



Extraction de connaissances à partir du corpus des Bulletins de Santé du Végétal: Mise à jour des échelles phénologiques

Catherine ROUSSEY, Marine COURTIN, Robert BOSSY, Stephan BERNARD, Xavier DELPUECH, Lionel DELBAC

Plan



Contexte : le projet D2KAB et le corpus des Bulletins de Santé du Végétal (BSV)

Enrichissement de ressource sémantique par extraction de terminologie

Annotation des BSV

Perspectives et Conclusion



le projet D2KAB et le corpus des Bulletins de Santé du Végétal



Partenaires



ACTA Matthieu Hirschy,



INRAE DIPSO Sophie Aubin, Sonia Bravo*, François-Xavier Sennesal,

ELZEARD Florence Amardeilh, Baptiste Darnala, Juliette Raphelle*

IFV Lionel Delbach, Xavier Delpuech, Philippe Larignon,



INRIA WIMMICS I3S Catherine Faron, Franck Michel,

LIRIS Nadia Yacoubi^(I3S),

INRAE LISC Stephan Bernard^(TSCF, CATI CODEX),



INRAE MAIAGE Robert Bossy, Marine Courtin,

INRAE MISTEA Clement Jonquet^(LIRMM), Catherine Roussey^(TSCF, CATI CODEX)



INRAE MISTEA/OpenSILEX Arnaud Charleroy^(CATI CODEX), Anne Tireau^(CATI CODEX),



INRAE PHIM Nicolas Sauvion,

UMR ABSYS Yvan Buisson, Nathalie Smits,

Anna Chepaikina*^(TSCF)



Plant Health
Institute
Montpellier

INRAE

Séminaire Sésame

Novembre 2024 / C Roussey

Bulletins de Santé du Végétal (BSV)



LICENCE OUVERTE
OPEN LICENCE



Les BSV sont distribués sur le web

Les BSV constituent une archive d'observations des parcelles agricoles françaises et décrivent :

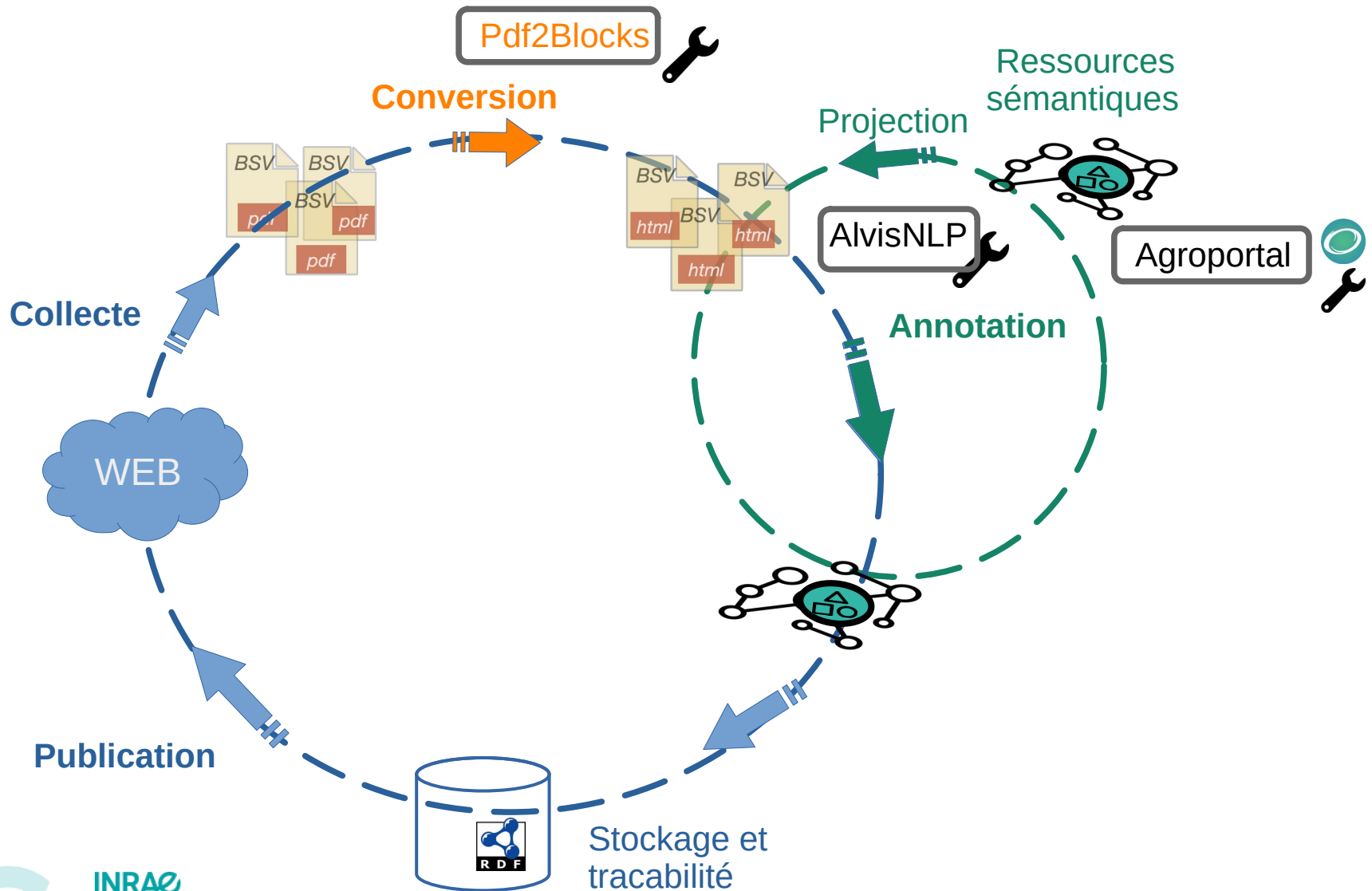
- Etat sanitaire des cultures,
- Observations des stades de développement,
- Observations des ravageurs et maladies avec la présence de symptômes,
- Evaluations du risque phytosanitaire, en fonction des périodes de sensibilité des cultures et des seuils de nuisibilité des ravageurs et maladies,
- Messages réglementaires.

→ Extraire ces informations

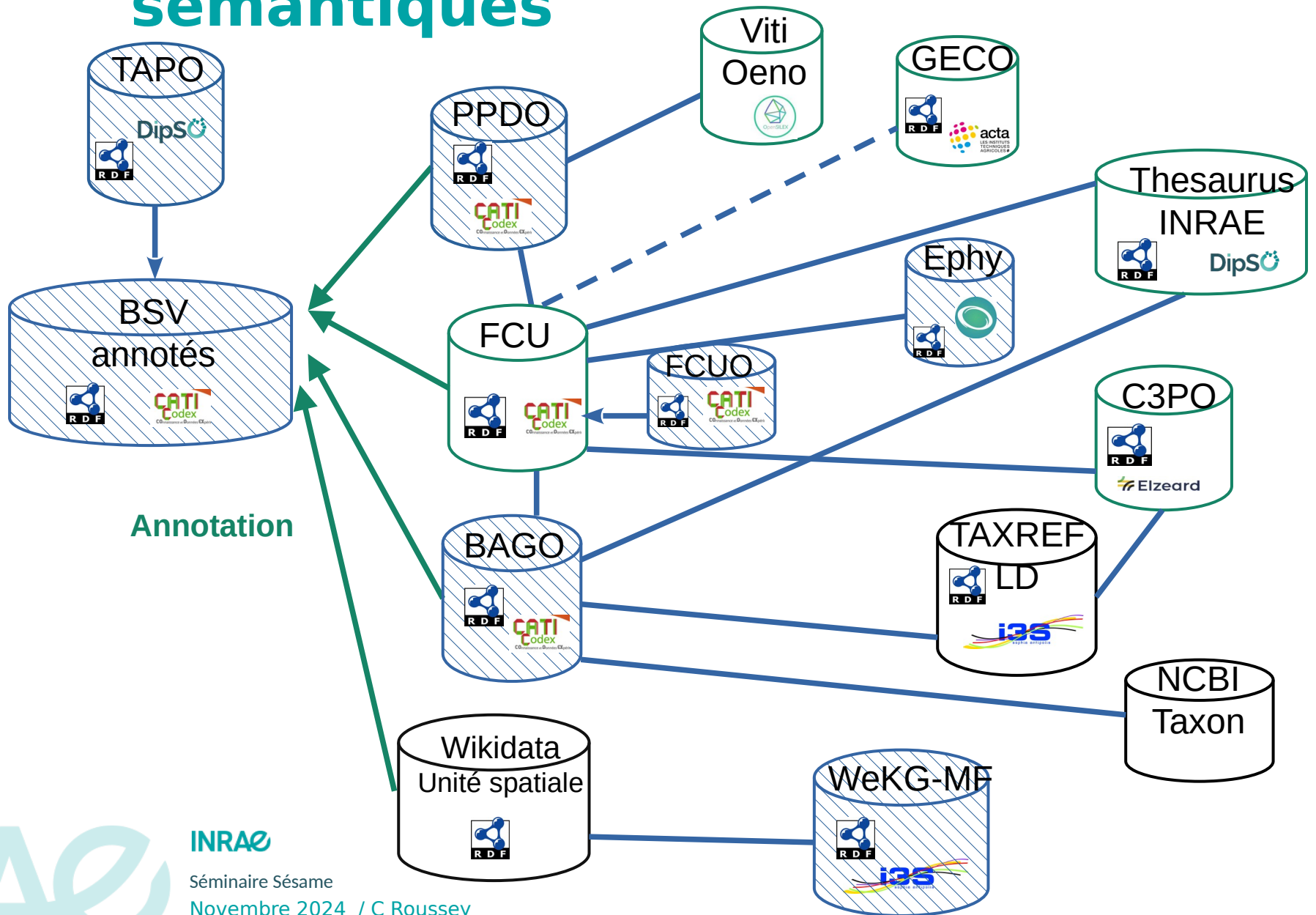
INRAE

Séminaire Sésame
Novembre 2024 / C Roussey

Cycle de vie des BSV ...



