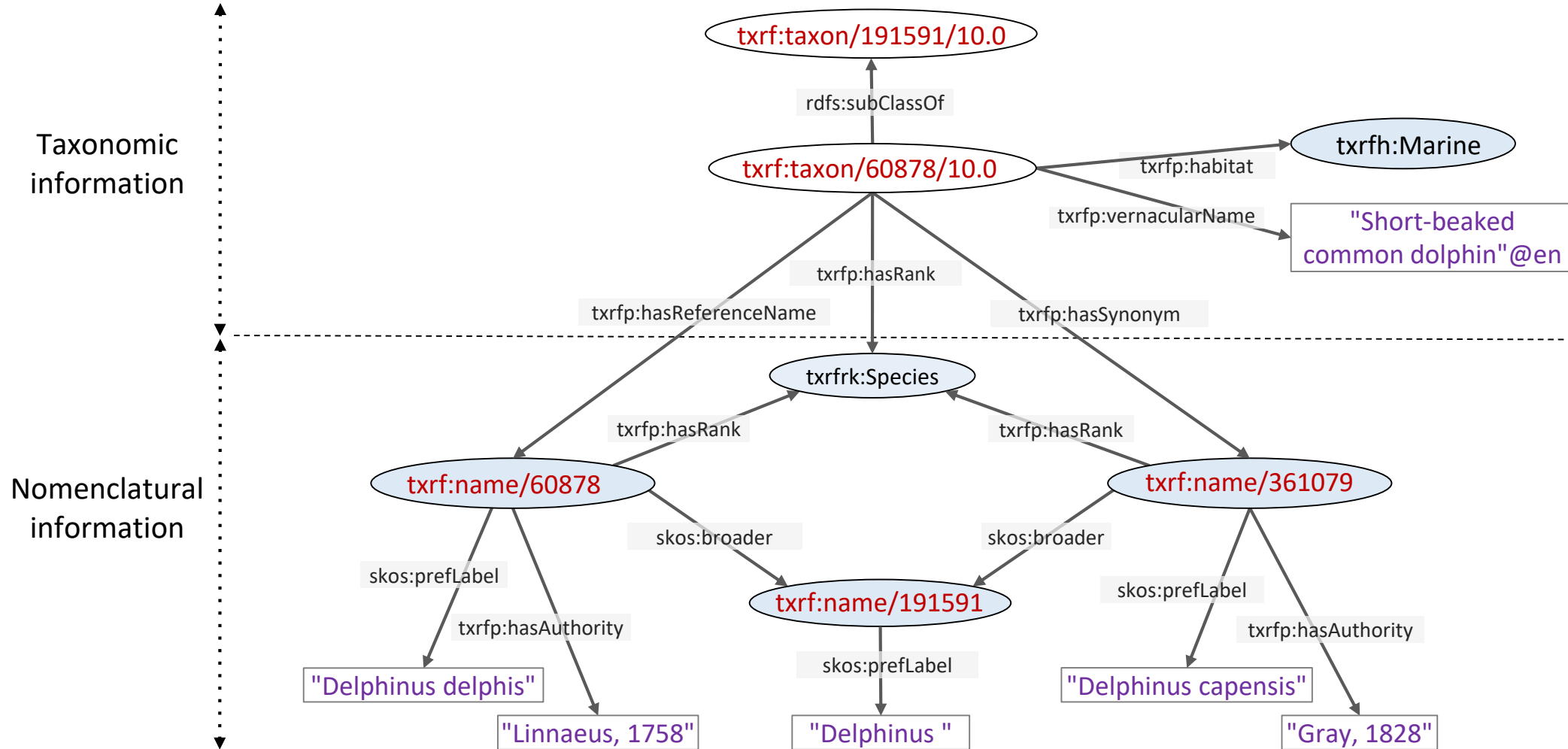
Retrieve all wheat traits related to resistance to fungal pathogens

```
SELECT *
FROM NAMED <http://ns.inria.fr/d2kab/ontology/wto/v3>
WHERE {
  GRAPH <http://ns.inria.fr/d2kab/ontology/wto/v3> {
    { ?body a ?class ; skos:prefLabel ?WTOtrait.
      ?class rdfs:subClassOf* <http://opendata.inrae.fr/wto/0000340>.
    }
    UNION
    { ?body rdfs:label ?WTOtrait ;
      rdfs:subClassOf* <http://opendata.inrae.fr/wto/0000340>.
    }
    UNION
    { ?body skos:prefLabel ?WTOtrait ; skos:broader* ?concept .
      ?concept a ?class.
      ?class rdfs:subClassOf* <http://opendata.inrae.fr/wto/0000340>.
    }
  }
}
```

