



Introduction à la formalisation des connaissances sous forme d'ontologies

Catherine Roussey, Stéphan Bernard, Ghislain Auguste Ateazing

► To cite this version:

Catherine Roussey, Stéphan Bernard, Ghislain Auguste Ateazing. Introduction à la formalisation des connaissances sous forme d'ontologies. Doctorat. Data Science School 2022, Bénin. 2022, pp.64. hal-04289831

HAL Id: hal-04289831

<https://hal.science/hal-04289831>

Submitted on 16 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



AfIA

Association française
pour l'Intelligence Artificielle

Introduction à la ➤ formalisation des connaissances sous forme d'ontologies

Catherine ROUSSEY, Stephan BERNARD, Ghislain ATEMEZING

équipe Copain, UR TSCF, INRAE Centre Clermont ARA

Axe Informatique, UMR MISTEA, INRAE Centre Occitanie Montpellier

Mondeca, Paris



INRAE



Plan

Definitions: connaissances, ontologie

Technologies du Web sémantique, graphe de connaissances → LOD

Ingénierie des ontologies

3 cas d'usages

1) Publication de données capteurs sur le LOD

2) Projet COSWOT : Système contextuel adaptatif

3) Projet D2KAB : Annotation sémantique de bulletins d'alerte agricole



Définitions

“Une donnée est une information qui peut être enregistrée, traitée, analysée ou communiquée, quelle que soit sa nature. Lorsque la donnée n’a pas encore été traitée ou contextualisée, on parle de donnée brute”. (Guide pratique pour la gestion des données de la recherche irstea)

34

Donnée: un élément d’information,

percevable par l’humain, 

manipulable par la machine



Information: donnée + sens

Sens : il faut fournir à l’humain du contexte



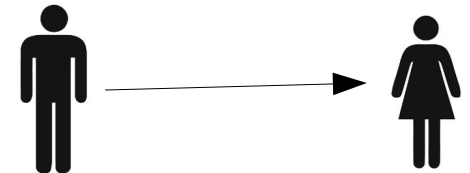
Type → structure de données



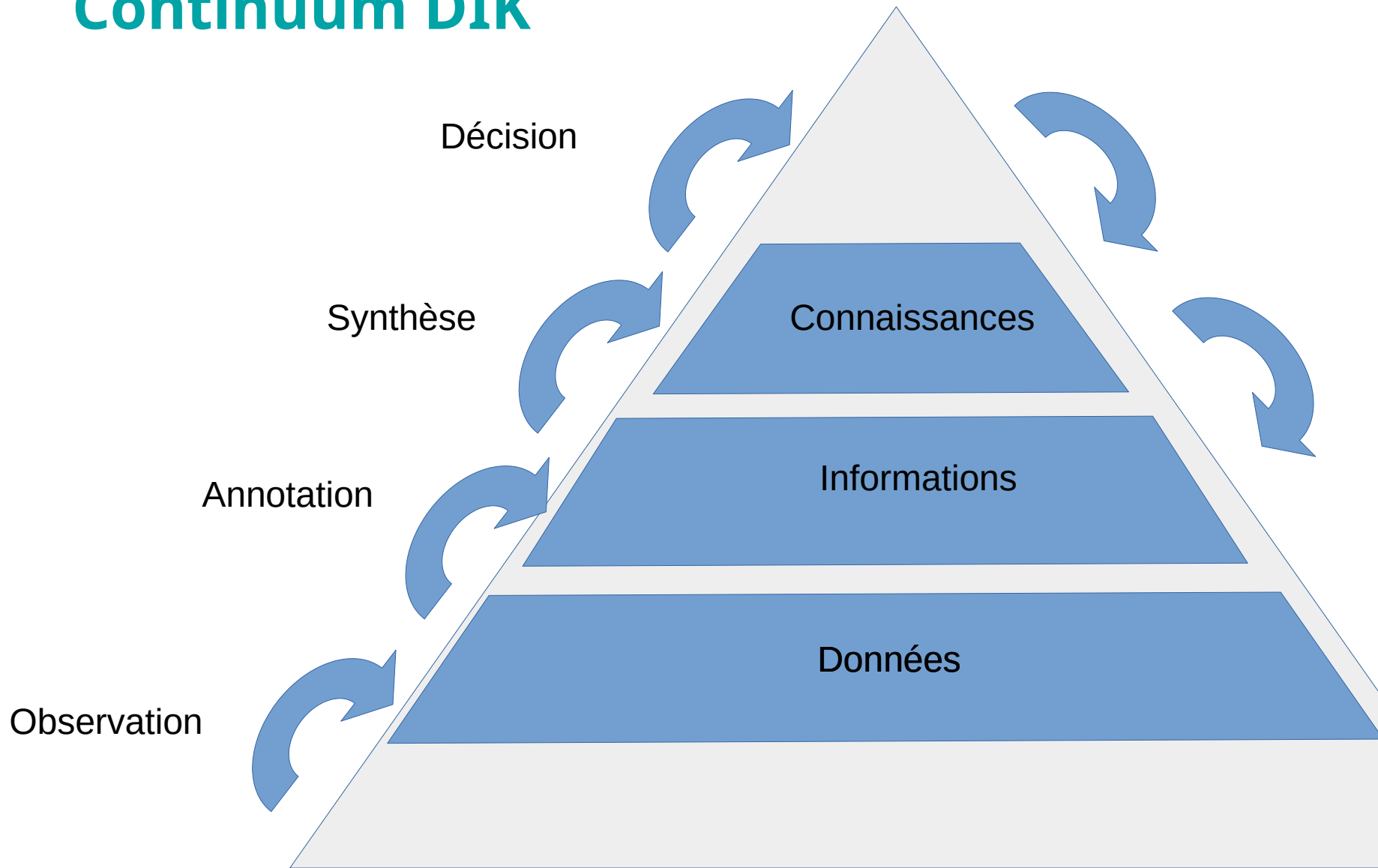
Connaissance: information formalisée dans un langage...

Pour inférer de nouvelles connaissances

“marié à ” est une relation symétrique



Continuum DIK



Ontologie: un peu d'histoire

Il était une fois, les systèmes experts une branche de l'Intelligence Artificielle
dans les années 90, évolution vers les systèmes à base de connaissances
Séparer les connaissances de résolution de problème des connaissances du domaine

Ontologie: formalisation des connaissances du domaine consensuelles et stables

Ontologie: "*spécification explicite d'une conceptualisation*" (Gruber, 1993)

Problème: spécification + conceptualisation = diagramme UML

Ontologie "*une spécification explicite et formelle d'une conceptualisation partagée*" (Studer et al., 1998)

Ontologie du web: les schémas de données formalisés dans un des langages des technologies du Web sémantique

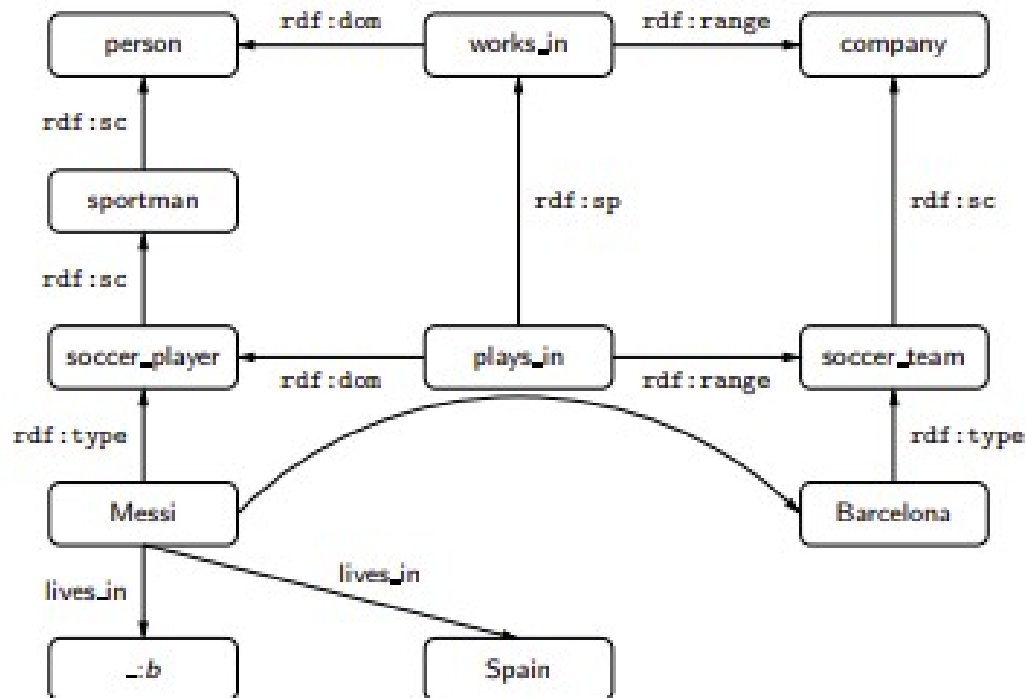
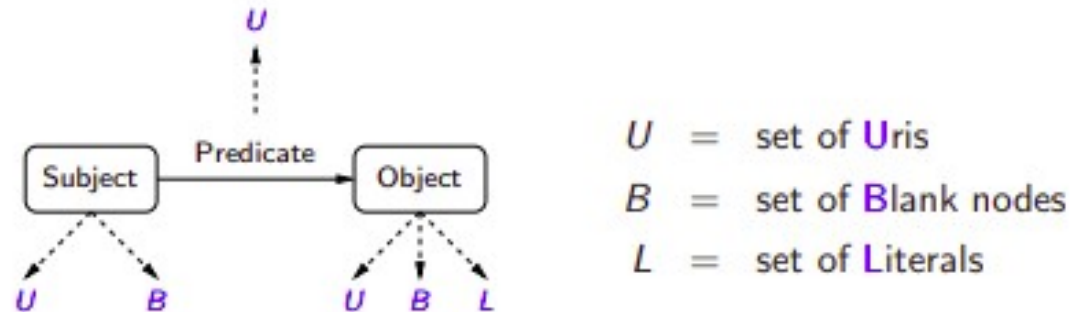


Resource Description Framework RDF

Credit Semantic Web Data/RDF/SPARQL

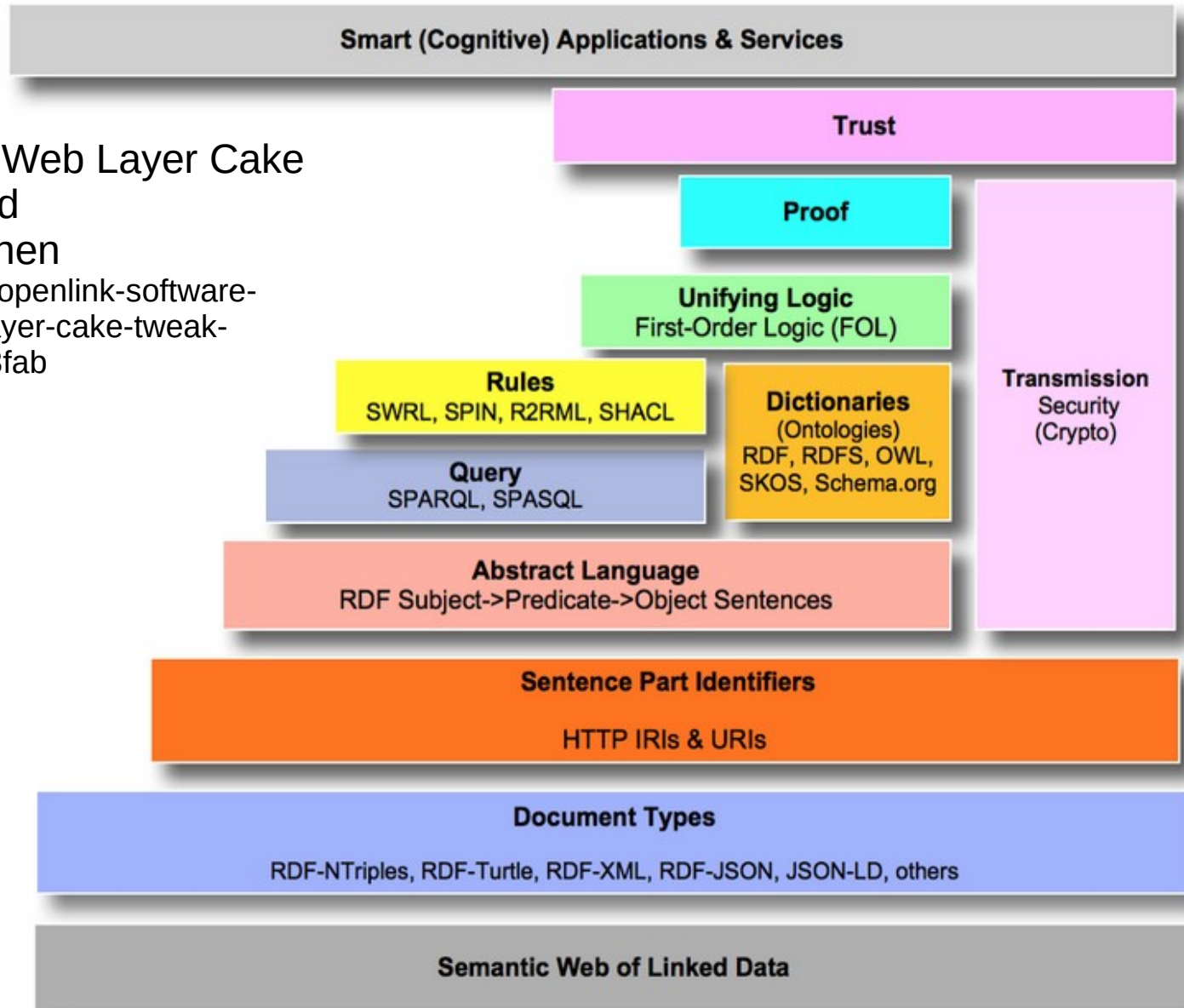
<http://homepages.inf.ed.ac.uk/libkin/teach/beijing2018/rdfslides.pdf>

Triplet RDF



Technologies Web Sémantique

Credit Semantic Web Layer Cake
Tweak, Explained
Kingsley Uyi Idehen
<https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab>



Ontologie du Web W3C

et graphe de connaissances

Dans le monde du web sémantique, une ontologie est l'ensemble des concepts et des relations utilisés pour décrire un domaine d'intérêt. Les mots ontologie et vocabulaire sont utilisés conjointement. Le mot ontologie est employé quand le vocabulaire de concepts et de relations est assez complexes et peut contenir par exemple des contraintes: conditions nécessaire et/ou suffisante d'appartenance. (W3C)

Une ontologie sert à (W3C):

- **Normaliser** les termes du domaine: leur associer un identifiant (URI), un label et une signification
- **Typer** les éléments de ce vocabulaire pour définir un schéma, documenté et réutilisable: des classes, des propriétés etc...
- Aider l'intégration de données multi-sources
- Organiser les connaissances d'un domaine: publication de ressources et de leur métadonnées descriptives sur le web (Linked Data)
- Produire des **inférences**