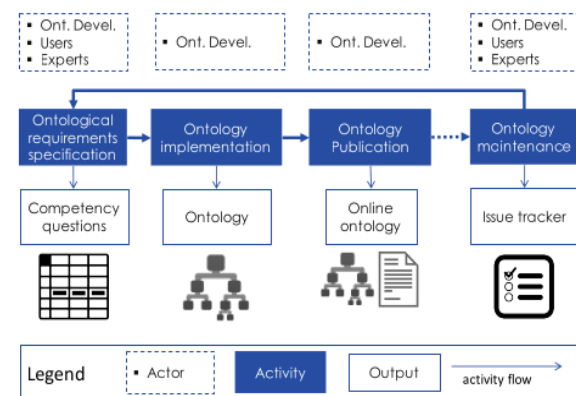
Un réseau de ressources sémantiques



Méthode de lifting de données de référence

Méthode « Linked Open Terms » (LOT)

- Questions de compétences (QC)
- Recherche de vocabulaires ou d'ontologies existantes
- **Extraction de terminologies à l'aide de méthode TALN**
- **Recherche de définitions et de sources de référence**
- **Choix du niveau de formalisation OWL ↔ RDFS ↔ SKOS**
- Diagramme UML / CHOWLK
- Production de fichiers csv avec des experts +
- **Production de définitions qui expliquent le voisinage de l'entité dans l'ontologie**
- **Peuplement avec plugin Cellfie de Protege**
- Validation avec un raisonneur plugin SWRL Tab + drools
- **Validation avec SKOSPLAY**
- Publication : depot GIT + AgroPortal + Sparql EndPoint + DOI sur RDG



Credit Espinoza-Arias, P., Poveda-Villalón, M. & Corcho, O. Using LOT methodology to develop a noise pollution ontology: a Spanish use case. J Ambient Intell Human Comput 11, 4557–4568 (2020).
<https://doi.org/10.1007/s12652-019-01561-2>

French Crop Usage (FCU) : Thésaurus des usages des plantes cultivées en France

QC :

- 1) Quelles sont les plantes cultivées en France ?
- 2) Pour quel usage est cultivée la plante ?...

URI : <https://opendata.inrae.fr/fcu-res/>

Git <https://forgemia.inra.fr/bsv/frenchcropusage>

Agroportal : agroportal.lirmm.fr/ontologies/CROPUSAGE

SPARQL EndPoint : <https://rdf.codex.cati.inrae.fr/fcu-res/snorql/>

DOI : <https://doi.org/10.15454/QHFTMX>

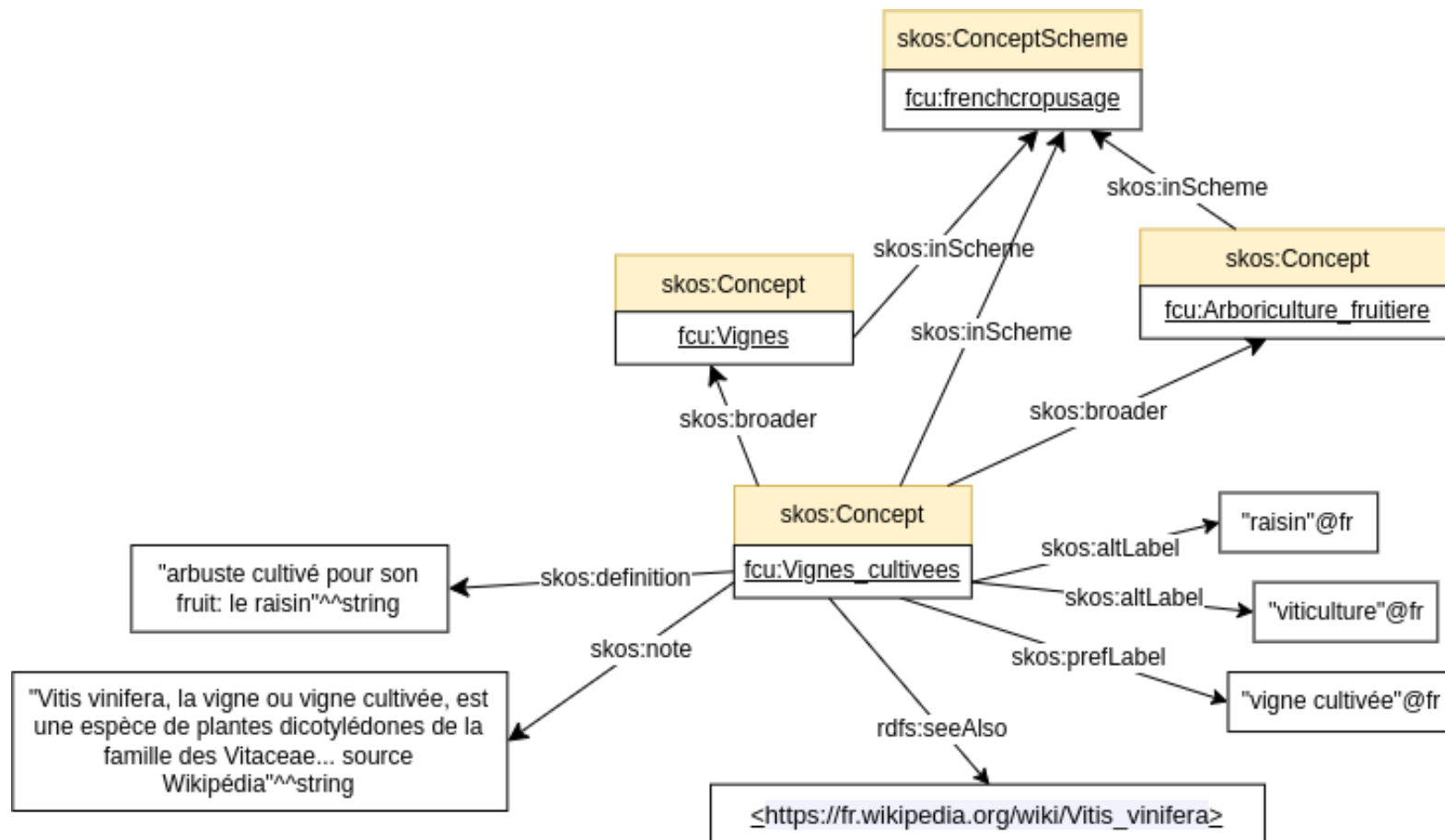
Version 3.3

Formalisation : SKOS + 1 classe OWL avec 707 concepts en français, anglais et code eppo

Plusieurs sources, plusieurs experts, ...



French Crop Usage (FCU) : Exemple



French Crop Usage Ontologie (FCUO) :

QC :

1) Quelles sont les types d'alignements possibles entre un taxon scientifique (une espèce) et le rôle d'une plante en agriculture (FCU usage) ?...