WTO v3.0

- SPARQL endpoint: <http://d2kab.i3s.unice.fr/sparql>
- Github:
<https://github.com/Wimmics/WheatGenomicsSLKG/tree/main/dataset/WTO-v3.0>

TaxRef-LD



SKOS concept
OWL class

txrf: <<http://taxref.mnhn.fr/lod/>>
txrfrk: <<http://taxref.mnhn.fr/lod/taxrank#>>
txrfh: <<http://taxref.mnhn.fr/lod/habitat#>>
txrfp: <<http://taxref.mnhn.fr/lod/property#>>

TaxRef-LD

- Entry point:
<http://taxref.mnhn.fr/lod/taxref-ld/10.0>
- Delphinus delphis example:
<http://taxref.mnhn.fr/lod/taxon/60878/10.0>
<http://taxref.mnhn.fr/lod/name/60878>
- SPARQL endpoint:
<http://taxref.mnhn.fr/sparql>
- Github: <https://github.com/frmichel/taxref-ld/>
- Publication:
S4BioDiv@ISWC2017: [hal-01617708](https://hal.archives-ouvertes.fr/hal-01617708)

OPENLINK SOFTWARE

Facets (new session) Description Metadata Settings

About: [Delphinus delphis](#) [Goto](#) [Sponge](#) [NotDistinct](#) [Permalink](#)

An Entity of Type : [owl:Class](#), within Data Space : [taxref.i3s.unice.fr:8890](#) associated with source [document\(s\)](#)

New Facet based on Instances of this Class

Attributes	Values
rdf:type	Class
equivalentClass	Delphinus delphis Delphinus delphis
label	Delphinus delphis
subClassOf	Delphinus
foaf:page	https://inpn.mnhn.fr/espece/cd_nom/60878?lg=en http://doris.ffessm.fr/fiche2.aspx?nom=60878 http://www.marinespecies.org/aphia.php?p=taxdetails&id=60878 http://www.departments.buckinham.ac.uk/marine/marine.asp?s=y&id=160878 http://www.iucnredlist.org/details/60878/0
has biogeographical status	Biogeographical status in Reunion Biogeographical status in Martinique Biogeographical status in French Guiana Biogeographical status in New Caledonia »more»
has habitat	Milieu marin
has taxonomic rank	Species
has reference name	Delphinus delphis
has synonym	Delphinus fulvifasciatus Delphinus fulvifasciatus Delphinus faina Delphinus marginatus Delphinus mitchellii »more»
has vernacular name	Common Dolphin Dauphin commun Dauphin commun à bec court Dauphin commun à bec long

OPENLINK SOFTWARE

Facets (new session) Description Metadata

About: [Delphinus delphis Linnaeus, 1758](#) [Goto](#) [Sponge](#) [NotDistinct](#) [Permalink](#)