Un graphe de connaissances est l'ensemble des faits décrit à l'aide d'une ontologie et publié sur le web



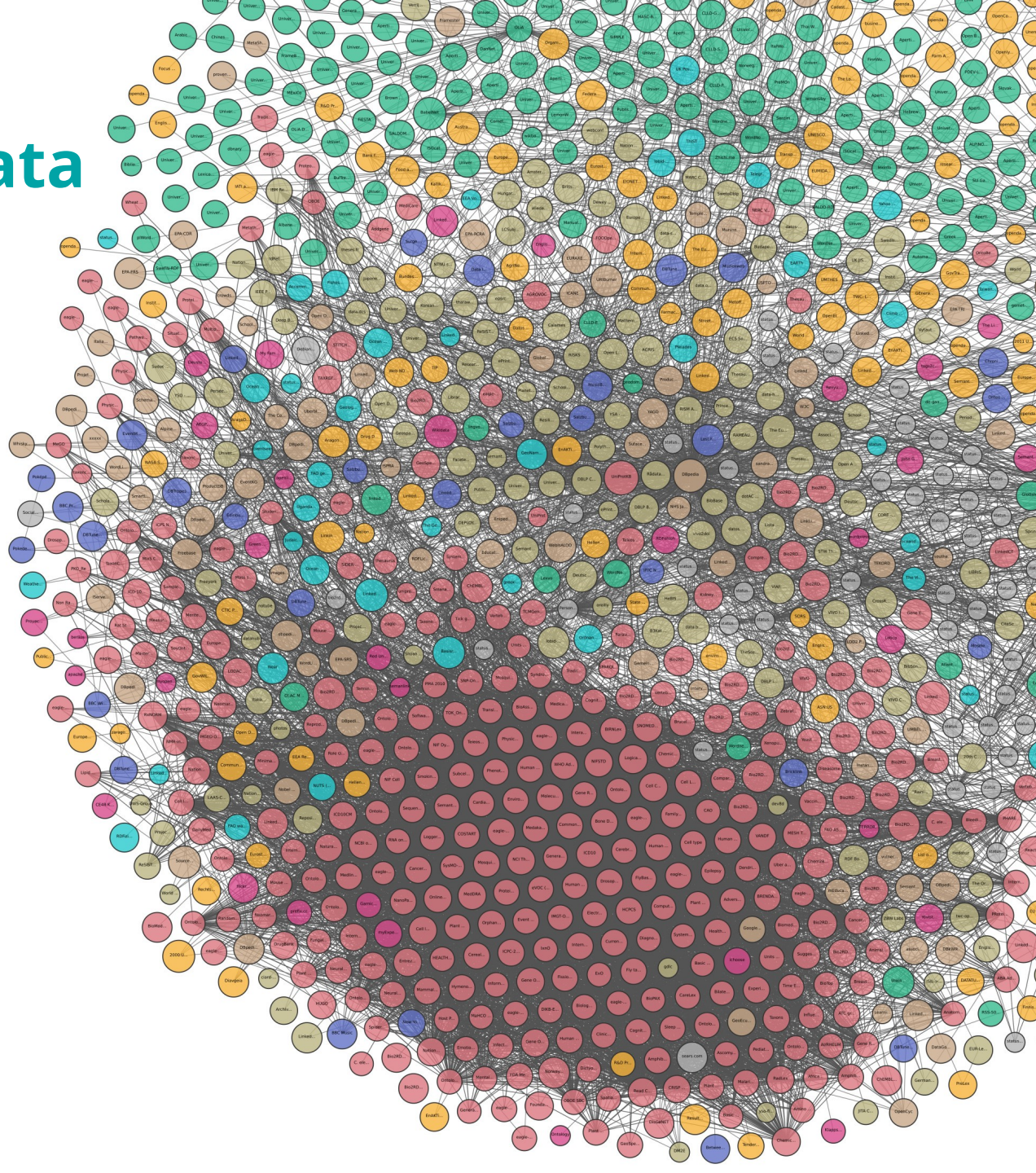
Linked Open Data (LOD)

Credit lod-cloud.net

Web de données Liées

Video europeana qu'est ce que le web des données

<https://vimeo.com/49231111>



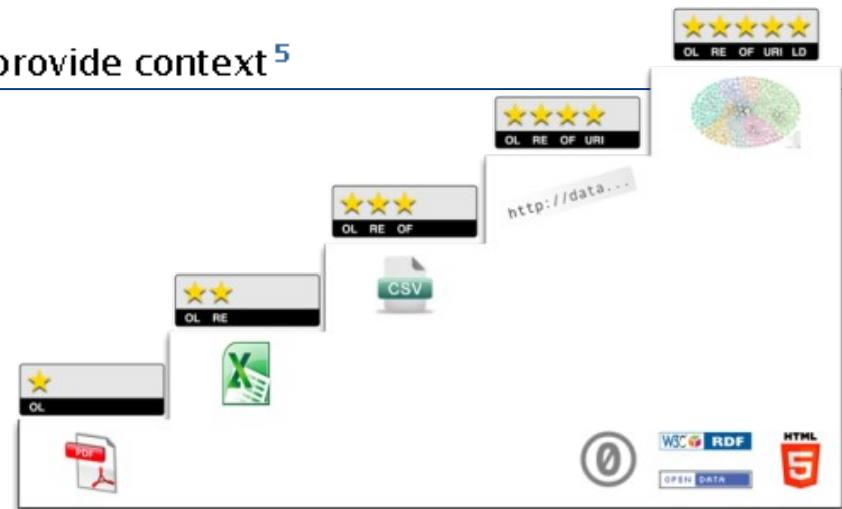
INRAE

Formalisation des connaissances
Octobre 2022 / C Roussey

Modèle 5 étoiles

Credit Tim Berners-Lee, <http://5stardata.info/>

- ★ make your stuff available on the Web (whatever format) under an open license¹
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)²
- ★★★ use non-proprietary formats (e.g., CSV instead of Excel)³
- ★★★★ use URIs to identify things, so that people can point at your stuff⁴
- ★★★★★ link your data to other data to provide context⁵



Données FAIR



Credit FAIR in a nutshell. Image: ARDC 2018 - CC-BY 4.0.

<https://librarycarpentry.org/Top-10-FAIR/2018/12/01/aus-government-data/>

INRAE

Formalisation des connaissances

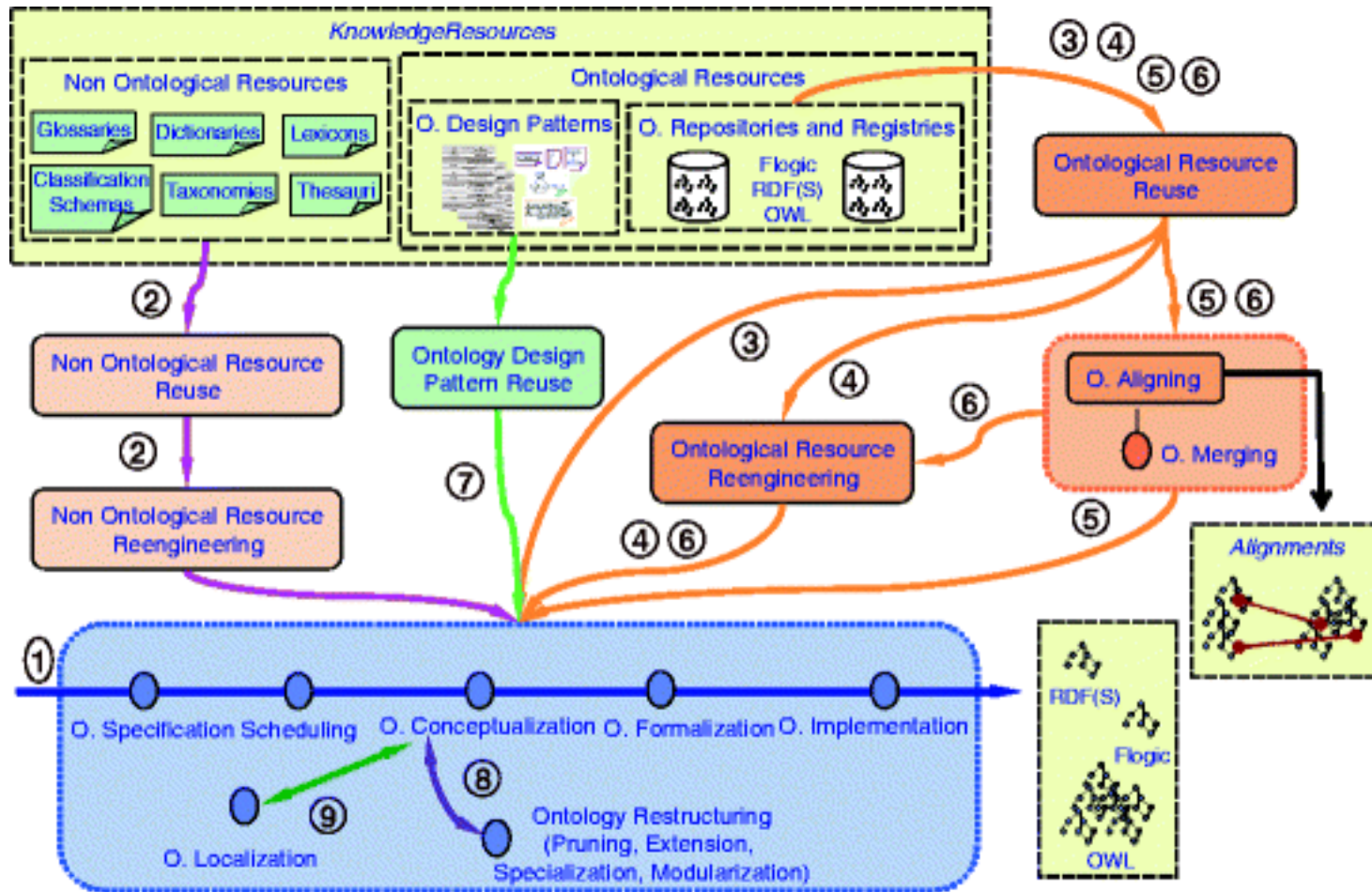
Octobre 2022 / C Roussey

Ingénierie des connaissances



Ingénierie des ontologies

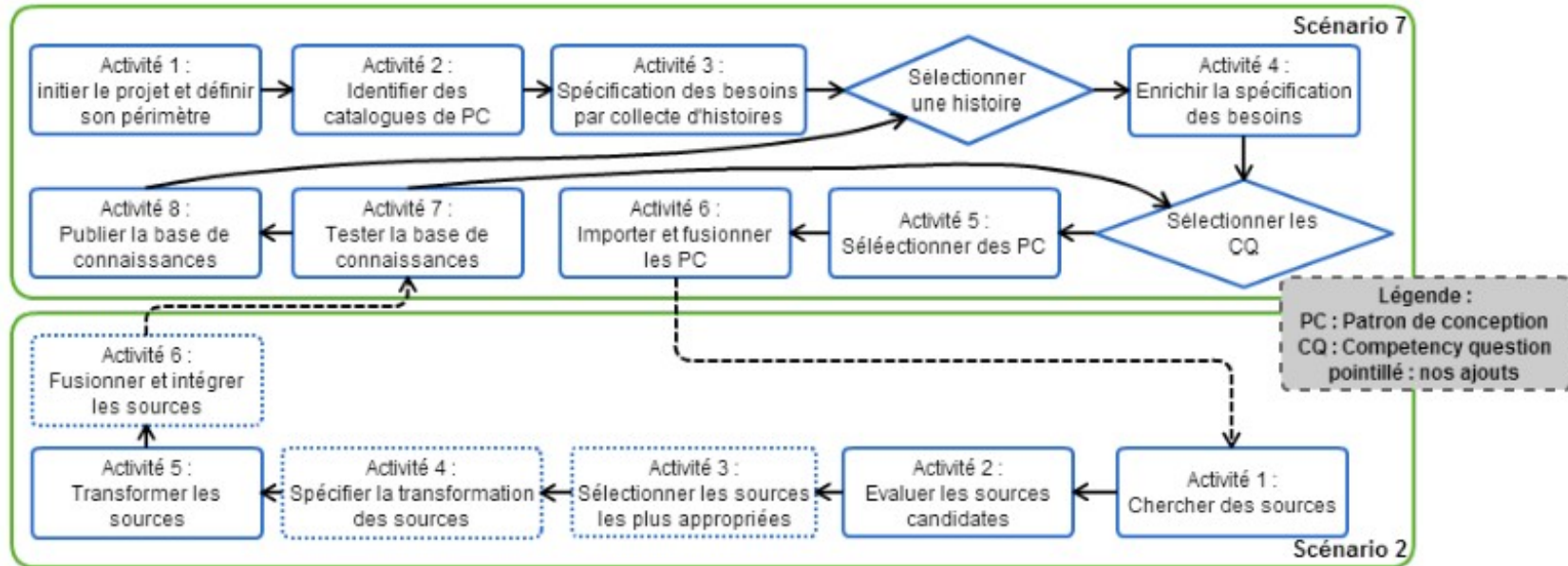
la méthodologie NeOn = un ensemble de méthodes



Crédit Suárez-Figueroa M.C., Gómez-Pérez A., Fernández-López M. (2012) The NeOn Methodology for Ontology Engineering. In: Suárez-Figueroa M., Gómez-Pérez A., Motta E., Gangemi A. (eds) Ontology Engineering in a Networked World. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24794-1_2