URI : <https://opendata.inrae.fr/fcu-def#>

Git <https://forgemia.inra.fr/bsv/frenchcropusage>

Agroportal : agroportal.lirmm.fr/ontologies/FCUO

SPARQL EndPoint : <https://rdf.codex.cati.inrae.fr/fcu-res/snorql/>

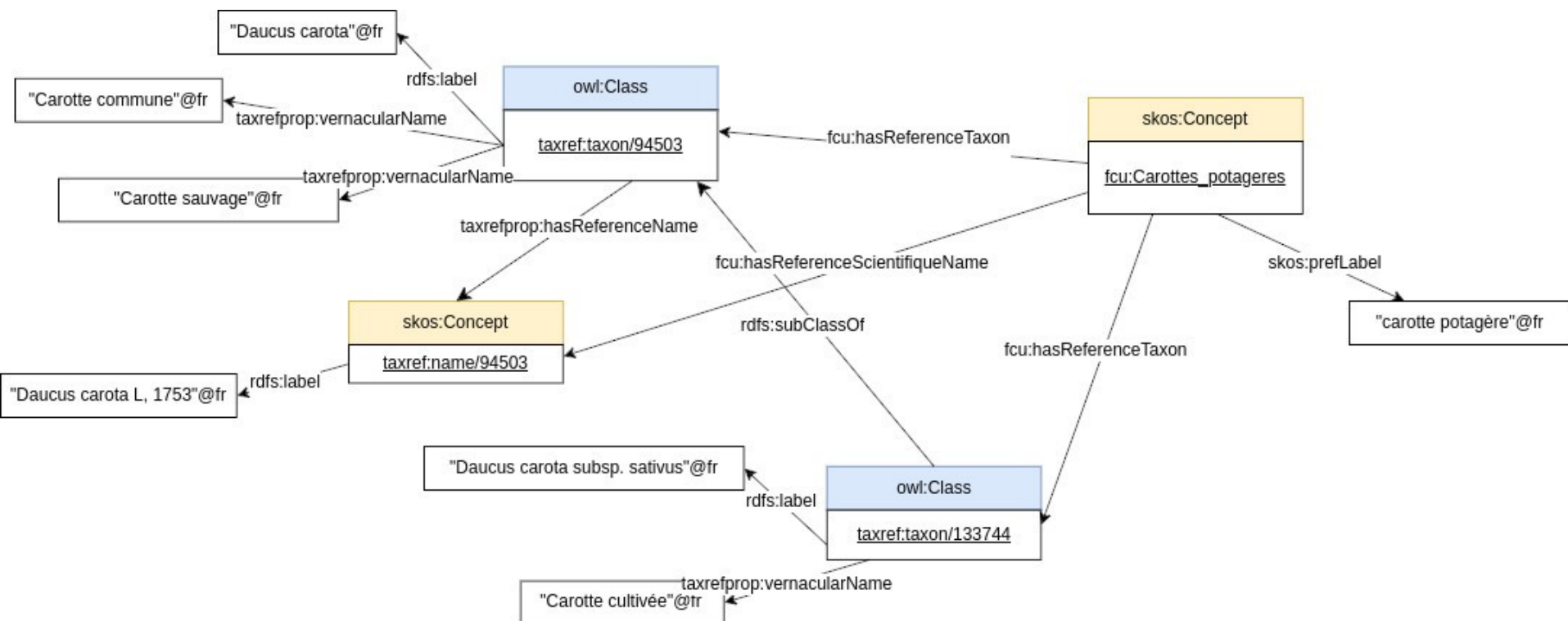
DOI : <https://doi.org/10.57745/SAANOA>

Version 1.0

Formalisation : OWL avec 12 propriété objet et 10 propriété d'annotation

F. AMARDEILH, S. AUBIN, S. BERNARD, S. BRAVO, R. BOSSY , C. FARON, F. MICHEL, J. RAPHEL, C. ROUSSEY. Combining different points of view on plant descriptions: mapping agricultural plant roles and biological taxa. *Frontiers Artificial Intelligence, Section AI in Food, Agriculture and Water, Research Topic Knowledge Graph Technologies: the Next Frontier of the Food, Agriculture, and Water Domains*. Volume 6, 2023, doi: 10.3389/frai.2023.1188036, <https://hal.science/hal-04289970>,

French Crop Usage Ontologie (FCUO) : Exemple



BBCH-based Plant Phenological Description Ontology

QC :

- 1) Quelles sont les stades d'une échelle phénologique ?
- 2) Quelle sont les plantes cultivées correspondant à l'échelle phénologique?...

URI : <https://opendata.inrae.fr/ppd-def>

Git <https://forgemia.inra.fr/bsv/phenologicalstages>

Agroportal : agroportal.lirmm.fr/ontologies/PPDO

SPARQL EndPoint : <https://rdf.codex.cati.inrae.fr/ppd-def/sparql>

DOI : <https://doi.org/10.15454/TIMQHW>

Version 1.3

Formalisation : SKOS + OWL

C. ROUSSEY, X. DELPUECH, M. RAYNAL, F. AMARDEILH, S. BERNARD, C. JONQUET, C. NOÛS.
Description sémantique des stades de développement phénologique des plantes, cas d'étude de la vigne.
Dans les Actes des 32e Journées Francophones d'Ingénierie des Connaissances IC 2021, adossée à la
14e Plate-forme Francophone d'Intelligence Artificielle, 28 juin - 2 juillet 2021, Bordeaux, France.
<https://hal.inrae.fr/emse-03260085>

Les KG des échelles phénologiques

14 échelles phénologiques :

- 2 échelles générales (ppd-gen),
- 5 échelles phénologiques de la vigne (ppd-gv <https://doi.org/10.57745/YVM7VH>),
- 3 échelles individuelles BBCH des légumes (ppd-lvfh, ppd-rsv, ppd-sol) ,
- 1 échelle pour les céréales (ppd-cr)
- 3 échelles pour les fruits (ppd-stf, ppd-pof) dont 1 échelle pour l'olivier (ppd-olt <https://doi.org/10.57745/E33CXX>).

→ des experts différents par échelle

Max 90 concepts / échelle : français, anglais, espagnole, allemand

URI : [https://opendata.inrae.fr/ppd-res/\[nomCulture\]](https://opendata.inrae.fr/ppd-res/[nomCulture])

<https://opendata.inrae.fr/ppd-res/grapevine>

Git <https://forgemia.inra.fr/bsv/phenologicalstages>

Agroportal des vues différenciées de PPDO

Sparql EndPoint : [https://rdf.codex.cati.inrae.fr/ppd-res/\[nomCulture\]/sparql](https://rdf.codex.cati.inrae.fr/ppd-res/[nomCulture]/sparql)



Bio AGgressor Ontology

QC :

- 1) Quel est le type du bioagresseur ?
- 2) Quels sont les agents pathogènes qui provoquent une maladie ?...

URI : <https://opendata.inrae.fr/bago-def#>

Git <https://forgemia.inra.fr/bsv/bio-agressor-ontology>

Agroportal : agroportal.lirmm.fr/ontologies/BAGO

SPARQL EndPoint : <https://rdf.codex.cati.inrae.fr/bag-def/sparql>

DOI : <https://doi.org/10.57745/QOQFPN>

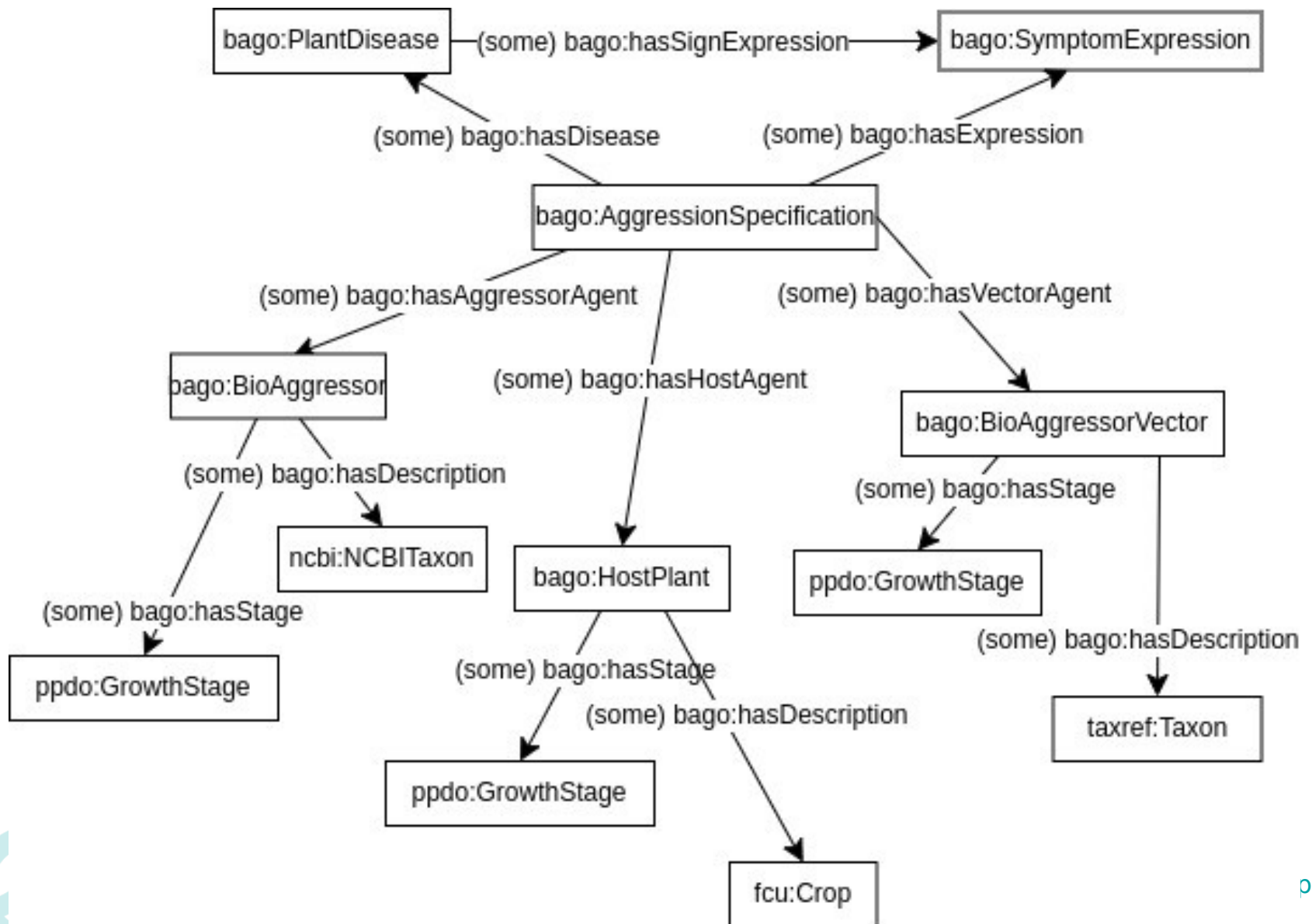
Version 1.0

Formalisation : OWL avec 69 classes, 38 propriétés objets, ...