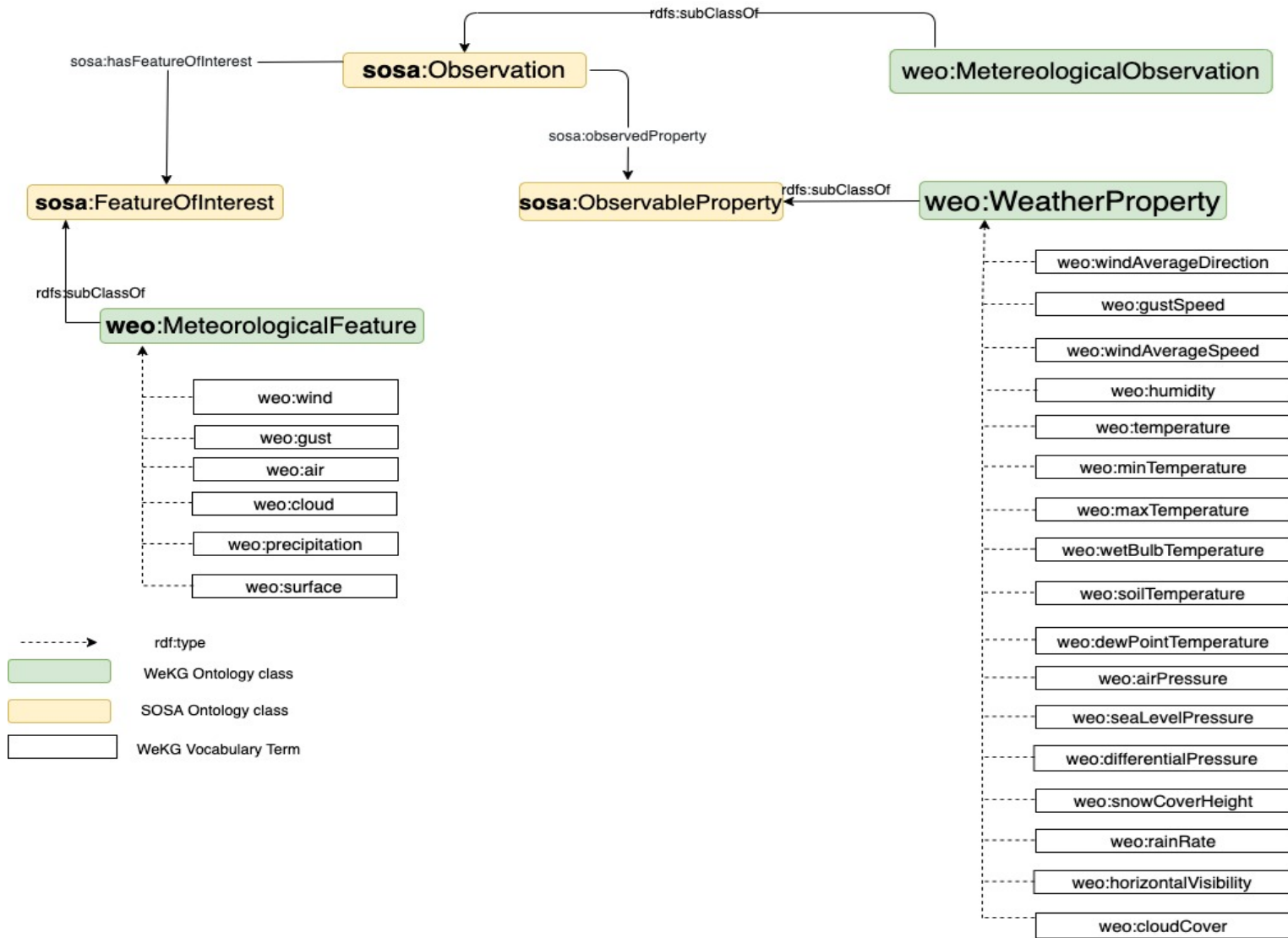
An Entity of Type : [http://rs.tdwg.org/ontology/voc/TaxonName#TaxonName](#), within Data Space : [taxref.i3s.unice.fr:8890](#) associated with source [document\(s\)](#)

Type: [Taxon Name](#) [New Facet based on Instances of this Class](#)