Ingénierie des ontologies

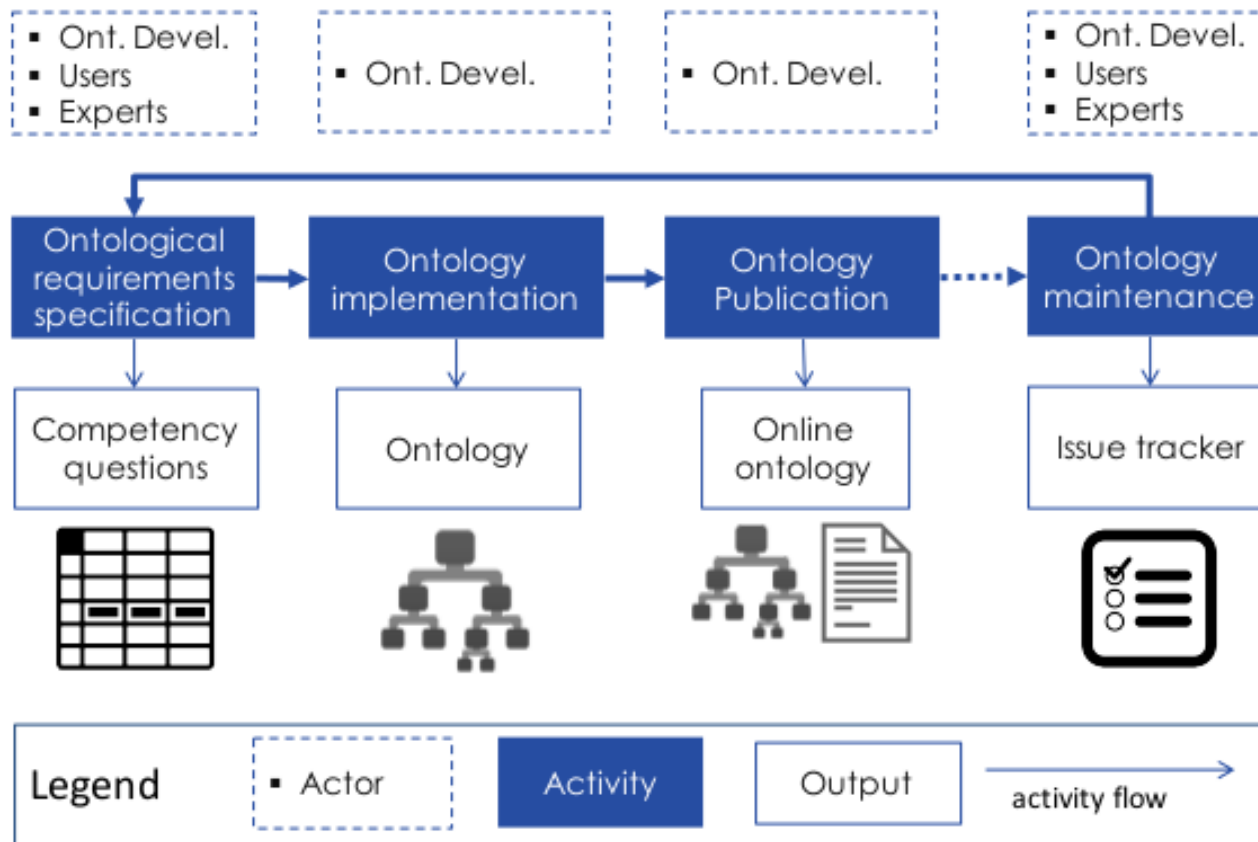
Neon propose plusieurs scénarios que l'on peut combiner



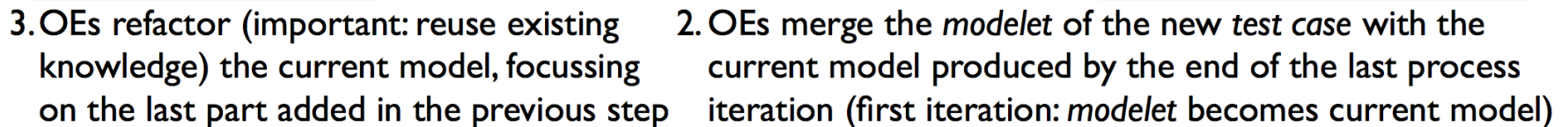
Crédit Fabien AMARGER, Vers un système intelligent de capitalisation de connaissances pour l'agriculture durable : construction d'ontologies agricoles par transformation de sources existantes
Thèse de doctorat en informatique, 2015, Université Toulouse 2

Ingénierie des ontologies

La méthode Linked Open Terms (LOT)



La méthode SAMOD



Ontology Design Pattern

Design Pattern : une modélisation de petite taille répondant à un problème précis

ODP sont présents :

- dans les ontologies fondationnelles BFO ou DoI Ce
- Dans les ontologies coeur de domaine (core domain) comme Semantic Sensor Network

Ils peuvent prendre différents noms et forme:

- Le Module AgronomicTaxon pour la Thèse de Fabien Amarger
- Le modelet dans la méthode SAMOD



Consensus : Comment l'atteindre ?

Thèse de Abdel Kader Keita

Aider les experts du domaine à ce mettre d'accord sur les définitions des concepts (méthode Delphi)

Illustrer les liens entre concepts par des images (photo, carte, ...)

Thèse de Fabien Amarger

Extraire des sources des informations communes et les qualifier :

- Détecter les candidats
- Évaluer leur degré de confiance
- Gérer les incohérences



Où trouver des ontologies FAIR ?

Organismes de normalisation

- World Wide Web Consortium (W3C) <https://www.w3.org/>
- European Telecom Standard Institut (ETSI) <https://www.etsi.org/>

Répertoires

- Linked Open Vocabulary (LOV) <https://lov.linkeddata.es/dataset/lov/>
- EMBL-EBI Ontology Lookup Service (OLS) -> biomedical
<https://www.ebi.ac.uk/ols/index>
- Agroportal -> agriculture, environnement, INRAE , ... <http://agroportal.lirmm.fr/>

Communauté de pratique

- Community group CROP ontology <https://www.cropontology.org/>
- Basic Formal Ontology <https://basic-formal-ontology.org/>



Publication de données

➤ capteurs sur le web de données liées

INRAE fournisseur de données

INRAE possède :

- des réseaux de capteurs pour observer une variété de phénomènes naturels: météo, hydro, pollution etc...
- des bases relationnelles pour stocker les données d'expérimentation, ...
- des bases bibliographiques: bulletins d'alertes, articles scientifiques, ...

Comment rendre accessible ces données ?

au personnel d'INRAE mais pas que...

Le site web BDOH permet de rechercher les fichiers de données capteurs (séries temporelles) des observatoires hydrologiques d'INRAE (ex-Irstea)



Hétérogénéité ...

Données de capteurs: observations automatisées qui doivent être structurées, complétées, validées et documentées.

Différents types de capteurs: observation d'une variété de phénomènes naturels

Différentes types de données: des mesures aux métadonnées descriptives des capteurs

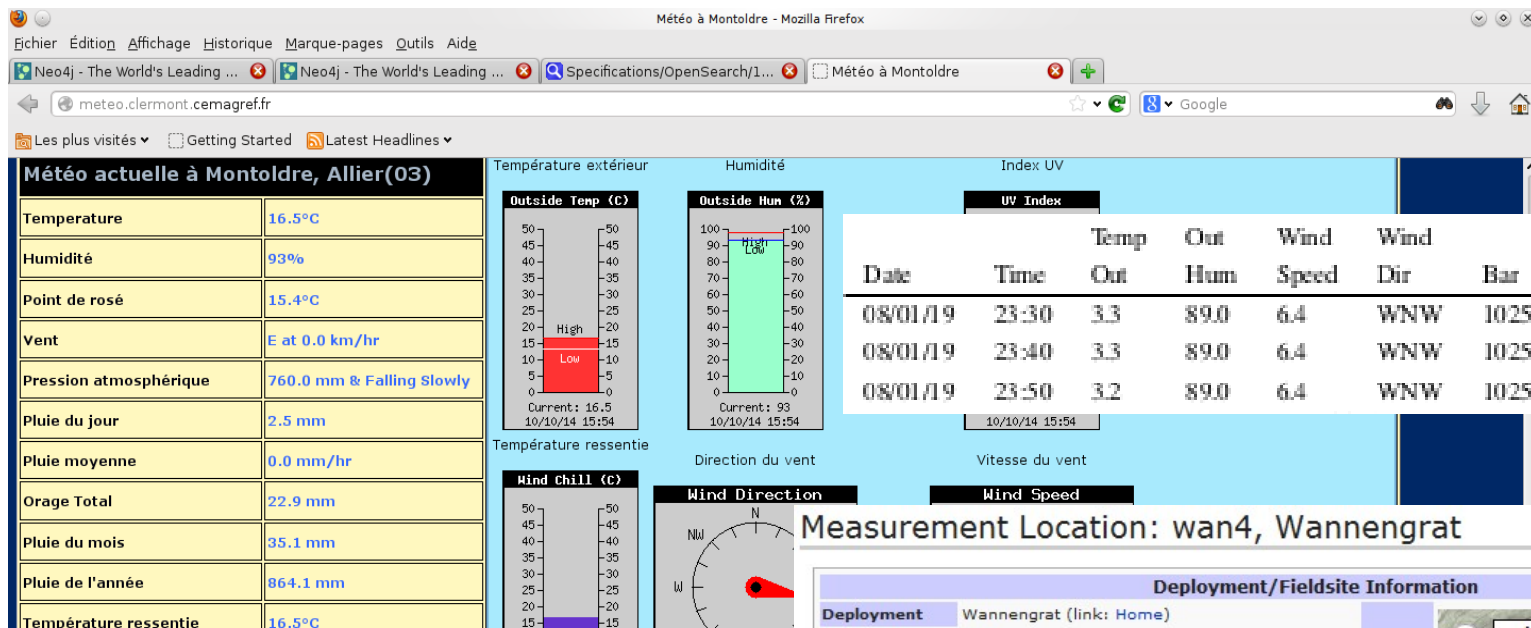
Différents formats / schéma de données / interface d'interrogation

Différents usages: agriculture, gestion des risques, etc...

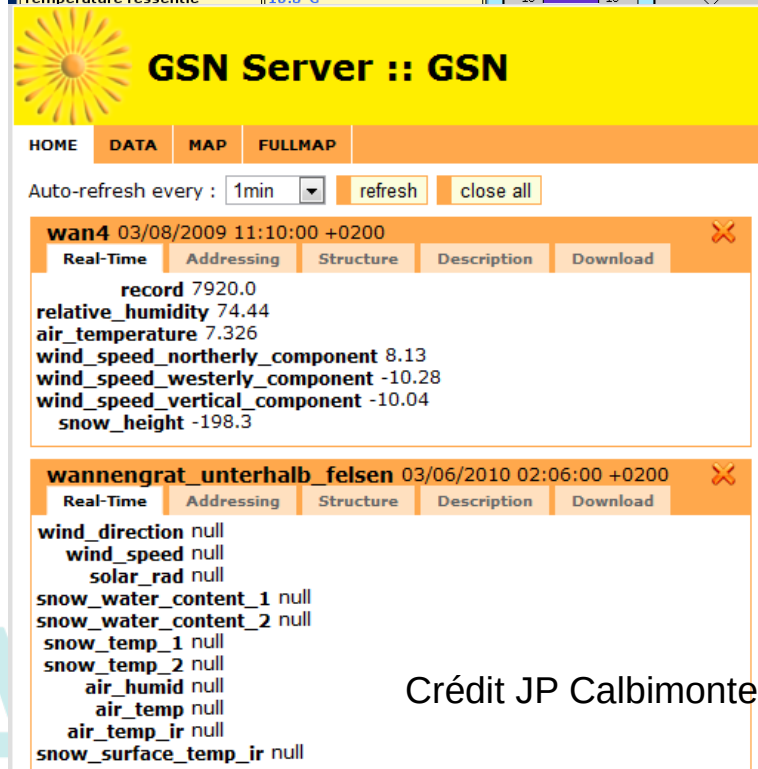
→ Difficulté à interroger de manière unifiée les données issues de différentes stations

→ Difficulté à intégrer des données de divers réseaux dans une même application



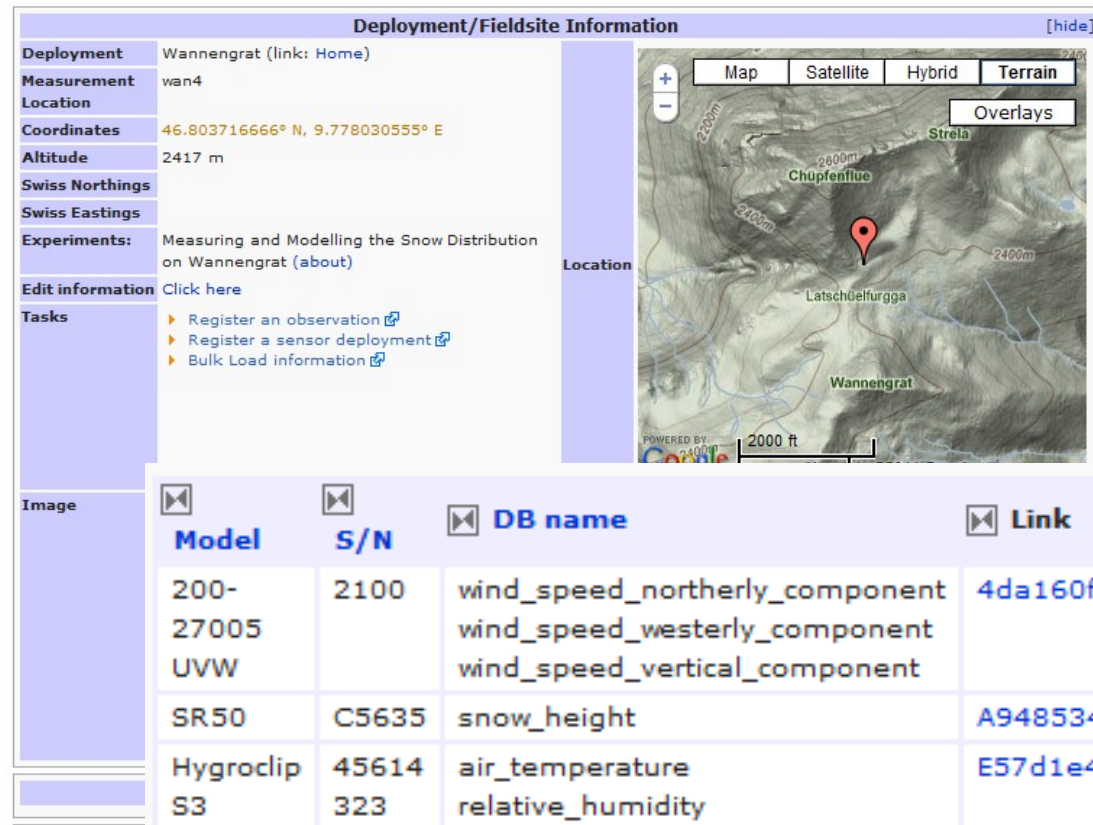


Crédit G André



Crédit JP Calbimonte

Measurement Location: wan4, Wannengrat



Publication de données capteurs

Accès uniformisé aux données et métadonnées des capteurs

« Standardisation » des données:

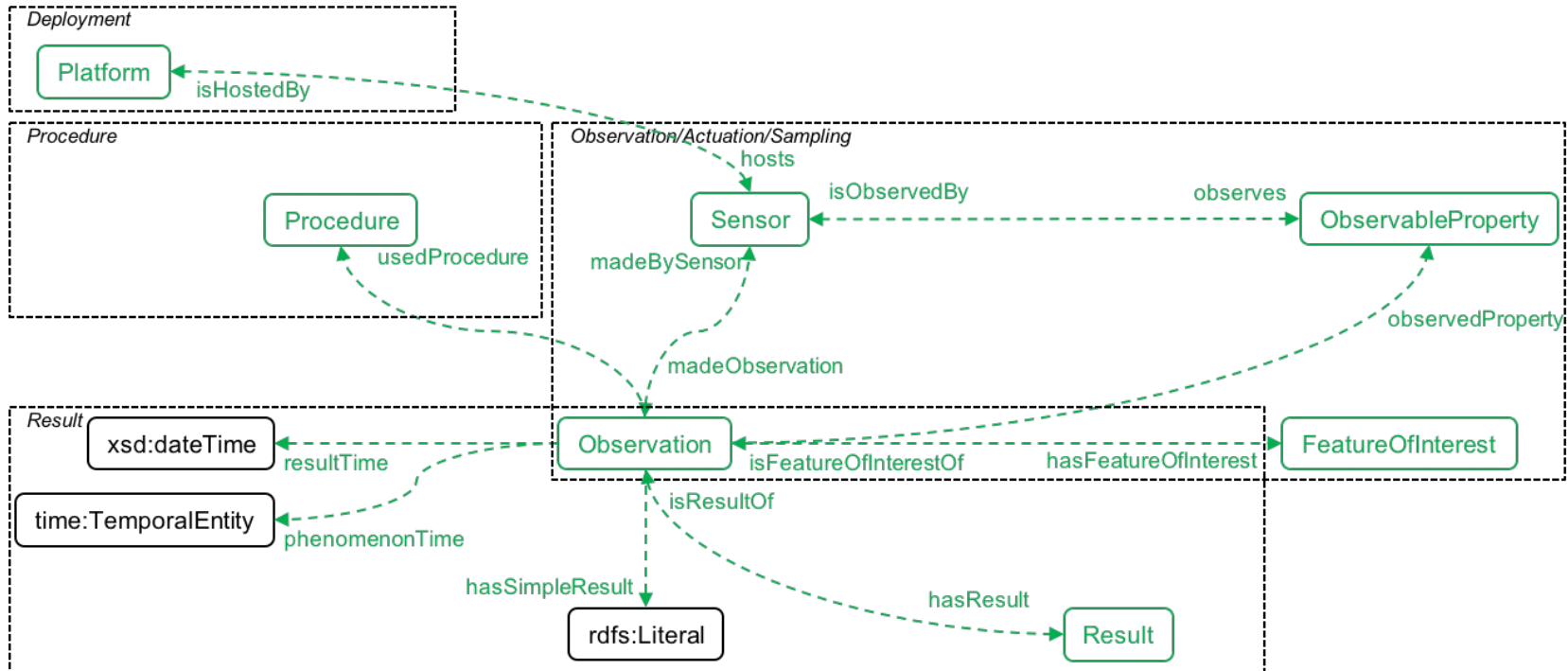
- Réutilisation de schémas (ontologies) existants déjà documentés
- Combinaison de différents schémas pour coller à notre besoin
- Facilité l'intégration de nos données avec d'autres jeux de données
- Facilité la découverte de notre jeu de données (visibilité)

Credit C. ROUSSEY, S. BERNARD, G. ANDRÉ, D. BOFFETY. Weather Data Publication on the LOD using SOSA/SSN Ontology. Semantic Web Journal, 2019
<http://www.semantic-web-journal.net/content/weather-data-publication-lod-using-sosassn-ontology-0>



Semantic Sensor Network (SSN)

Sensor Observation Sample Actuator (SOSA)



Credit W3C working group
<https://www.w3.org/TR/vocab-ssn/>

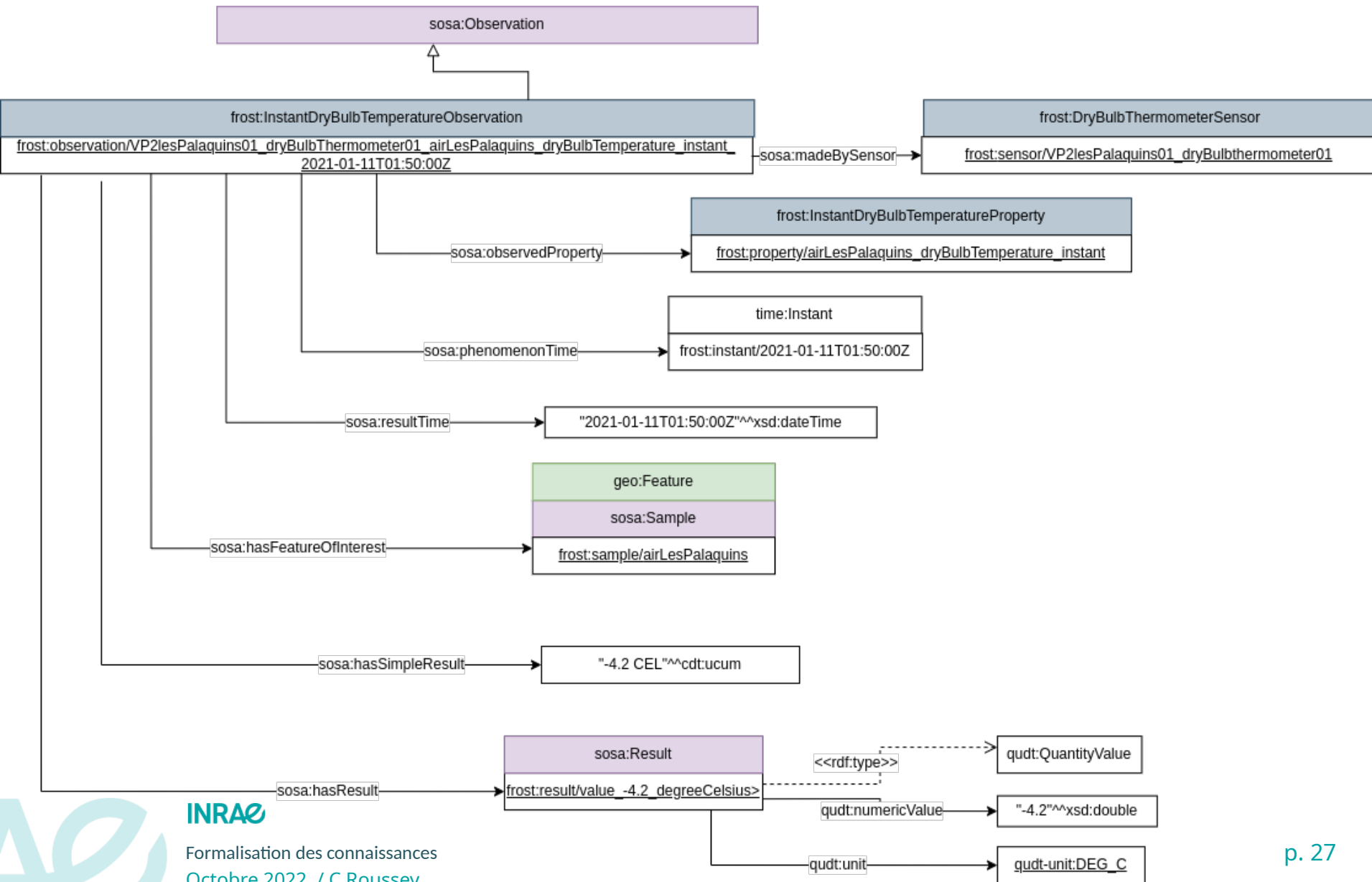
Réseau d'Ontologies

Semantic Sensor Network : le squelette

- Sensing Device
Ontology for meteorological sensor (aws)
- Feature of Interest
SWEET: set of classes
- Observable Property
Thesaurus Climate and Forecast (skos)
- Platform location
GeoSPARQL and Location Core Vocabulary (geosparql + locn)
- Observation result and phenomenon time
W3C Time Ontology (time)
- Observation value
Library of Quantity Kind and Units (qu + dim)

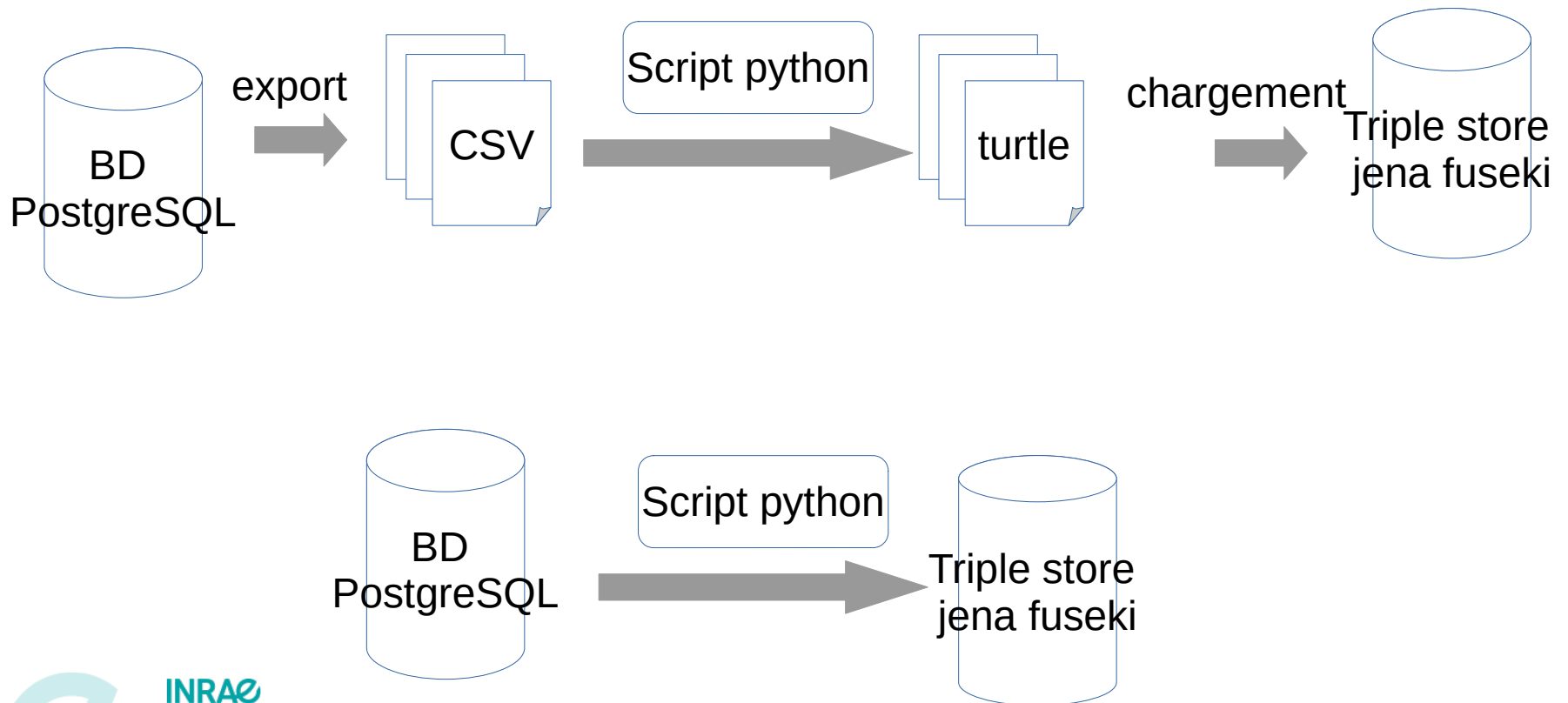


Observation



BD to triple store

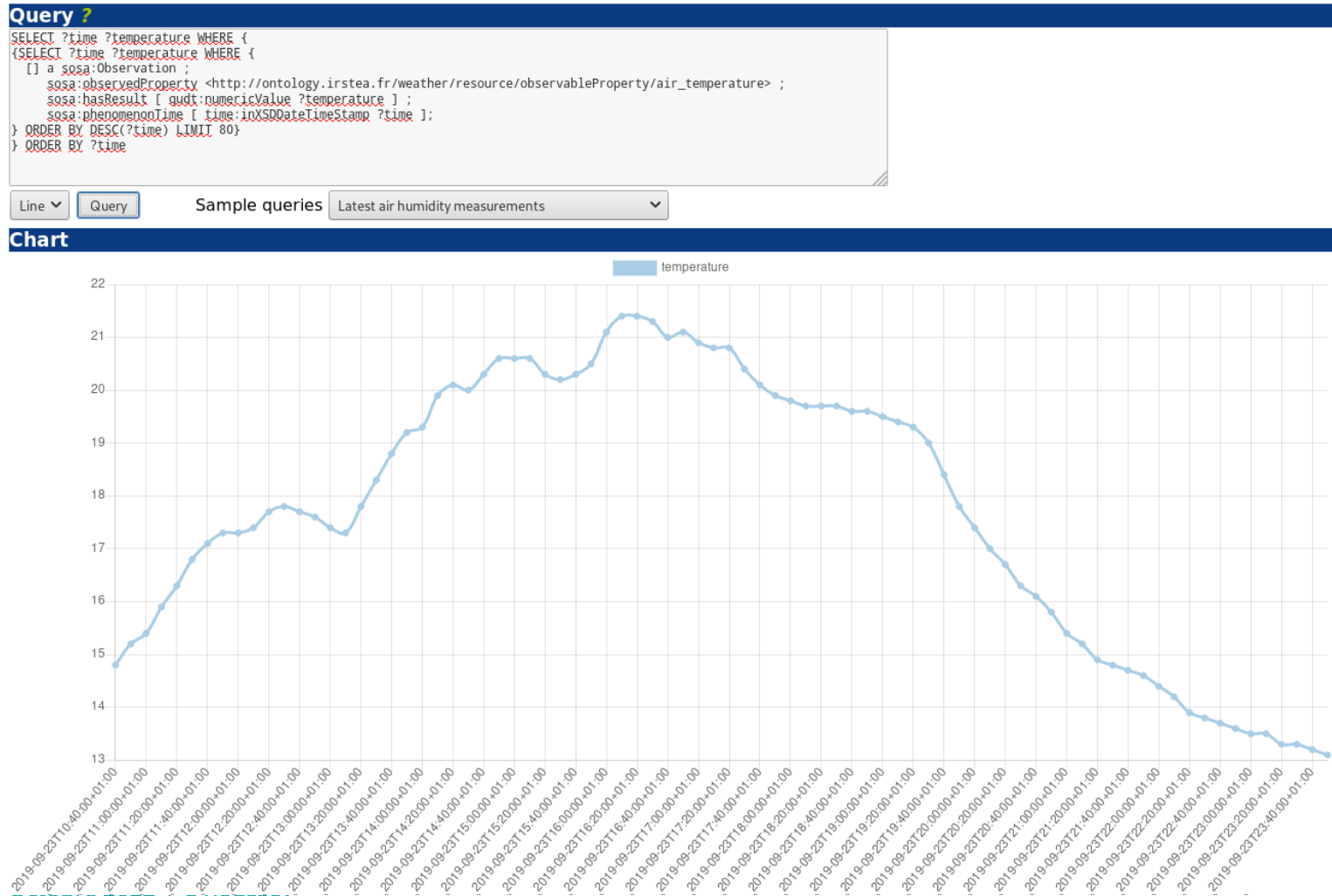
Credit Stephan Bernard stephan.bernard@inrae.fr



Interface d'interrogation

A SPARQL query interface using snorql is available at <http://ontology.irstea.fr/weather/snorql/>.