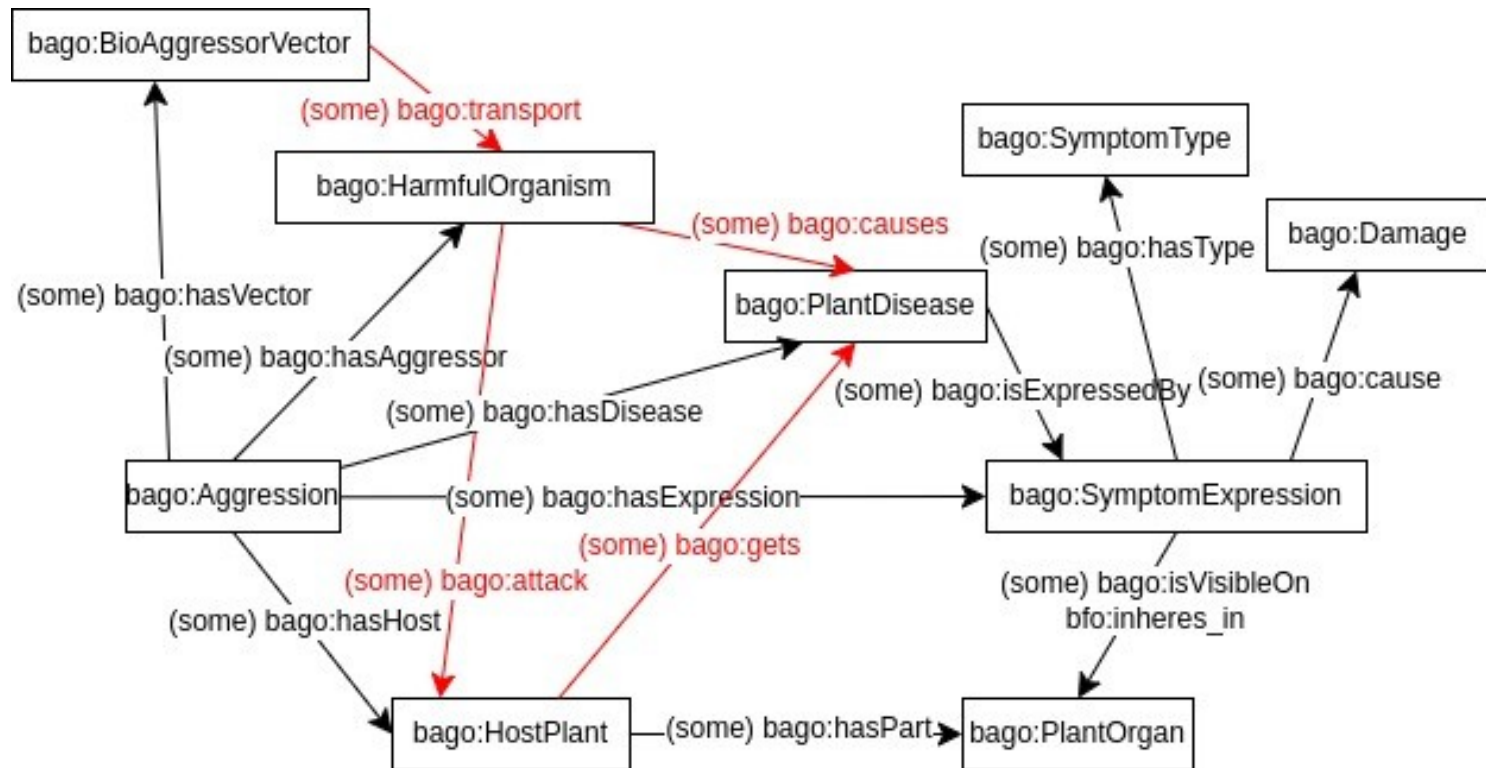
Sera ensuite dérivée par culture comme PPDO

F. AMARDEILH, S. BERNARD, R. BOSSY, M. COURTIN, M. HRSCHY, P. LARIGNON, C. ROUSSEY, N. SAUVION. Une ontologie pour modéliser les bioagresseurs des plantes. Dans les Actes des 35e Journées Francophones d'Ingénierie des Connaissances IC 2024, adossée à la 17e Plate-forme Francophone d'Intelligence Artificielle, 3-5 juillet 2024, La Rochelle, France

Bio AGgressor Ontology : Exemple



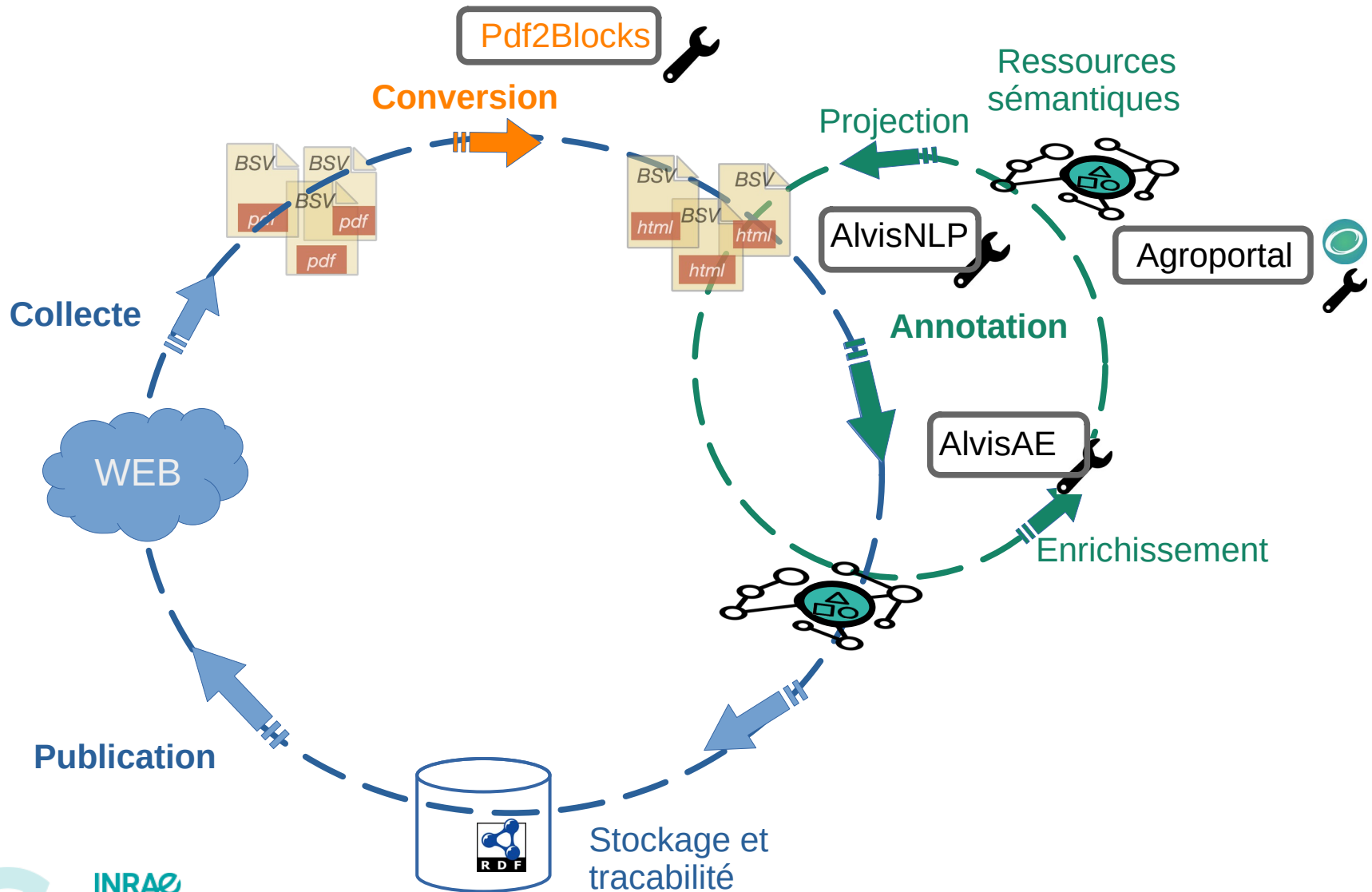
Bio AGgressor Ontology : Exemple





Enrichissement de ressource sémantique par extraction de terminologie : Cas du KG ppd-gv

Cycle de vie des BSV ...



Phénologie

La phénologie est l'étude du développement saisonnier des végétaux déterminé par l'influence des variations du climat. Pour se faire on observe la date d'apparition des stades phénologiques.

Un stade phénologique (stade de développement des plantes) caractérise une phase de développement de la plante pendant son cycle de vie.

Un stade est déterminé en fonction de plusieurs observables :

- Quel organe de la plante est observé (feuille, fleur, fruit) ?
- Quelle est la taille observée de l'organe ? Combien d'organes sont observés ?
Quelle est la couleur observée de l'organe? ...

Une échelle phénologique est un ensemble de stades fixés par convention.