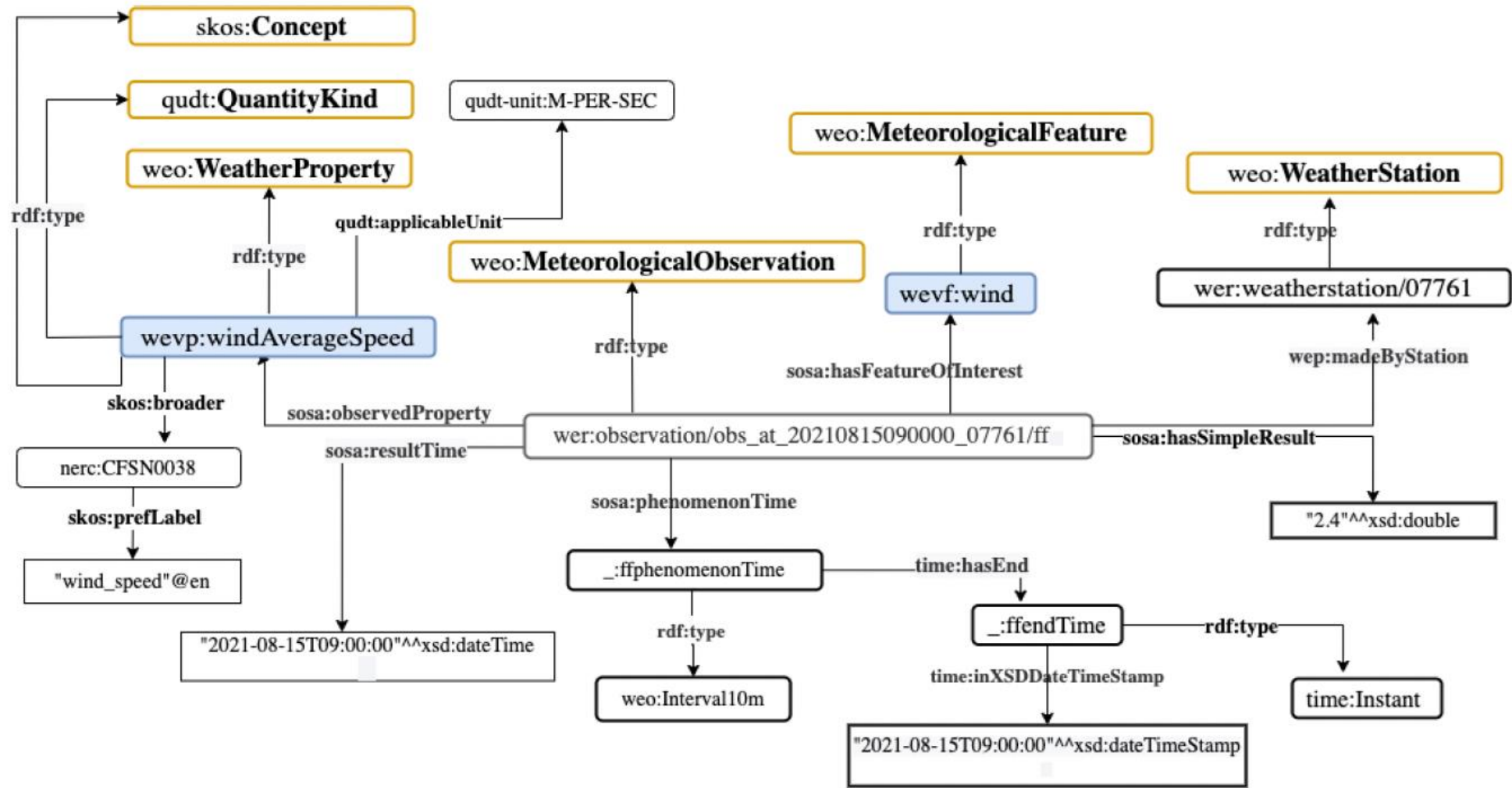
Attributes	Values
rdf:type	skos:Concept Taxon Name
label	Delphinus delphis Linnaeus, 1758
skos:broader	Delphinus
foaf:homepage	https://inpn.mnhn.fr/espece/cd_nom/60878?lg=en
skos:prefLabel	Delphinus delphis
skos:exactMatch	Delphinus delphis Linnaeus, 1758 Delphinus delphis Linnaeus 1758 urn:lsid:marinespecies.org:taxname:137094
dct:identifier	60878
foaf:page	http://doris.ffessm.fr/ref/espece/1351 ↗ http://zoobank.org/7f44dfdd-ed8b-4c24-9b07-2d705a7623be ↗ http://apiv3.iucnredlist.org/api/v3/taxonredirect/6336 ↗ http://www.marinespecies.org/cetacea/aphia.php?p=taxdetails&id=60878 ↗ https://inpn.mnhn.fr/espece/cd_nom/60878?lg=en ↗ »more»
schema:identifier	GBIF id SANDRE id WoRMS id GBIF id SANDRE id »more»
WoRMS ID (Wikidata)	137094
has scientific name authority	Linnaeus, 1758
has taxonomic rank	Species