A SPARQL query interface that display numeric results as a graph is available at <http://ontology.irstea.fr/weather/snorql/test/>



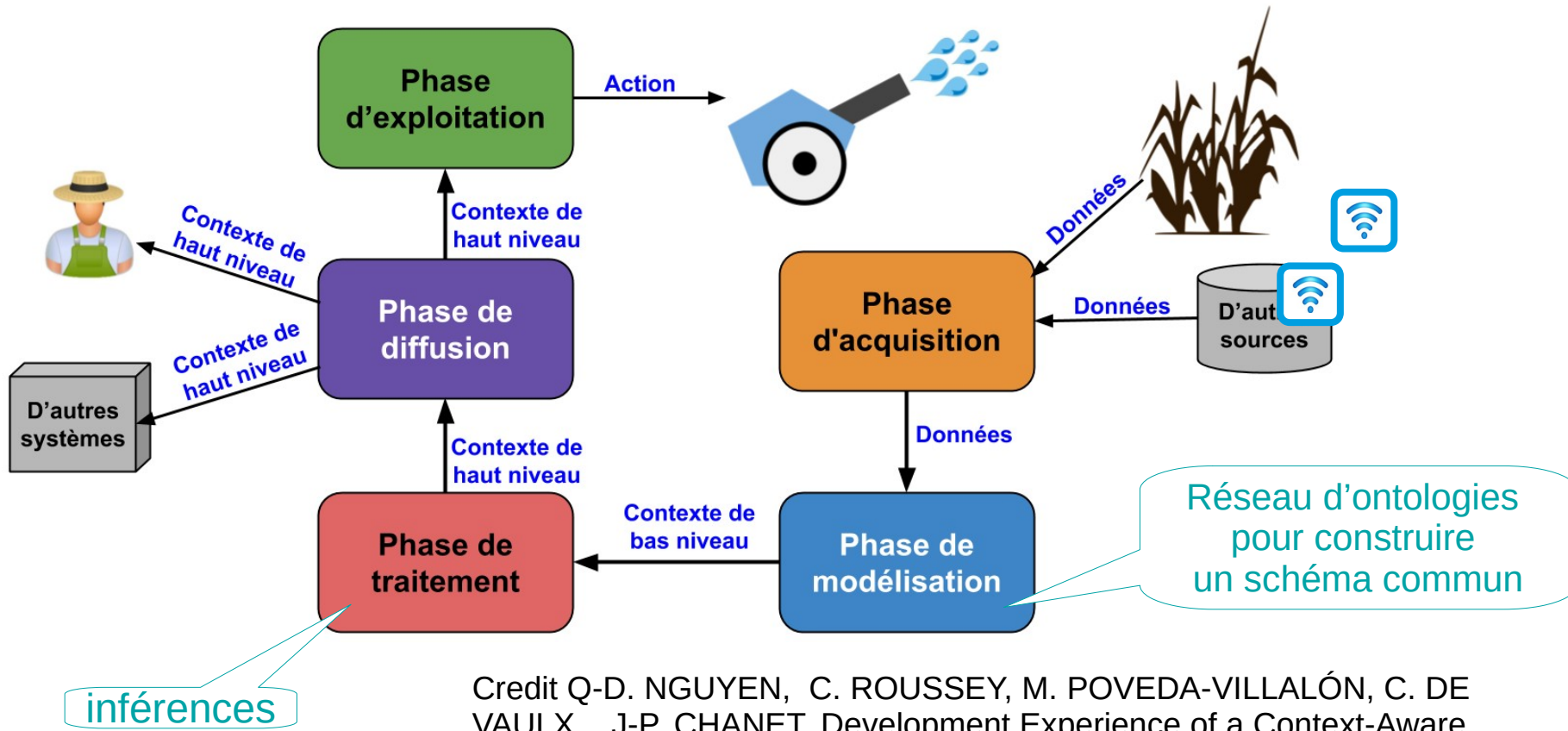


Ontologies pour les systèmes ➤ contextuels adaptatifs

Système contextuel adaptatif

Un système contextuel est un système qui utilise le contexte pour fournir des informations et des services appropriés à l'utilisateur. (Abowd et al., 1999)

Implanter la boucle observation → decision → action → observation



Credit Q-D. NGUYEN, C. ROUSSEY, M. POVEDA-VILLALÓN, C. DE VAULX, J-P. CHANET. Development Experience of a Context-Aware System for Smart Irrigation Using CASO and IRRIG Ontologies. Applied Science 2020, 10(5), 1803; <https://doi.org/10.3390/app10051803>

Système contextuel adaptatif

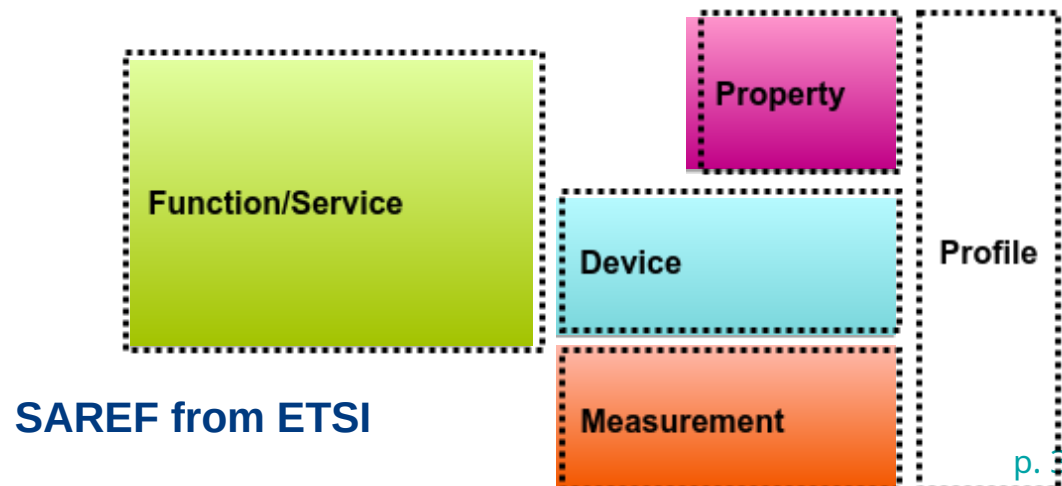
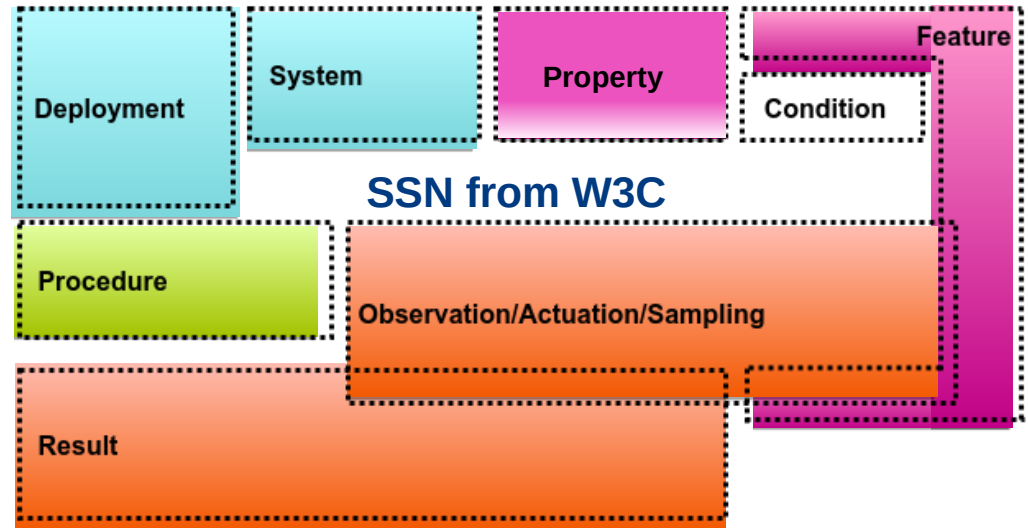
Combiner ontologie d'observation + ontologie de service

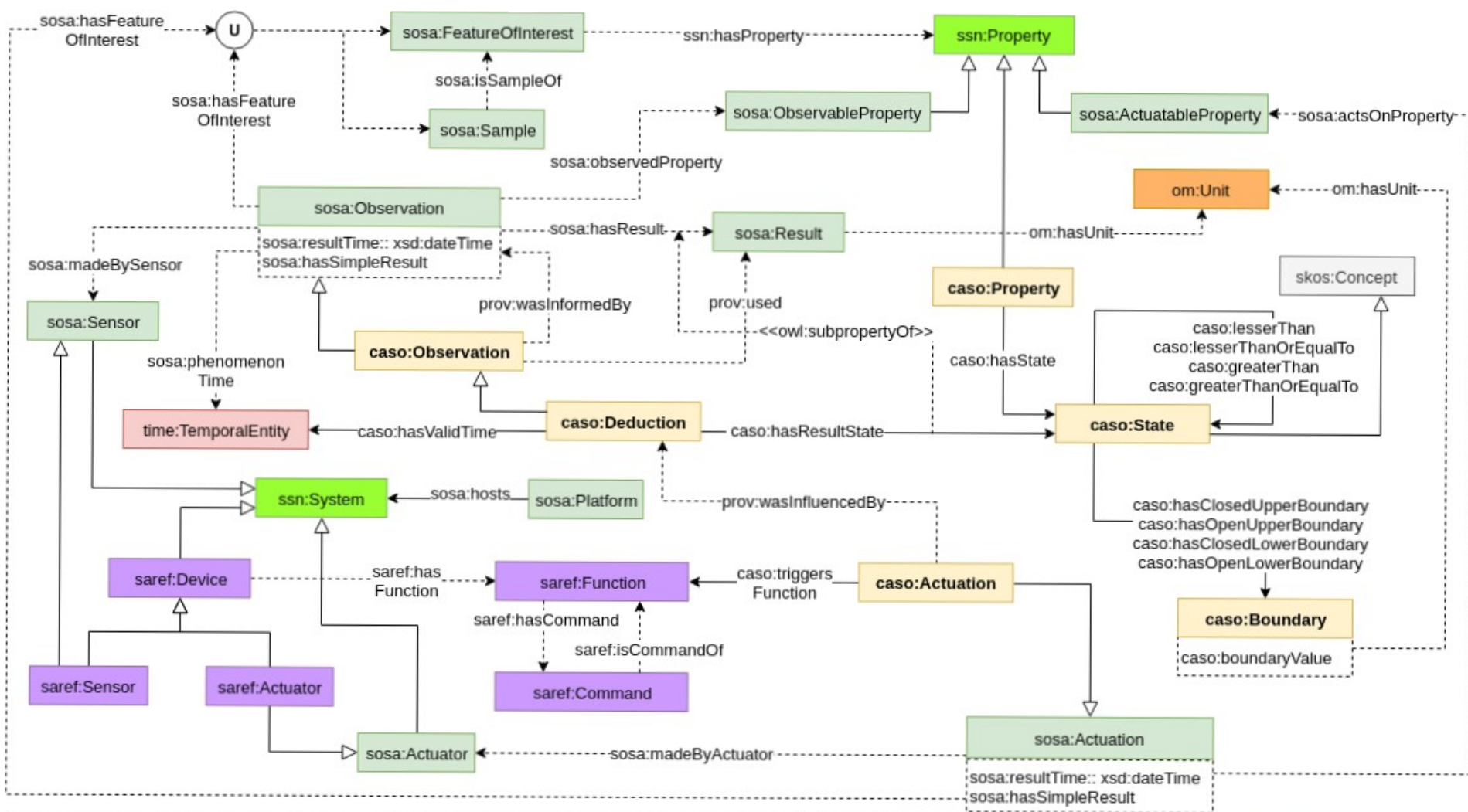
SSN décrit une situation de mesure :
qui, quoi, quand, comment.

Smart Appliance Reference (SAREF)
décrit des appareils connectés de
tous les domaines.

des ontologies noyaux auxquelles
d'autres ontologies se connectent
pour définir le schéma de données
de l'application cible.

Ces deux ontologies s'appliquent à
des domaines variés: agriculture,
santé, domotique.





Légende

Classe
Attribut dont "domain" est la classe attachée

- sous classe de —>
- propriété d'objet avec "domain" et "range" définis —>
- <<stéréotype>>
- ... propriété d'objet applicable à la classe attachée ...>

Classe d'Union (Anonyme)



Préfixes

owl: <http://www.w3.org/2002/07/owl#>
 sosa: <http://www.w3.org/ns/sosa/>
 ssn: <http://www.w3.org/ns/ssn/>
 saref: <http://www.w3id.org/saref#>
 time: <http://www.w3.org/2006/time#>
 om: <http://www.ontology-of-units-of-measure.org/resource/om-2/>
 prov: <http://www.w3.org/ns/prov>
 skos: <http://www.w3.org/2008/05/skos>