Four seasons of grapevine source

Les échelles phénologiques de la vigne

Plusieurs échelles phénologiques existent pour la vigne:

- 1952 Échelle Baggiolini = 10 stades
- 1977 Échelle Eichhorn and Lorenz (EL) = 24 stades
- 1992 Échelles BBCH \cong 50 stades
- 1993 Échelle Baggiolini = 16 stades

Échelles Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie (BBCH) :

- 30 échelles individuelles BBCH spécifiques à une culture: échelle individuelle BBCH de la vigne
- Échelle générale BBCH: pour les cultures qui n'ont pas d'échelle individuelle

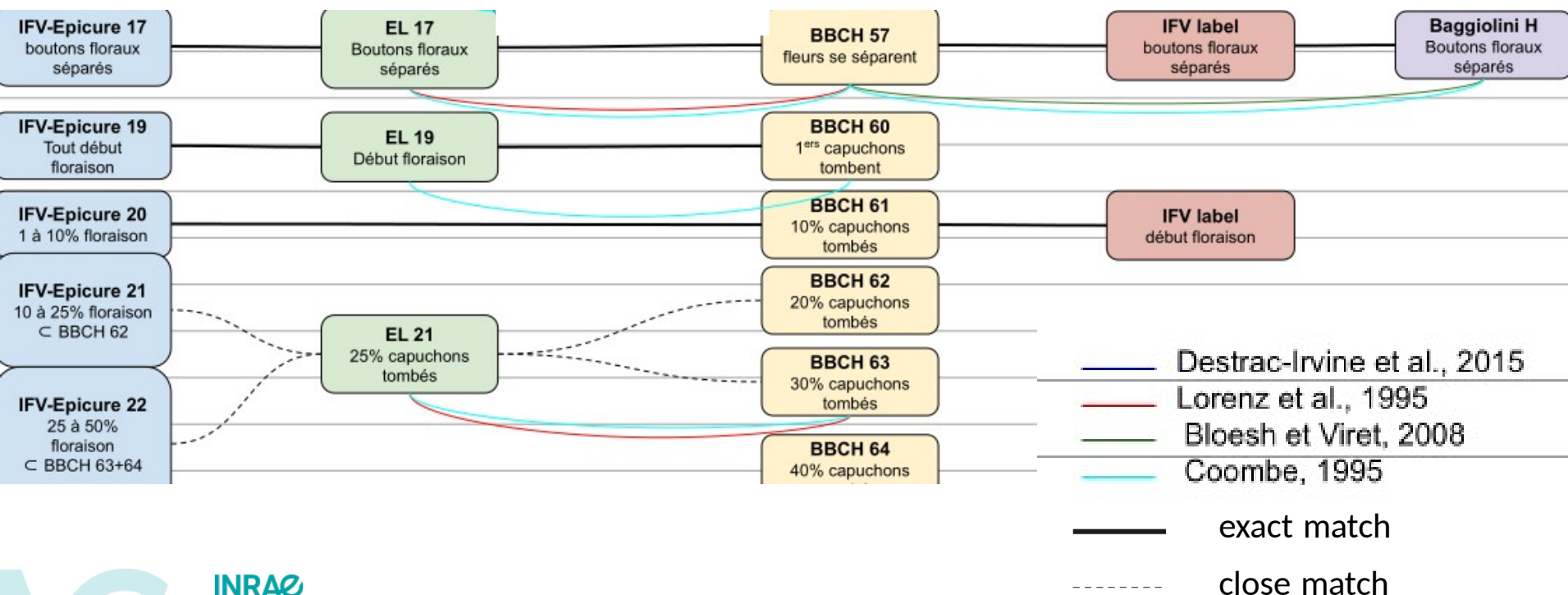


Spécification des alignements entre les échelles de la vigne

5 échelles existantes, qui sont alignées partiellement :

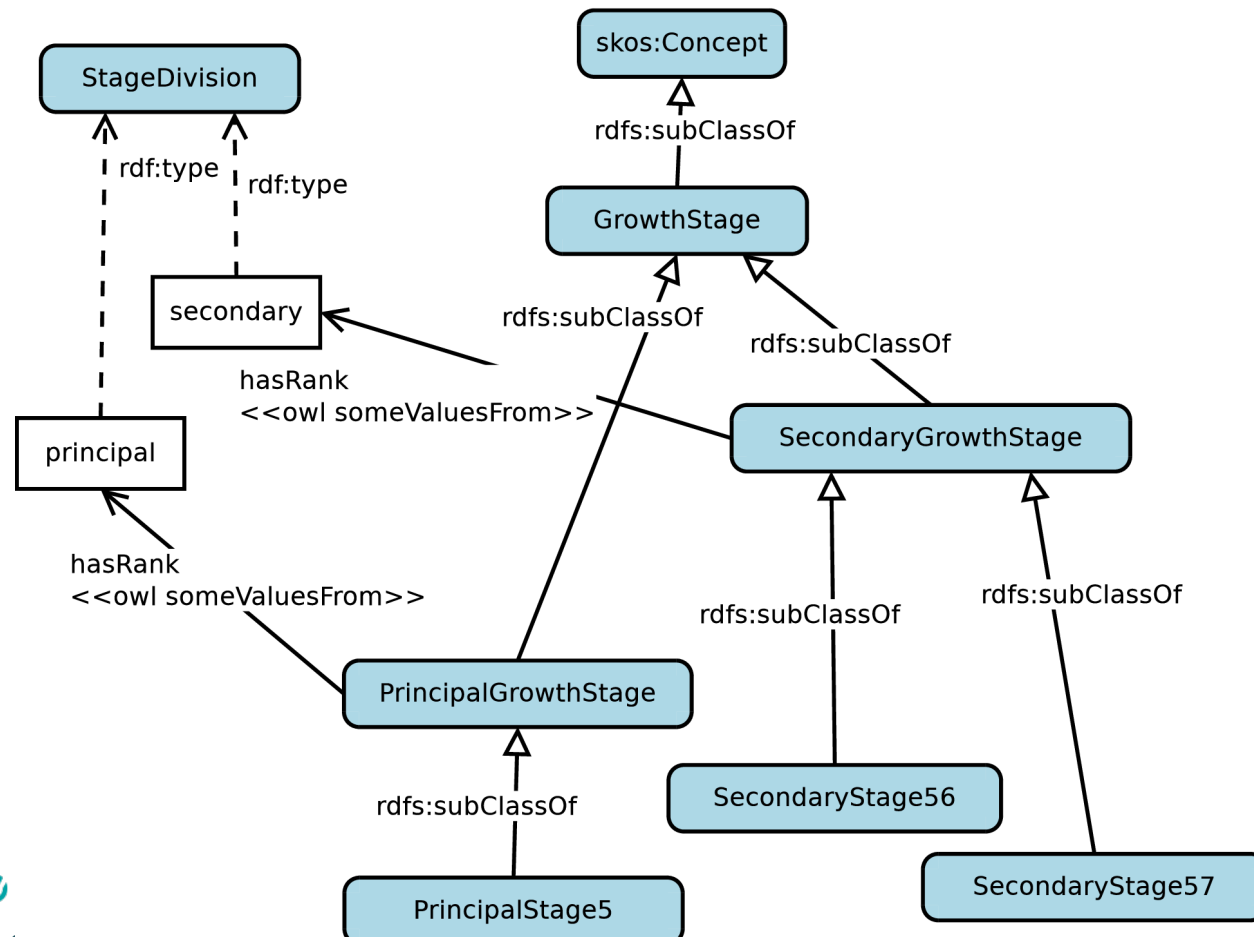
le type d'alignement n'est pas spécifié : exacte, proche, générique

certaines échelles sont alignées deux à deux sans vérification de la cohérence de l'ensemble des alignements