WeKG-MF

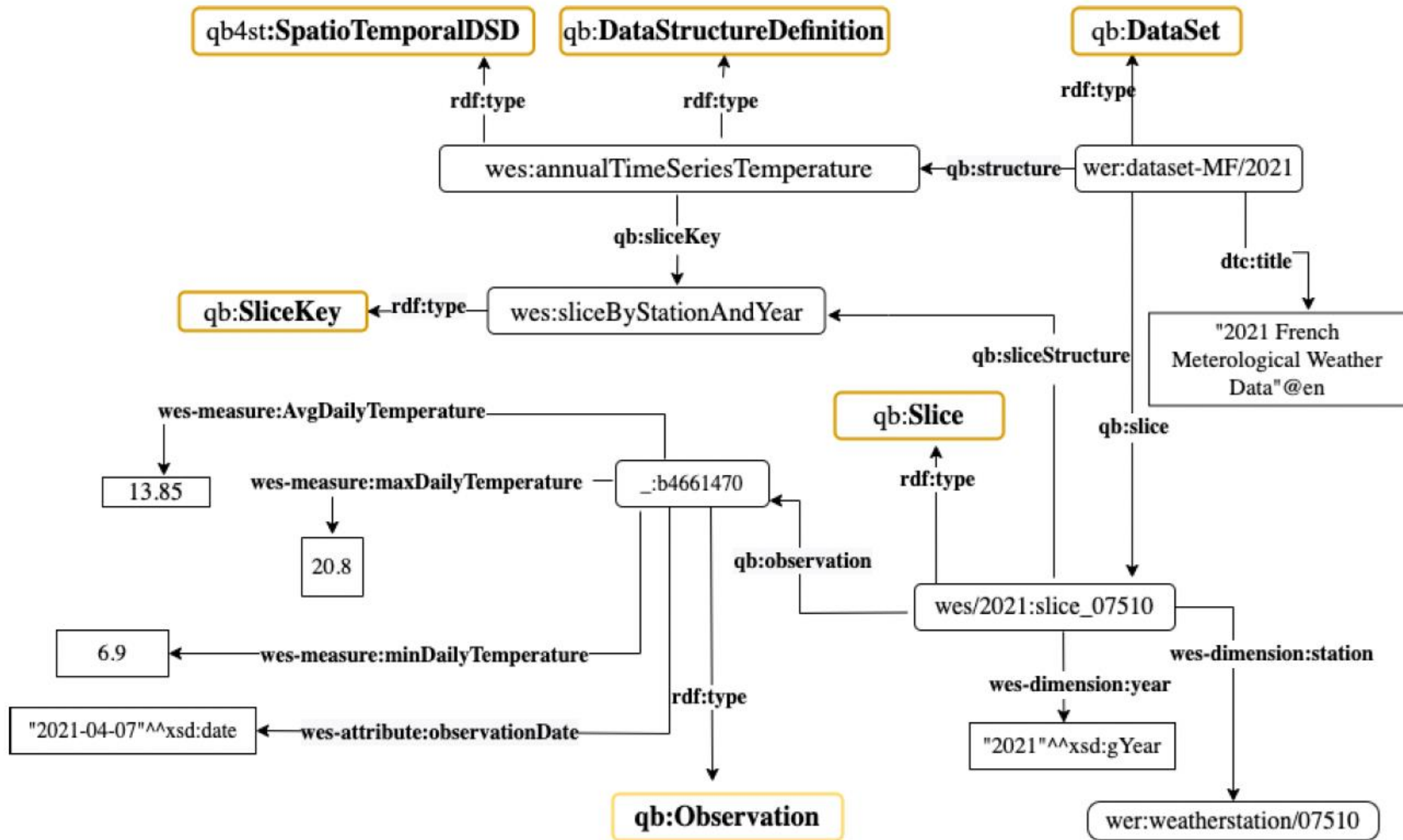
WeKG ontologie et vocabulaire



WeKG-MF



WeKG-MF



WeKG-MF: questions de compétences

- Which are the weather stations located in a given French region?
- What is the closest weather station to a specific spatial location?
- At what time of the day was the highest value of air temperature observed?
- Calculate the rain cumulation during the period of time from {date1} to {date2} in a specific location.
- Calculate the weekly average min/max temperatures in Celsius for a region {inseeRegion} during the period from {date1} to {date2}?

WeKG-MF: requêtes SPARQL

At what time of the day was the highest value of the air temperature?