Ontology

caso: <https://w3id.org/def/caso#>

Conception: une méthode d'aggrégation pour la pluie

$\text{getRainDailyTotalQuantity}(p, d) = \sum \text{getRainQuantity}(p, t1, t2)$
on the condition that $[t1, t2] \subset [d \text{ 06:00:00}, d+1 \text{ 06:00:00}[$

where

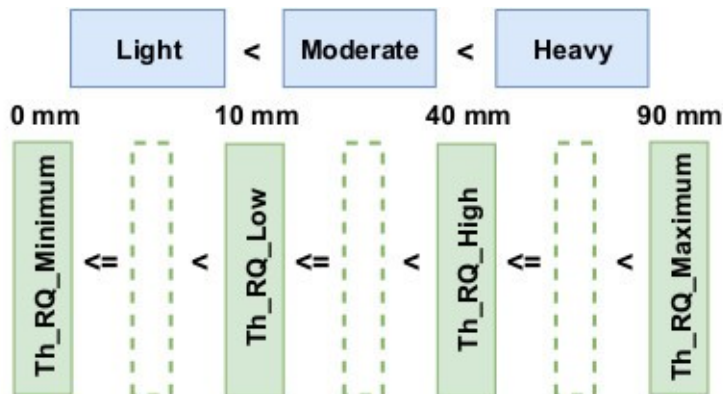
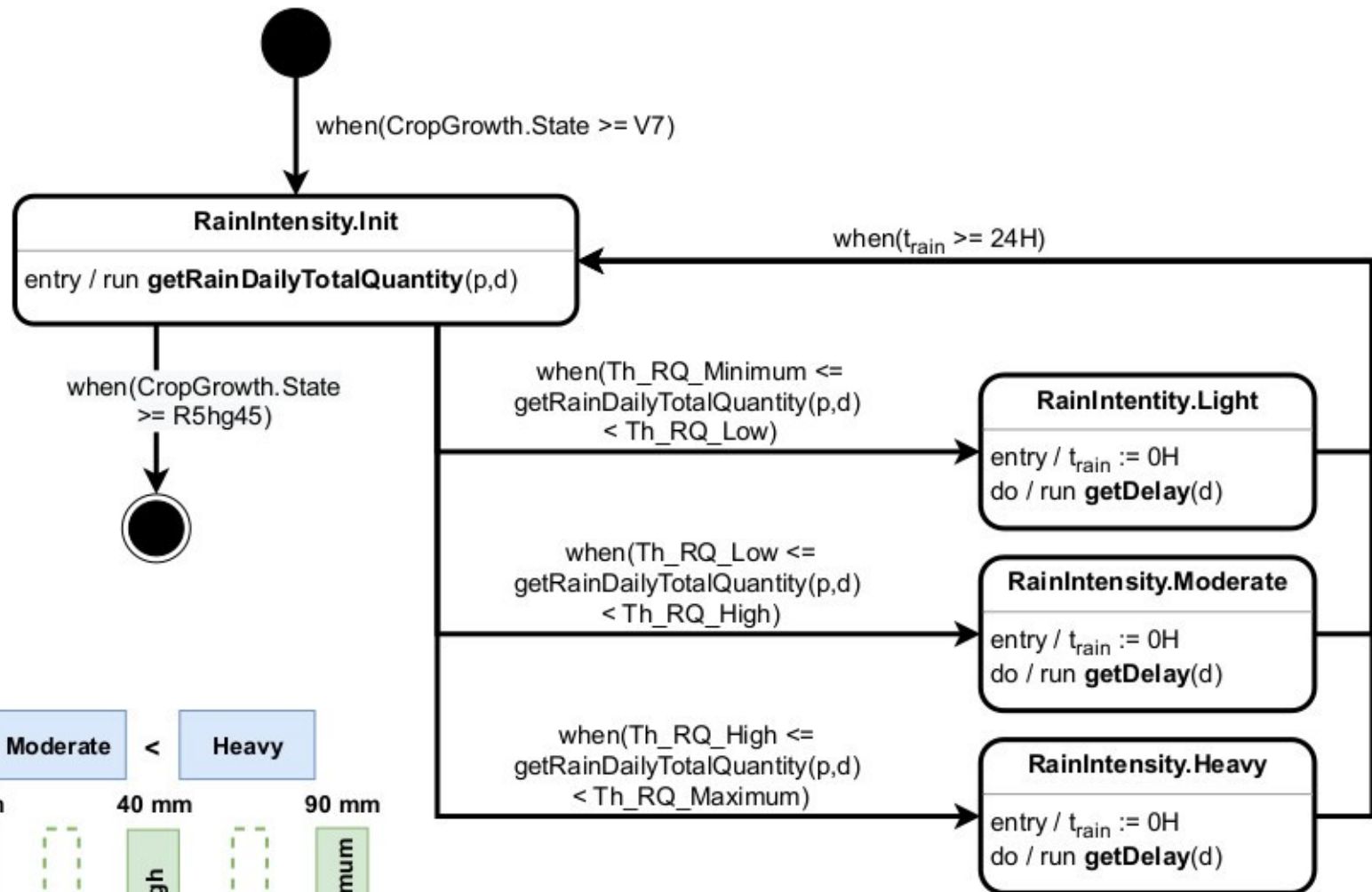
d: a specific day, e.g., 14/08/2013

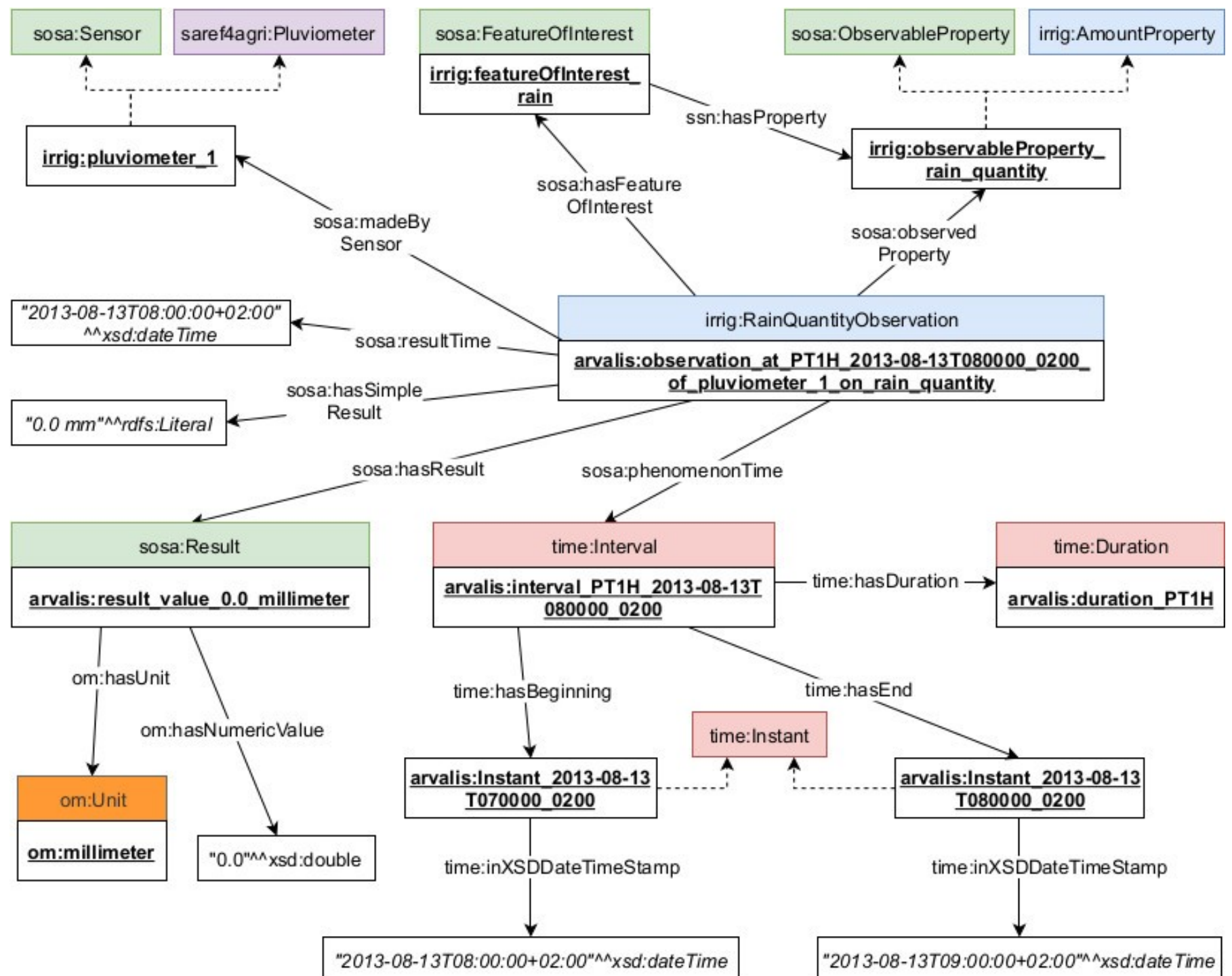
p: a pluviometer

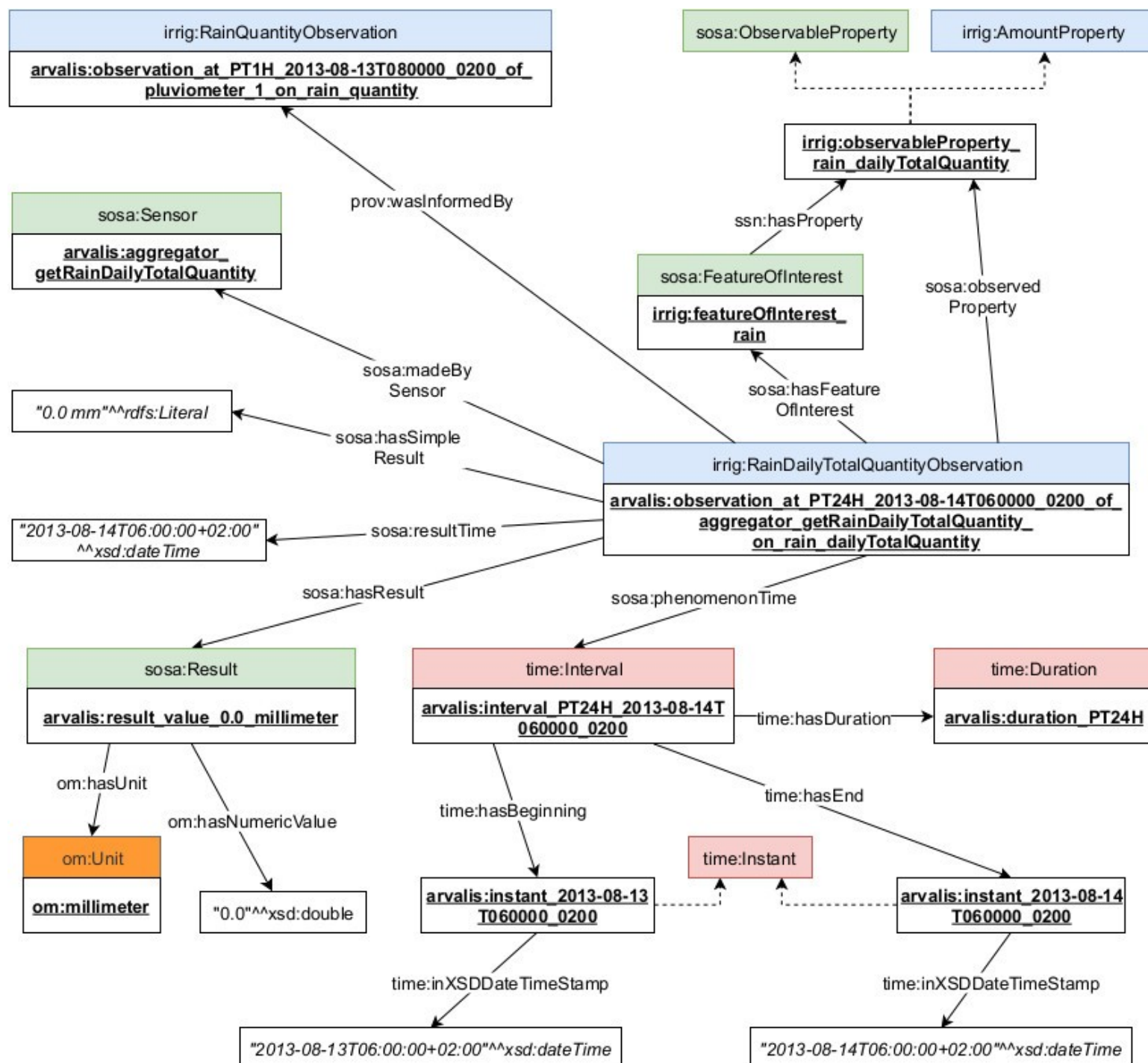
ti : a specific instant of day d, e.g., 08:00:00 on 14/08/2013

$\text{getRainQuantity}(p, t1, t2)$: the rainfall quantity value measured by the pluviometer p during the interval $[t1, t2[$.

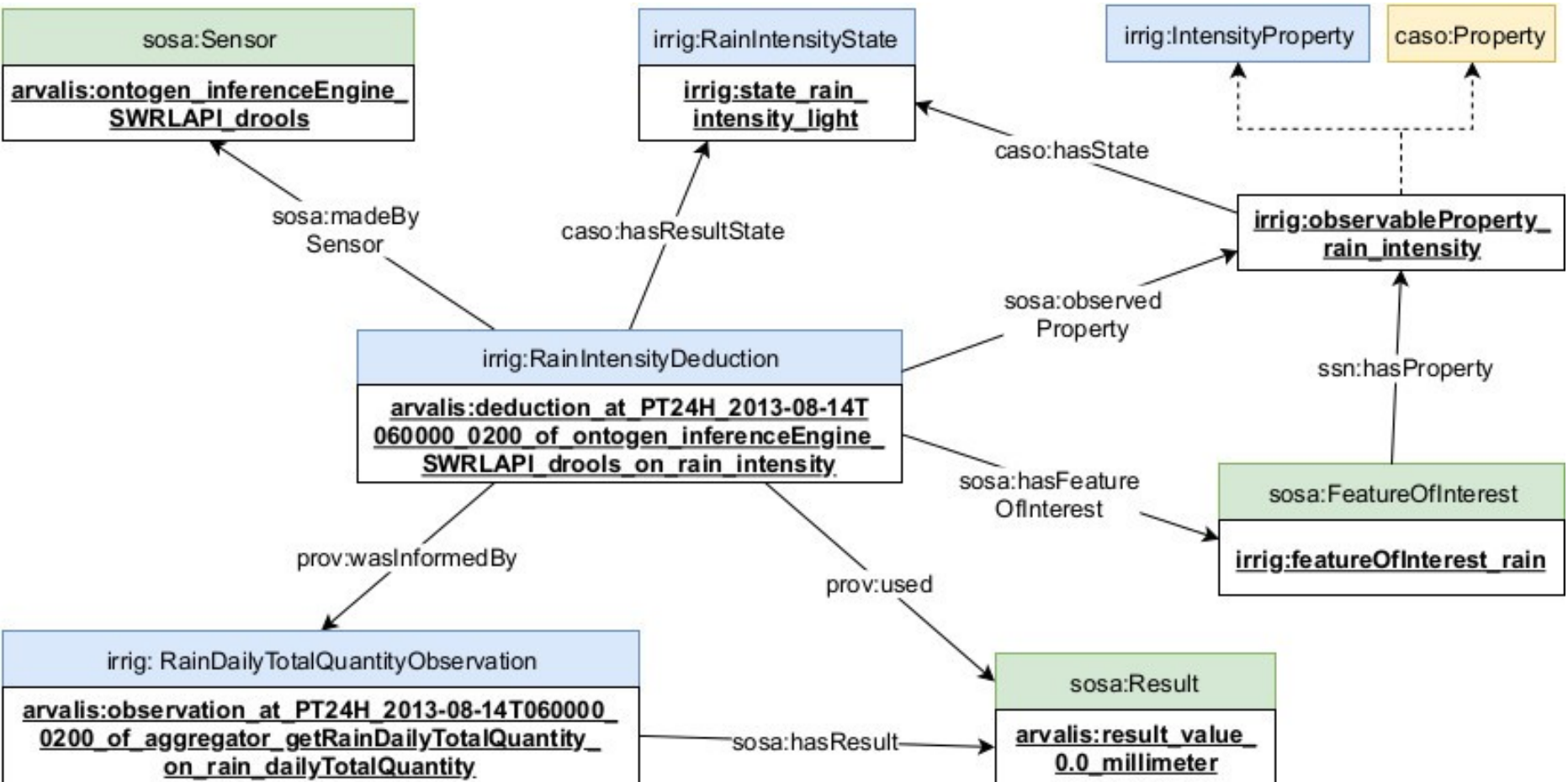
Conception: : diagramme d'état pour déduire les états de la pluie