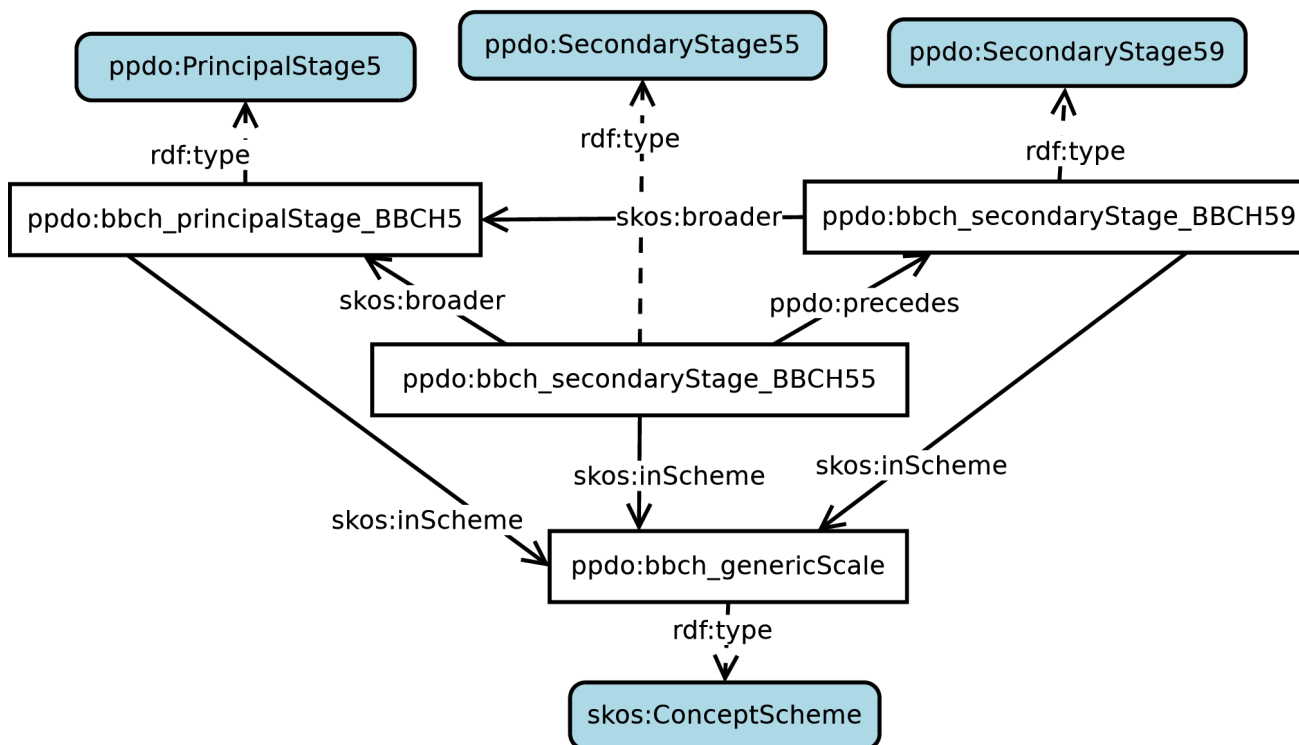


BBCH-based Plant Phenological Description Ontology

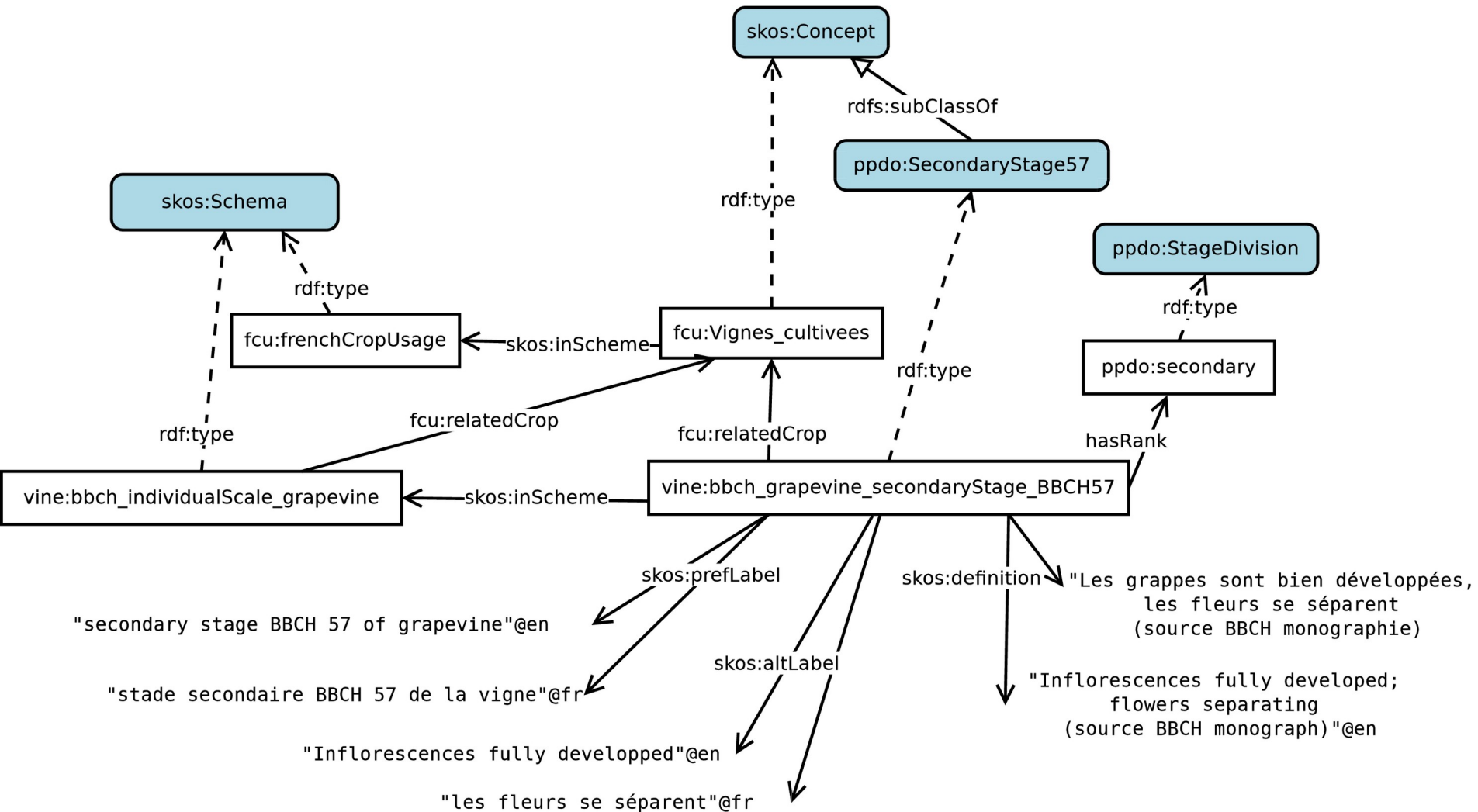
Création d'une classe pour chaque stade existant dans une des échelles BBCH



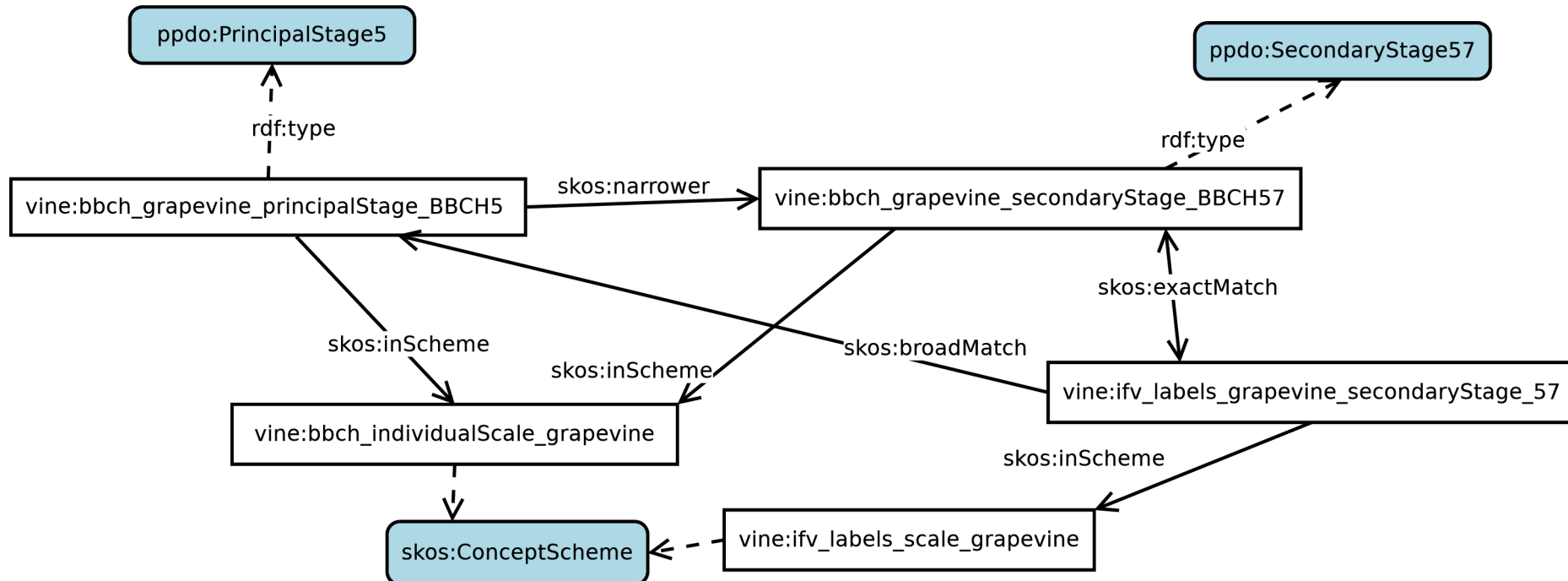
Stades de l'échelle générale BBCH



Stade de l'échelle individuelle BBCH de la vigne



Alignements entre stades



Annotation manuelle des BSV

- Un corpus de BSV Viti Alsace sélectionné
- Corriger les annotations automatiques de AlvisNLP
- Prise en main AlvisAE formation de 8 experts viti
- un guide d'annotation des BSV [Doi :10.57745/I5YVJH]
- Un dépôt git dédié
<https://forgemia.inra.fr/bsv/bsv-campagne-annotation-manuelle>
 - Plan AlvisNLP, la documentation, ...
 - Issues pour résoudre les problèmes des annotateurs
- 5 experts viti ont annotés 5 BSV chacun
- Correction semi automatique des annotations
- 8 jours de travail (experts viti + experts TALN)