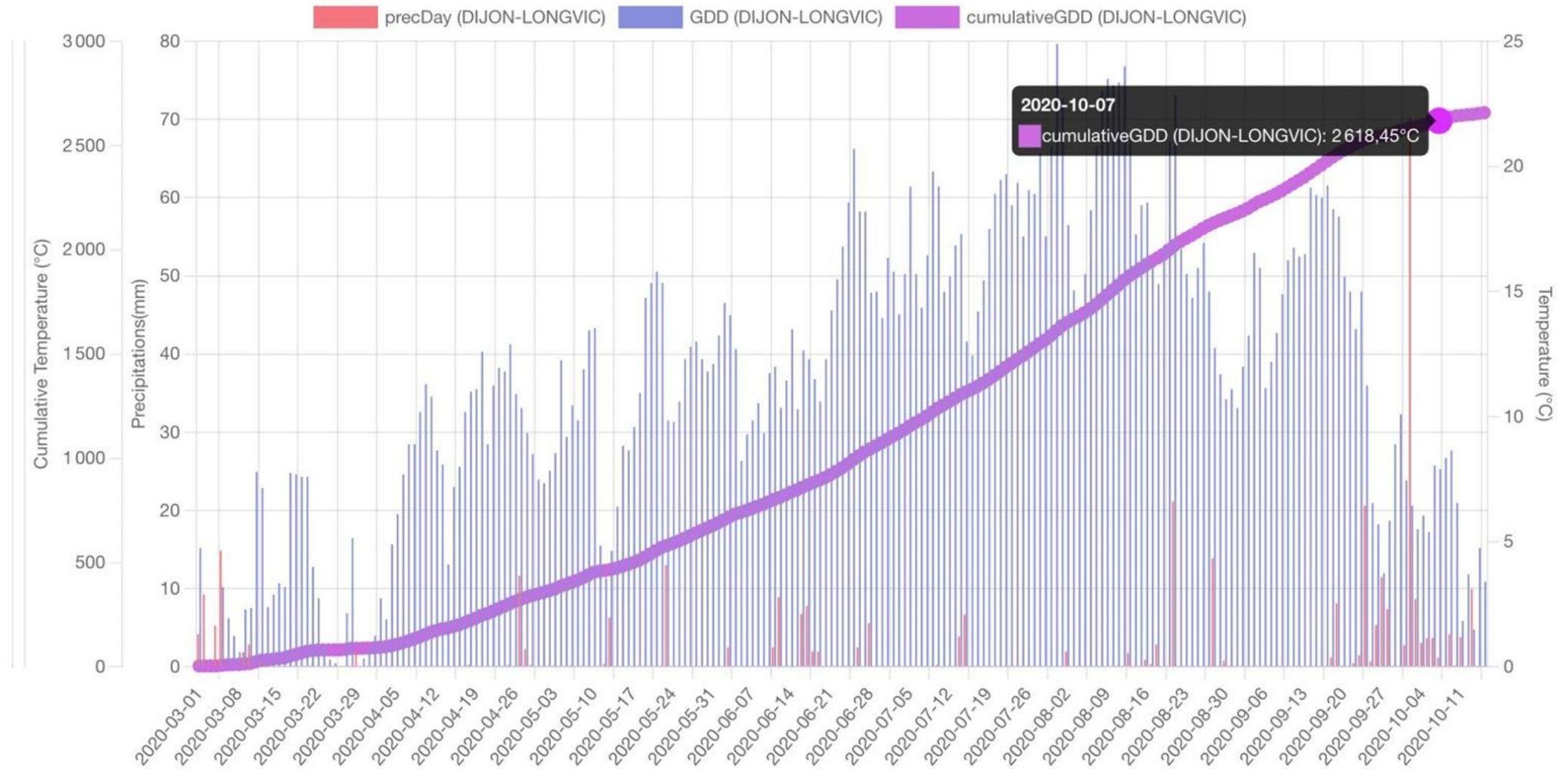
```
SELECT ?date ?hour ?station ((?temp_max- 273.15) as ?tempMaxInCelsius) WHERE {
  { SELECT ?date ?s (MAX(?v) as ?temp_max) WHERE {
    ?obs a weo:MeteorologicalObservation; sosa:observedProperty wevp:airTemperature ;
      sosa:hasSimpleResult ?v; wep:madeByStation ?s ;sosa:resultTime ?t .
    BIND(xsd:date("2022-05-01") as ?date)
    FILTER( xsd:date(?t) = ?date)
  } GROUP BY ?s ?date
}
?obs a weo:MeteorologicalObservation; sosa:observedProperty wevp:airTemperature ;
sosa:hasSimpleResult ?temp_max; wep:madeByStation ?s ; sosa:resultTime ?t .
?s rdfs:label ?station .
FILTER(xsd:date(?t)= ?date)
BIND(HOURS(?t) as ?hour)
} ORDER BY ?station
```

Requêtes SPARQL sur WeKG-MF

Calculate daily average air temperatures recorded after April, 1st 2019 in PACA

```
SELECT ?date (AVG(?temp_avg) as ?avg_temp)
WHERE {
  ?s a qb:Slice ;
  wes-dimension:station ?station ;
  qb:observation [ a qb:Observation ;
    wes-attribute:observationDate ?date ;
    wes-measure:avgDailyTemperature ?temp_avg ] .
  FILTER(xsd:date(?date) >= xsd:date("2019-04-01"))
  ?station a weo:WeatherStation ;
  dct:spatial [ wdt:P131 <http://www.wikidata.org/entity/Q15104> ] .
} GROUP BY ?date
```