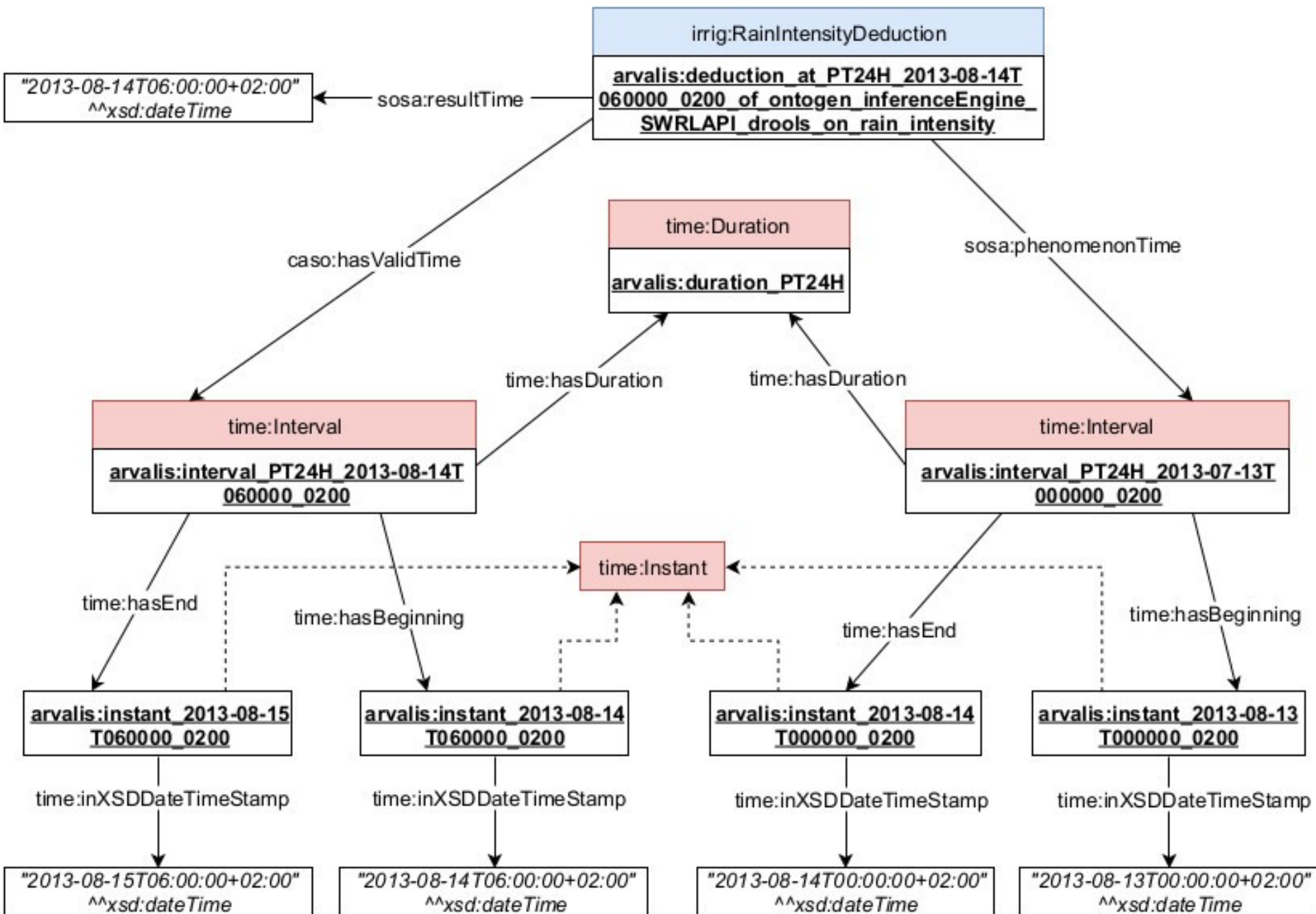






Pluie: Observation et Deduction

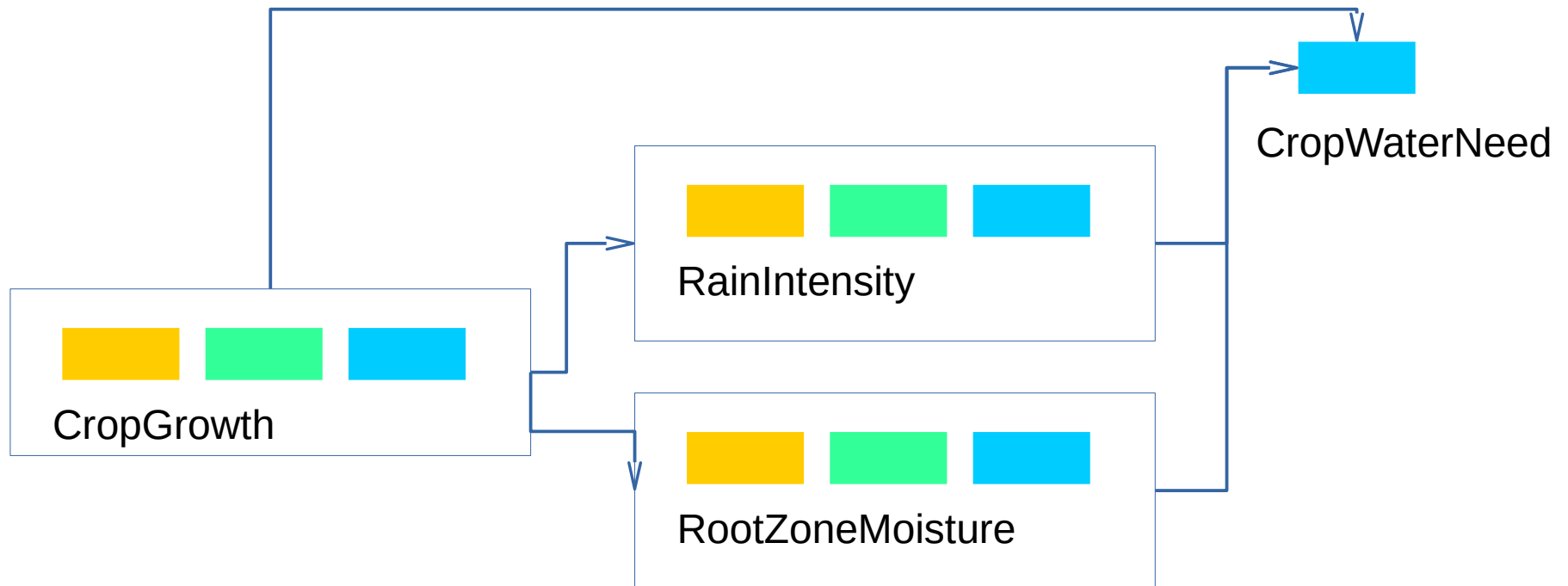




Validation: Une seule règle pour déduire l'état de la pluie

```
irrig:RainIntensityDeduction(?deduction_rain_intensity) ^  
prov:used(?deduction_rain_intensity, ?result_observation_rain_total_quantity) ^  
om:hasNumericalValue(?result_observation_rain_total_quantity, ?  
value_result_observation) ^  
caso:hasState(irrig:observableProperty_rain_intensity, ?rain_intensity_state) ^  
caso:hasOpenUpperBoundary(?rain_intensity_state, ?boundary_upper) ^  
caso:boundaryValue(?boundary_upper, ?value_boundary_upper) ^  
caso:hasClosedLowerBoundary(?rain_intensity_state, ?boundary_lower) ^  
caso:boundaryValue(?boundary_lower, ?value_boundary_lower) ^  
swrlb:lessThan(?value_result_observation, ?value_boundary_upper) ^  
swrlb:greaterThanOrEqual(?value_result_observation, ?value_boundary_lower)  
->  
caso:hasResultState(?deduction_rain_intensity, ?rain_intensity_state)
```

Les chaînes de traitements





> Annotation sémantique de bulletins d'alertes agricoles

Projet D2KAB



D2KAB



Le projet ANR D2KAB 2019-2023 « Data to Knowledge in Agronomy and Biodiversity » (www.d2kab.org) a pour objectif :

- d'améliorer la visibilité et l'échange de jeux de données et de référentiels par le biais de l'Agroportal
- → rendre ces données FAIR par le biais des technologies web sémantique : ontologie, SKOS, graphe de connaissances
- Exploiter ces données FAIR dans des applications dédiées

5 scénarios : agronomie, agriculture, écosystème, biodiversité

Copain a la charge des scénaris :

T4.2 : publication de jeux de données et de référentiels dans le domaine de l'agriculture

T4.3 : navigateur augmenté pour les bulletins de santé du végétal

T4.3 Bulletins de Santé du Végétal (BSV)



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

Colza
A RETENIR CETTE SEMAINE

- 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 siliques 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 : 66%
- F1 premières fleurs ouvertes : 20%
- F2 allongement de la hampe florale, nombreuses fleurs ouvertes : 4%

Méligèthes

Prévisions météorologiques du mercredi 06 au mardi 12 avril :

Temps très ensoleillé et sec pendant toute cette période. Les températures sont en hausse avec des minimales ne descendant pas en dessous de 9°C et des maximales atteignant 20 à 22°C (source météo).

Météo

Risque méligèthes

Risque faible, Risque élevé

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

un état sanitaire des cultures

stades de développement

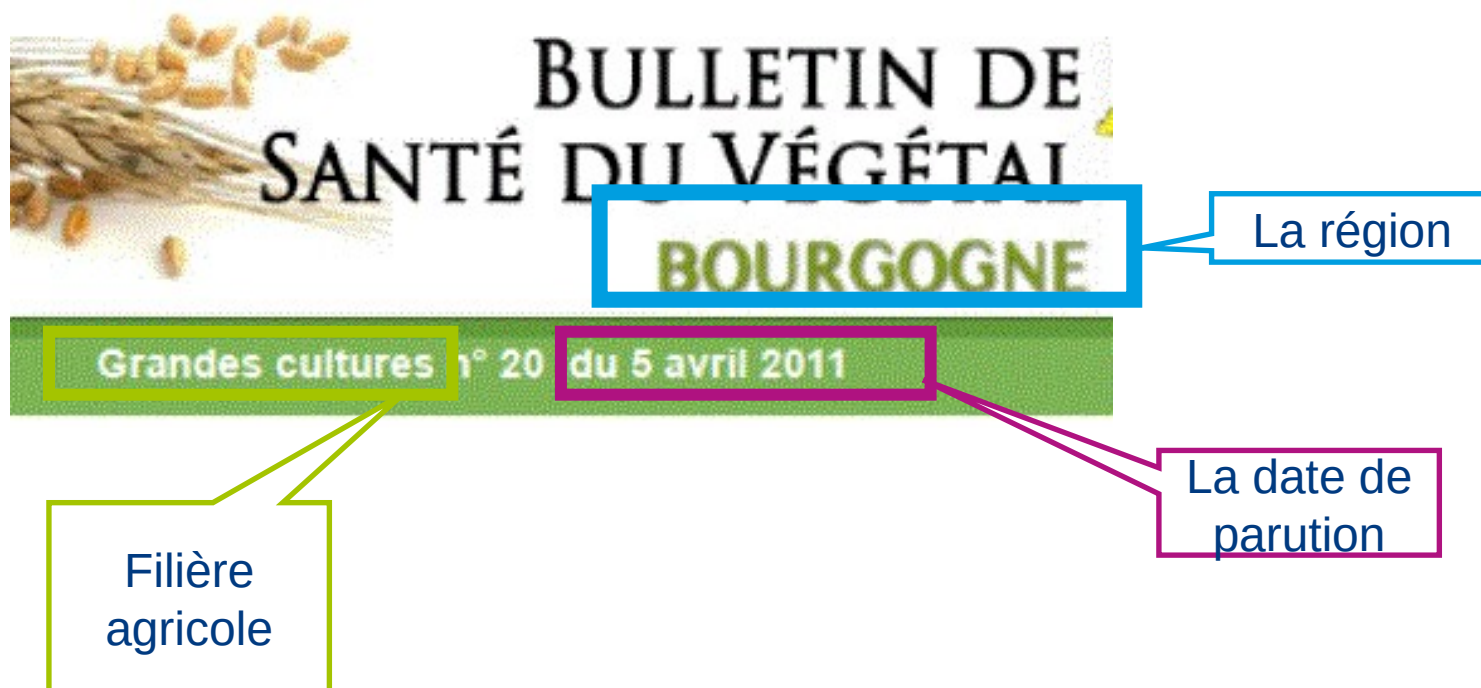
observations des ravageurs et maladies

présence de symptômes

une évaluation du risque phytosanitaire, en fonction des périodes de sensibilité des cultures et des seuils de nuisibilité des ravageurs et maladies

des messages réglementaires

Indexation des BSV



Spécification des besoins

Les agronomes souhaitent interroger une archive d'observations des cultures

la culture

Stade des colzas

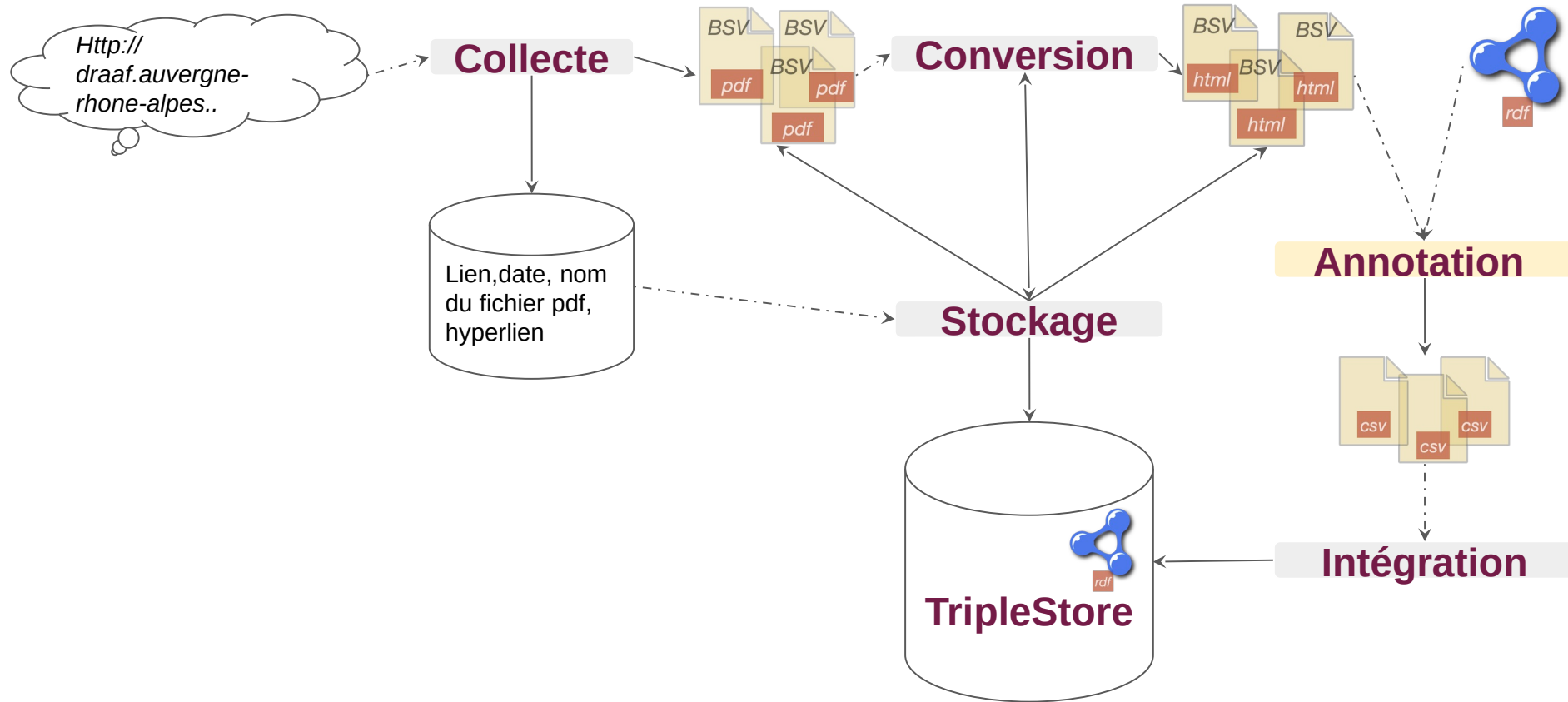
Le stade de développement de la culture

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 : 66%
- F1 premières fleurs ouvertes : 20%
- F2 allongement de la hampe florale, nombreuses fleurs ouvertes : 4%