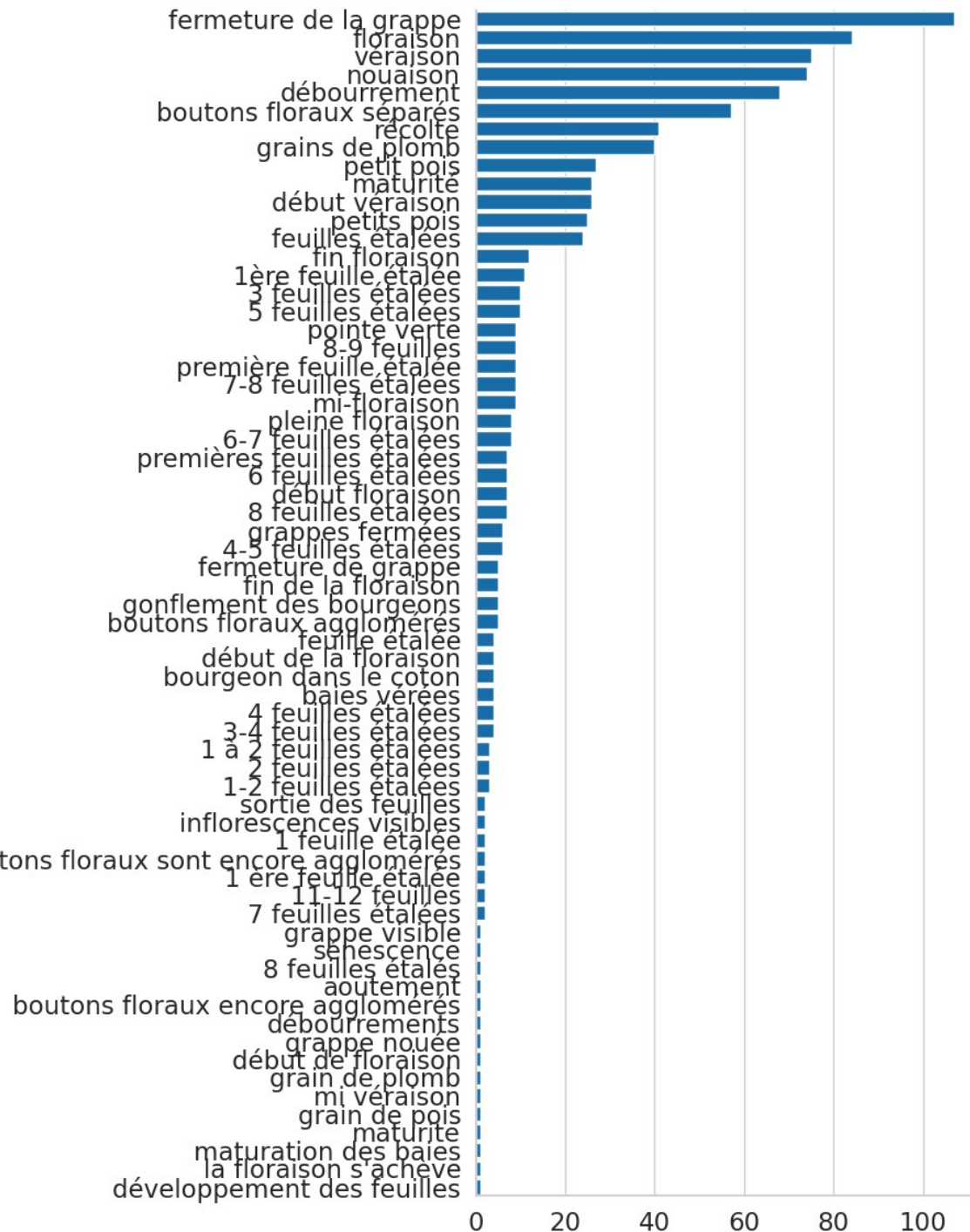
Résultats

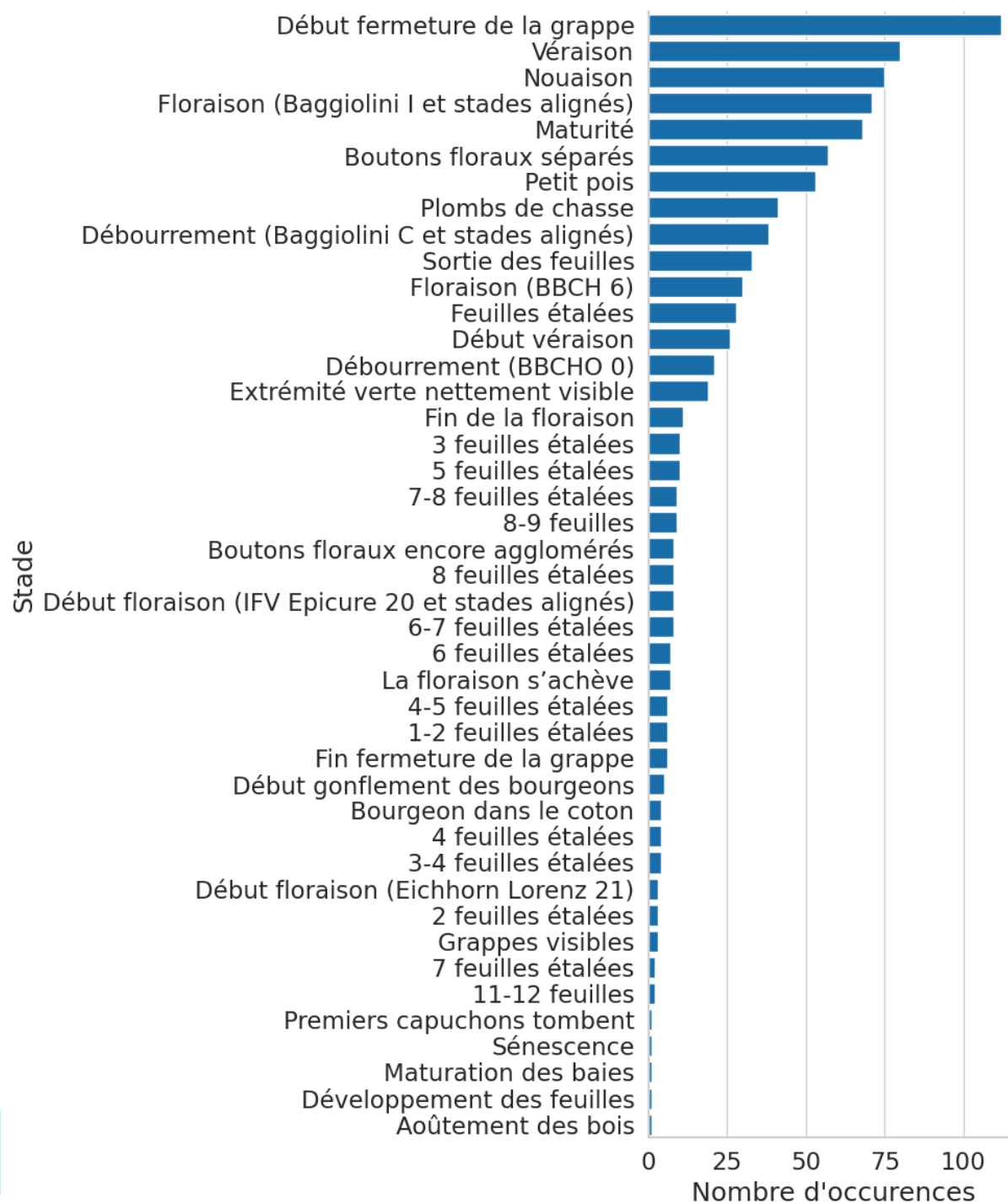
900 mentions de stades identifiées et corrigées pour 25 BSV annotés

Les labels "fermeture de la grappe" (107 occurrences), "floraison" (84 occurrences) et "véraison" (75 occurrences) = 30% des mentions

- Des raccourcies :
 - "éclatement" → "éclatement des bourgeons",
 - "coton" → "bourgeon dans le coton",
 - "BFS" → "boutons floraux séparés", ...
- Des stades relatifs : "préfloraison", "préfermeture", ...
- Des stades locaux : "assouplissement des pellicules"