Visualisation interactive des paramètres (agro)météorologiques dans WeKG-MF



WeKG-MF

- SPARQL endpoint: <http://weakg.i3s.unice.fr/sparql>
- Github: <https://github.com/Wimmics/weather-kg>
- WeKG ontology & vocabulary: <https://weakg.i3s.unice.fr/ontology/>
- Interface de visualisation: <https://nadiaya2019.github.io/DemoKGViz-1/>
- Publications:
 - ICWE 2022 [hal-03619869](https://hal.archives-ouvertes.fr/hal-03619869)
 - ESWC 2022 [hal-03657694](https://hal.archives-ouvertes.fr/hal-03657694)
 - VOILA@ISWC2022 [hal-03890035](https://hal.archives-ouvertes.fr/hal-03890035)
 - WWW 2023 [hal-04230234](https://hal.archives-ouvertes.fr/hal-04230234)

Interroger l'écosystème de KG D2KAB

Interroger l'écosystème de KG D2KAB

Interroger différents KG dans différents SPARQL endpoints

```
select * where {  
  service <http://ontology.inrae.fr/frenchcropusage/sparql> {  
    ?x rdfs:label ?l ; rdf:type ?t  
  }  
  ?y rdf:type ?t ; ?y rdfs:label ?l2  
}
```

Interroger l'écosystème de KG D2KAB

D2KAB SPARQL federated query service

<<http://corese.inria.fr/d2kab/federate>> a st:Federation ;
st:definition (liste des endpoint de D2KAB).

<http://corese.inria.fr/d2kab/federate?query=select * where {...} limit 10>

Interroger l'écosystème de KG D2KAB

Interroger différents KG (nommés) dans le même SPARQL endpoint

```
select * where {  
  graph <http://ns.inria.fr/d2kab/graph/ricegenomicsslkg> {  
    ?a oa:hasBody ?trait .  
    ...  
  }  
  graph <http://ns.inria.fr/d2kab/ontology/wto/v3> {  
    ?trait skos:prefLabel "resistance to Leaf rust" .  
  }  
  ...  
}
```

Interroger l'écosystème de KG D2KAB: WheatGenomicsSLKG + RiceGenomicsSLKG

Quels sont les taxons communs aux littératures sc. sur blé et riz

```
SELECT distinct ?uri ?label
WHERE {
  graph <http://ns.inria.fr/d2kab/graph/ricegenomicsslkg> {
    ?s1 oa:hasBody ?uri .
  }
  graph <http://ns.inria.fr/d2kab/graph/wheatgenomicsslkg> {
    ?s2 oa:hasBody ?uri .
  }
  ?uri rdfs:label ?label.
}
ORDER BY ?label
```

Interroger l'écosystème de KG D2KAB: WheatGenomicsSLKG + RiceGenomicsSLKG

Quels sont les gènes qui sont mentionnés à proximité d'un taxon de pathogène du genre "streptomyces"?

```
SELECT distinct ?paper (GROUP_CONCAT(distinct ?geneName; SEPARATOR="-") as ?genes) ?ncbiTaxon
FROM <http://ns.inria.fr/d2kab/graph/wheatgenomicsslkg>
FROM <http://ns.inria.fr/d2kab/graph/ricegenomicsslkg>
FROM <http://purl.obolibrary.org/obo/ncbitaxon/ncbitaxon.owl>
WHERE {
  ?a1 a oa:Annotation ; oa:hasTarget [ oa:hasSource ?source1 ] ;
  oa:hasBody [ a d2kab:Gene ; skos:prefLabel ?geneName ].
  ?source1 frbr:partOf+ ?paper .
  ?a2 a oa:Annotation; oa:hasTarget [ oa:hasSource ?source2 ] ;
  oa:hasBody ?ncbitaxonURI .
  ?source2 frbr:partOf+ ?paper .
  ?paper a fabio:ResearchPaper ; dct:title ?titleURI .
  GRAPH <http://purl.obolibrary.org/obo/ncbitaxon/ncbitaxon.owl> {
    ?ncbitaxonURI rdfs:subClassOf* <http://purl.obolibrary.org/obo/NCBITaxon_1883>;
    rdfs:label ?ncbiTaxon . }
}
```

Interroger l'écosystème de KG D2KAB: WheatGenomicsSLKG + BSV

Quels sont les articles de Pubmed et les BSV qui parlent du blé tendre?

```
SELECT distinct ?paper ?bsv ?taxLabel ?fcuCropName ?taxrefClass
FROM <http://ns.inria.fr/d2kab/graph/wheatgenomicsslkg>
FROM <http://ns.inria.fr/d2kab/graph/alignments-fcu-taxref>
WHERE {
  { SELECT distinct ?paper ?taxon WHERE {
    ?annot a oa:Annotation; oa:hasTarget [ oa:hasSource ?source ] ; oa:hasBody ?taxon .
    ?taxon a d2kab:Taxon; skos:prefLabel ?label .
    ?source frbr:partOf+ ?paper . ?paper a fabio:ResearchPaper ; dct:title ?source .
    FILTER(CONTAINS(?label, "Triticum aestivum"))
  }
  LIMIT 100
}

SERVICE <http://taxref.i3s.unice.fr/sparql> {
  ?taxrefClass owl:equivalentClass ?taxon ; rdfs:label ?taxLabel .
  ?fcuCropName taxref:candidateAlignment_eppo|taxref:candidateAlignment_geves ?taxrefClass .

SERVICE <http://ontology.inrae.fr/bsv/sparql> {
  ?bsv a d2kab_inrae:Bulletin ; dul:isRealizedBy ?s ; dct:spatial ?w ; dct:date ?date_bsv .
  ?aa a oa:Annotation ; oa:hasTarget [ oa:hasSource ?s ] ; oa:hasBody ?fcuCropName .
}
LIMIT 20
```


Interroger l'écosystème de KG D2KAB: WeKG-MF + BSV

Calcul de la somme des GDD pour la vigne entre 1er avril et date du dernier BSV paru en 2019 dans la région PACA

```
SELECT ?phb ?date_bsv (SUM(?avg_temp-10) as ?sum_GDD) WHERE {  
  SERVICE <http://ontology.inrae.fr/bsv/sparql> {  
    SELECT ?phb ?date_bsv WHERE {  
      ?phb a d2kab_inrae:Bulletin ; dct:spatial <http://www.wikidata.org/entity/Q15104>;  
      dct:date ?date_bsv ; dul:isRealizedBy ?phb_html . ?phb_html dce:format "text/html" .  
      ?rs a oa:ResourceSelection ; oa:hasSource ?phb_html .  
      ?aa a oa:Annotation ; oa:hasTarget ?rs ; oa:hasBody fcu:Vignes .  
      FILTER (xsd:date(?date_bsv) >= xsd:date("2019-04-01")) }  
    ORDER BY DESC(?date_bsv) LIMIT 1 }  
  SERVICE <http://weakg.i3s.unice.fr/sparql> {cf. slide 32}  
  FILTER(xsd:date(?date) <= xsd:date(?date_bsv) )  
}
```

Interroger l'écosystème de KG D2KAB: WeKG-MF + WheatObservationsKG

Rechercher la station météo la plus proche d'une parcelle dans une investigation et calculer des paramètres agroclimatiques relativement à la culture qui fait l'objet de l'investigation (blé ou maïs)

Interroger l'écosystème de KG D2KAB et au-delà: WeKG-MF + Coffee-WeKG

- Quelles sont les propriétés observables communes?
- Comparer les mesures observées à la même date en des lieux différents (e.g. température, taux d'humidité)?

Description de l'écosystème D2KAB: IndeGx, KartographI, Metadatamatics

- SPARQL endpoint: <http://prod-dekalog.inria.fr/sparql>
- KartoGraphI: <http://prod-dekalog.inria.fr/?graphSetIndex=d2kab>
- Metadatamatics: <https://wimmics.github.io/voidmatic/>
- Github: <https://github.com/Wimmics/IndeGx>
- Publications:
 - IndeGx: JWS (76) 2023 [hal-03946680](https://hal.archives-ouvertes.fr/hal-03946680)
 - Metadatamatics: WWW2023 [hal-04171196](https://hal.archives-ouvertes.fr/hal-04171196)
 - KartoGraphI: ESWC 2022 [hal-03652865](https://hal.archives-ouvertes.fr/hal-03652865)

Questions?

Catherine Faron, Nadia Yacoubi, Franck Michel, Olivier Corby, Fabien Gandon,
Pierre Maillot, Arnaud Barbe, Nicolas Audoux, Frédéric Météreau, Aymeric Rebuffel,
partenaires D2KAB