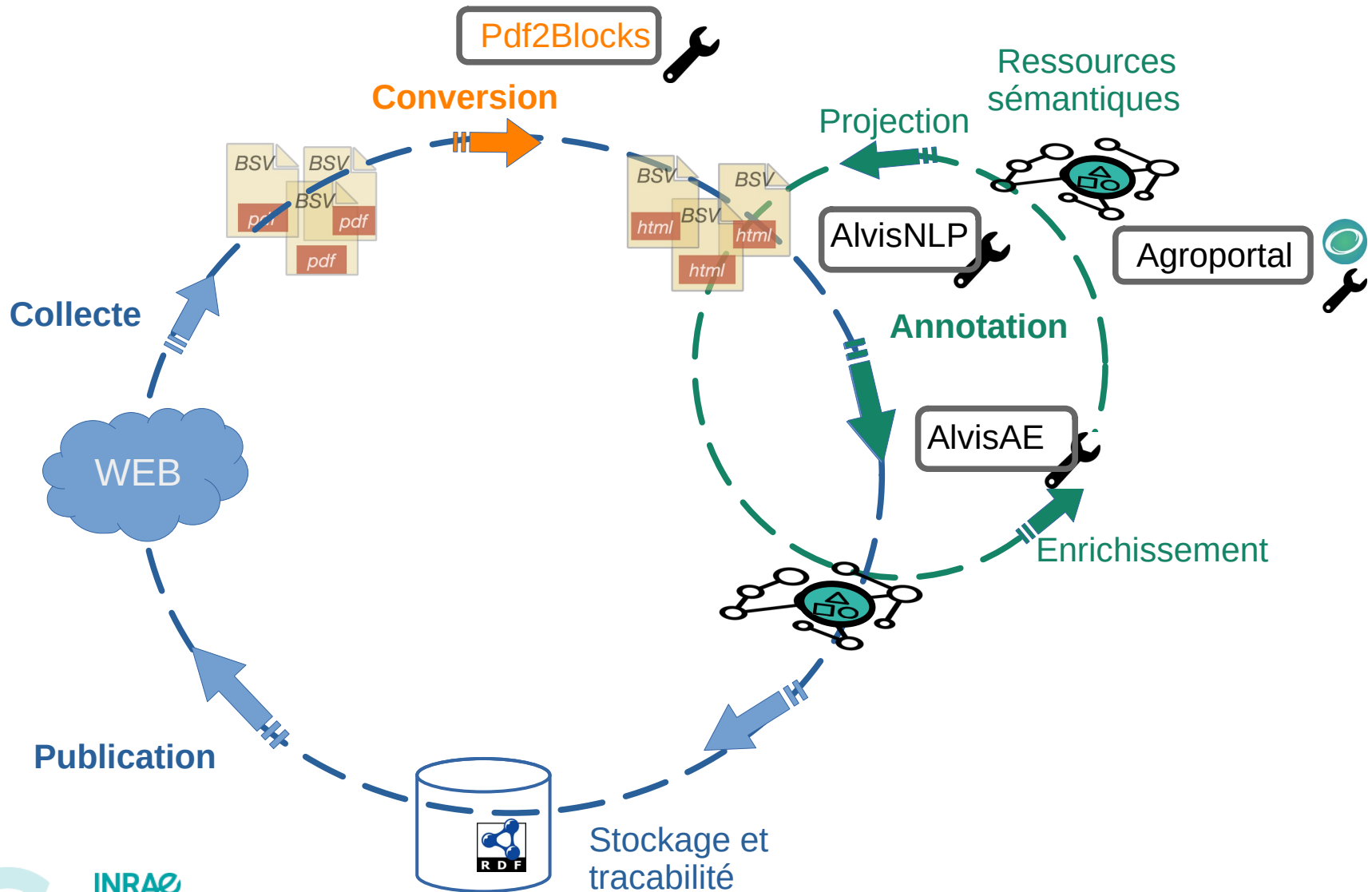




Annotation des Bulletins de Santé du Végétal



Cycle de vie des BSV ...



Les Corpus de BSV

GIT : <https://forgemia.inra.fr/bsv/corpus-bsv/>

| Corpus | nb BSV | Culture | spatial | temps | commentaire |
|---------------------|-----------|-------------------|--------------------|------------------|----------------------------|
| Vespa | 19 448 | toute | France | 2009-2016 | collecte semi automatique |
| Test VESPA | 497 | toute | France hors DomTom | 2009-2016 | sélection automatique |
| Test ALEA | 150 | toute | France hors DomTom | 2009-2016 | sélection aléatoire |
| D2KAB | > 22 000 | toute | France | 2009-2023 | collecte automatique |
| Test D2KAB | 230 | Viti, GC, Maraïc. | France hors DomTom | 2019 | sélection manuelle |
| Test Viti Alsace | 187 | Viti | Alsace | 2011-2022 | sélection manuelle |
| Test Annotés | 25 | Viti | Alsace | 2011-2022 | sélection manuelle |
| Bilans Viti | 18 | Viti | régions viticoles | 2021 | sélection semi automatique |



pdf2blocks v1.2

Conversion pdf 2 html





BULLETIN DE SANTÉ DU VÉGÉTAL CORSE

VITICULTURE N°5 – 18 juillet 2019

ÉCOPHYTO

AGRICULTURES & TERRITOIRES
CHAMBRE D'AGRICULTURE CORSE

SOMMAIRE

- Stades phénologiques
- Mildiou
- Oïdium
- Pyrâle du Daphné
- Vers de la grappe
- Flavescente dorée
- Prévision météo
- Liens utiles

ANIMATEUR FILIERE : CRVI
Rédacteur : Gilles Salvo

Structures partenaires :
CA2B, SCA UVIB, Domaine Comte Peraldi, JC Montell, Clos Capitano, CANICO, CAP

Directeur de publication :
Pierre Acquaviva,
Président de la Chambre d'Agriculture de Corse
15 Avenue Jean Zuccarelli
20200 BASTIA
Tel : 04 95 32 84 40
Fax : 04 95 32 84 43
<https://corse.chambres-agriculture.fr>
Crédit photo : CRVI de Corse CA2B

A RETENIR

Oïdium : rester vigilant
Tordeuses de la grappe : se préparer au vol de 3^{ème} génération
Pyrâle du Daphné (ou des agrumes) : surveiller les populations !
Flavescente dorée : rechercher les symptômes

• STADES PHENOLOGIQUES

La majorité des cépages se trouvent au stade fermeture de la grappe (L).
Le début de véraison apparaît sur cépages précoces, toujours avec quelques jours de retard par rapport à l'année 2018 :



Stade L : fermeture de la grappe



Stade M : début véraison

• MILDIOU

Biologie : voir BSV antérieurs
Observation : Pas d'évolution de la maladie. Les rares foyers ont séché, et on n'observe pas de rot brun.
Les pluies de la mi-juillet ont pu engendrer de nouvelles contaminations ou repiquages, dont les symptômes seront visibles à partir de la fin de cette semaine. Cela concerne particulièrement le nord de la côte orientale (25 à 100 mm), ainsi que la région Patrimoine / Cap corse (30 à 50 mm). La Balagne (15 mm) ainsi que le sud de l'île (moins de 10 mm) sont moins impactés.

Fax : 04 95 32 84 43 <https://corse.chambres-agriculture.fr>
Crédit photo : CRVI de Corse, CA2B

Action pilotée par le Ministère chargé de l'agriculture, avec l'appui financier de l'Office National de l'Eau et des Milieux Aquatiques, par les crédits issus de la redevance pour pollutions diffuses attribués au financement du plan ECOPHYTO.

STADES PHENOLOGIQUES

La majorité des cépages se trouvent au stade fermeture de la grappe (L). Le début de véraison apparaît sur cépages précoces, toujours avec quelques jours de retard par rapport à l'année 2018 :

Stade L : fermeture de la grappe

Stade M : début véraison

MILDIOU

Biologie : voir BSV antérieurs

Observation : Pas d'évolution de la maladie. Les rares foyers ont séché, et on n'observe pas de rot brun. Les pluies de la mi-juillet ont pu engendrer de nouvelles contaminations ou repiquages, dont les symptômes seront visibles à partir de la fin de cette semaine.

```
<p blk="10"><small class="misc">Structures partenaires : CA2B, SCA UVIB, Domaine Com
<p blk="11"><small class="misc">Directeur de publication : Pierre Acquaviva, Présidé
<p blk="12"><small class="misc">Action pilotée par le Ministère chargé de l'agricult
<h3 class="title 3" blk="13">STADES PHENOLOGIQUES</h3>
<p class="paragraph" blk="14">La majorité des cépages se trouvent au stade fermeture
<h4 class="title 4" blk="15">Stade L : fermeture de la grappe</h4>
<h4 class="title 4" blk="16">Stade M : début véraison</h4>
<h3 class="title 3" blk="17">MILDIOU</h3>
<p class="paragraph" blk="18">Biologie : voir BSV antérieurs<br />Observation : Pas
<footer class="bottom page" blk="19">Bulletin de Santé du Végétal Corse<br /></foot
<footer class="bottom page" blk="20">Viticulture N°5 18 juillet 2019<br /></footer>
<footer class="bottom page" blk="21">1/5<br /></footer>
<hr /><!-- ===== Page 2 ===== -->
<h3 class="title 3" blk="22">OÏDIUM</h3>
```

Git <https://gitlab.irstea.fr/copain/pdf2blocs>

<https://zenodo.org/records/6605450>

Textual Annotation and Provenance Ontology (TAPO)

Une extension (spécialisation de OA) pour les BSV

QCs : Classer les annotations en fonction du type d'activité qui les à produite (génération automatique ou manuelle) et du type d'entités annotées (Stade, Culture, ...)

1) Quelles sont les annotations de plantes cultivées ?....

URI : <https://opendata.inrae.fr/tap-def>

Git <https://forgemia.inra.fr/bsv/textual-annotation-and-provenance>

DOI : <https://doi.org/10.57745/1RWGZK>

Version 1.0

Formalisation : OWL



Conclusion et Perspectives

Annotation d'entités,

besoin d'annotation de relations binaires entre entités

puis reconstruire un graphe de relations N-aires (des observations)

Publier les annotations manuelles et automatiques du corpus des 25 BSV Viti Alsace : date, localité, culture, cépage, stade, bioagresseurs.

Mettre au point des méthodes pour compléter, valider, enrichir les annotations automatiques afin d'identifier des observations de stades et des observations d'agression

Publier les annotations automatiques du corpus D2KAB de 22 000 BSV

Et jouer avec...

Transformer les annotations automatiques en index des BSV

Proposer des interfaces de recherche des BSV adaptées aux communautés d'utilisateurs



Variables scientifiques

Les stades sont utilisés de manières différentes dans les variables scientifiques

- Date d'observation d'un stade pour une culture (plante) donnée
- % de la parcelle cultivée ayant atteint un stade (si $>$ à 50 % par convention on dit que la parcelle a atteint ce stade)
- Période d'un stade pour une culture (plante ou parcelle) donnée (date de début d'observation et date de fin d'observation)
 - calculer la durée de la période
 - déterminer la période d'une phase de développement de la culture (plante ou parcelle) pendant laquelle on fera d'autres mesures.
- Autres mesures où la culture (plante ou parcelle) doit être à un stade fixé
 - modifie l'objet scientifique





Questions ?



INRAE

Séminaire Sésame

Novembre 2024 / C Roussey