Schéma global





pdf

htm
|

Pluies du 25 au 28 Juillet :
Le passage pluvio-orageux du week-end dernier a concerné l'ensemble des vignobles de Bourgogne Franche Comté. Les cumuls sont très variables :

5 à 73 mm en 21
17 à 38 mm en 71
7 à 21 mm en 89
61 à 91 mm en 39

STADES VEGETATIFS

Dans les situations les plus précoces, les premières baies vérées sont observées plus régulièrement mais la plupart des situations sont encore au stade grappe fermée. Les pluies de ce week-end, là où elles ont été significatives, vont permettre d'enclencher la véraison.

ACCIDENTS CLIMATIQUES

Les fortes températures de la semaine dernière combinées à des hygrométries basses ont provoqué des dégâts d'échaudage et parfois de phytotoxicité dans de nombreuses parcelles quel que soit le vignoble. Les pertes oscillent entre 5 et 50% selon les vignes.

MILDIU

- Situation au vignoble :

221 parcelles ont été observées sur le réseau BSV.

Aucune évolution n'a été notée depuis la semaine dernière.

- Etat des contaminations en cours :

Avec les pluies des derniers jours qui ont concerné quasiment tout le vignoble de Bourgogne-Franche Comté, de nouvelles contaminations se sont produites. Les symptômes correspondant s'exprimeront entre le milieu et la fin de cette semaine.

Analyse de risque mildiou :

A début véraison, les grappes atteignent le stade de moindre sensibilité. Le jeune feuillage reste quant à lui toujours très sensible.

Compte tenu de l'état sanitaire et des stades phénologiques, le risque est faible pour l'ensemble du vignoble Bourguignon et modéré pour le Jura. Toutefois, en fonction des sorties de taches et des conditions météorologiques à venir, il pourrait évoluer.

INRAE

Mars 2022 / TSCF

Viticulture n°16 du 30 juillet 2019

STADES VEGETATIFS

Dans les situations les plus précoces, les premières baies vérées sont observées plus régulièrement mais la plupart des situations sont encore au stade grappe fermée. Les pluies de ce week-end, là où elles ont été significatives, vont permettre d'enclencher la véraison.

ACCIDENTS CLIMATIQUES

Les fortes températures de la semaine dernière combinées à des hygrométries basses ont provoqué des dégâts d'échaudage et parfois de phytotoxicité dans de nombreuses parcelles quel que soit le vignoble. Les pertes oscillent entre 5 et 50% selon les vignes.

MILDIU

- Situation au vignoble :

221 parcelles ont été observées sur le réseau BSV.

Aucune évolution n'a été notée depuis la semaine dernière.

- Etat des contaminations en cours :

```
<h1>Viticulture n°16 du 30 juillet 2019</h1>
<h2>STADES VEGETATIFS</h2>
<p>Dans les situations les plus précoces, les premières baies vérées sont observées plus régulièrement mais la plupart des situations sont encore au stade grappe fermée. Les pluies de ce week-end, là où elles ont été significatives, vont permettre d'enclencher la véraison.</p>
<h2>ACCIDENTS CLIMATIQUES</h2>
<p>Les fortes températures de la semaine dernière combinées à des hygrométries basses ont provoqué des dégâts d'échaudage et parfois de phytotoxicité dans de nombreuses parcelles quel que soit le vignoble. Les pertes oscillent entre 5 et 50% selon les vignes.</p>
<h2>MILDIU</h2>
<h3>- Situation au vignoble :</h3>
<p>221 parcelles ont été observées sur le réseau BSV.</p>
<p>Aucune évolution n'a été notée depuis la semaine dernière.</p>
<h3>- Etat des contaminations en cours :</h3>
<p>Avec les pluies des derniers jours qui ont concerné quasiment tout le vignoble de Bourgogne-Franche Comté, de nouvelles contaminations se sont produites. Les symptômes correspondant s'exprimeront entre le milieu et la fin de cette semaine.</p>
```



LICENCE OUVERTE
OPEN LICENCE



Spécification des besoins

- 57 Competency Questions ont été définies
- Exemples <http://ontology.irstea.fr/pmwiki.php/Site/BSVSparqlSamples>

RC2: Combien de bulletins contiennent les
corpus ?

RC18: Combien d'annotations de concepts du
thésaurus FCU ont été produites pour le document
D ?

Thésaurus usage des plantes cultivées en France FCU

Issu du projet VESPA, construit à la main à l'aide de plusieurs sources : le larousse agricole, wikipédia français, le registre parcellaire, les statistique de l'agreste. Point de départ les céréales.

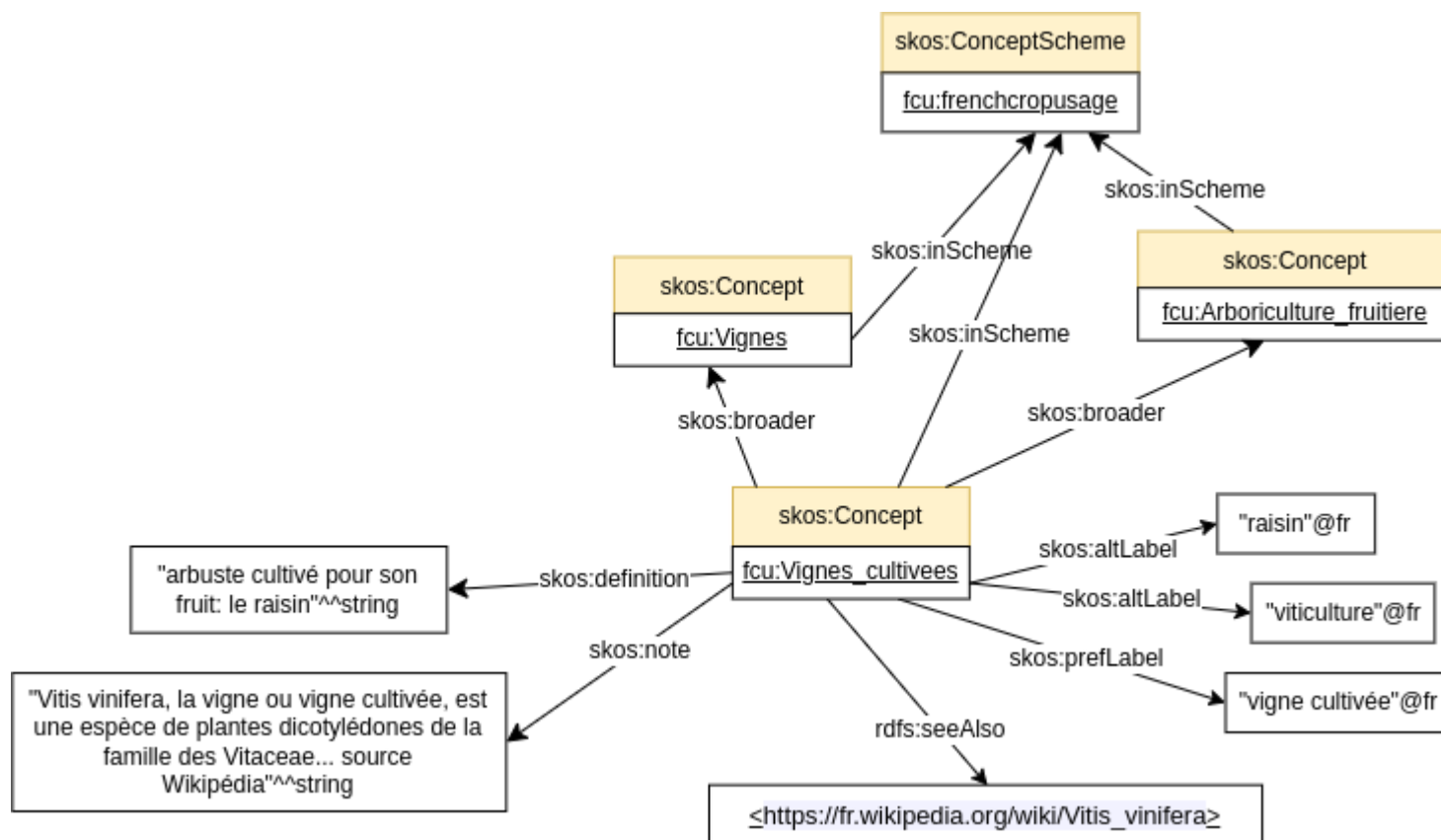
Licence ouverte, un DOI fournit par data.inrae, Disponible sur agroportal et sur le gitlab irstea <https://gitlab.irstea.fr/copain/frenchcropusage> et sur ontology.irstea.fr

Mise à jour pendant toute la durée du projet D2KAB : Version 2.1 et 2.2, la version 3 bientôt disponible

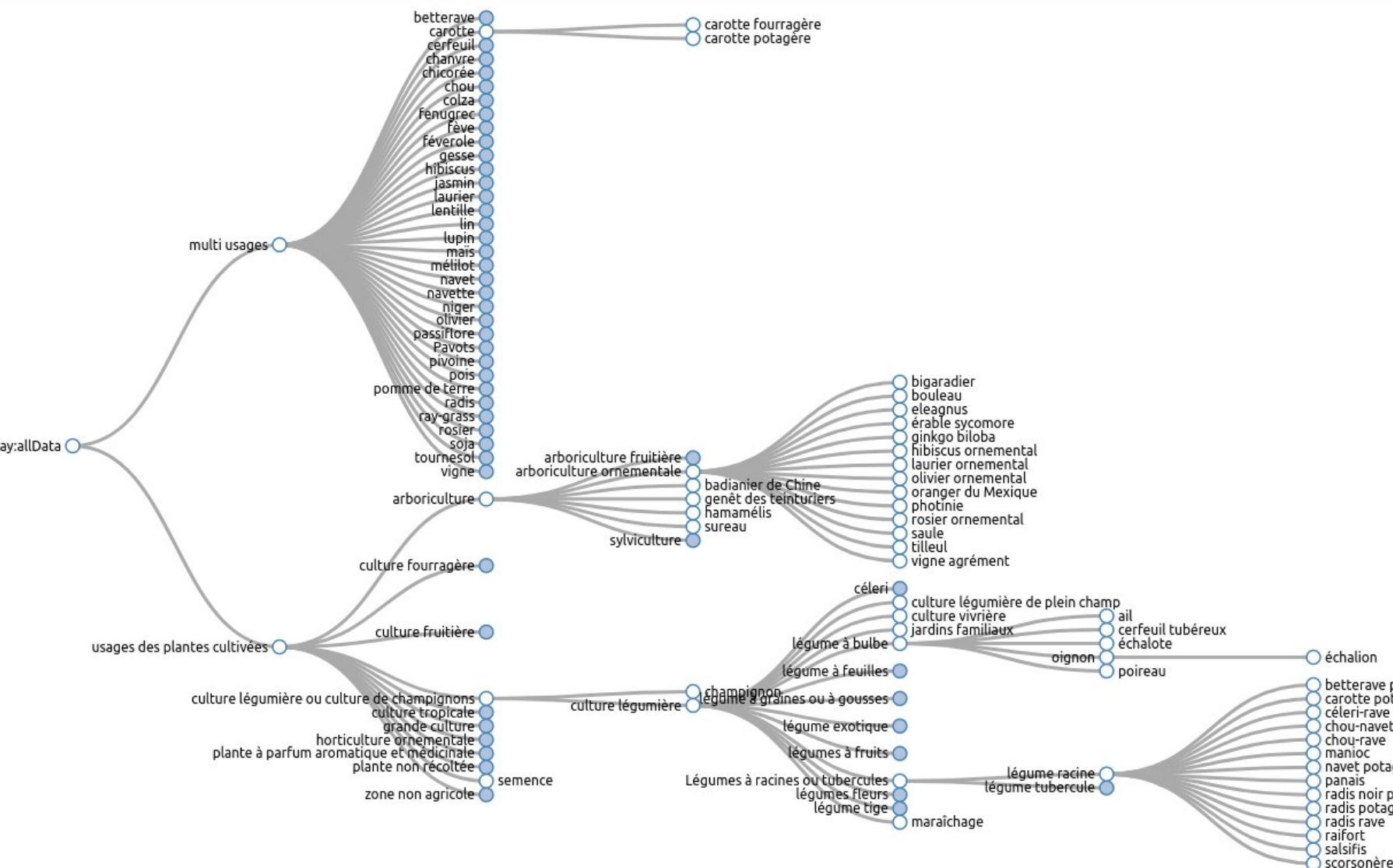
- Maraîchage ; travail en commun ave Elzeard à l'aide de plusieurs nouvelles sources Bonduelle, FranceAgriMer, Encyclopédia Universalis, AM Nagelseim, la ferme du Bec Hellouin
- Fourrage : travail avec un expert à partir des fiches du GNIS
- Arboriculture / sylviculture : extractions de Anna Chepaikina, expertise acta
- Plante de service : travail en cours avec Elzeard, Acta et INRAE.
- Réutiliser dans la base GECCO, la publication sur le web de la base Ephy fait par D2KAB, la serre des savoirs d'Elzeard.



Thésaurus French Crop Usage FCU



Thésaurus French Crop Usage FCU





Spécification des besoins

DÉFINITION DES MODÈLES

- Un corpus est constitué de plusieurs documents
- Un document peut être présenté sous différent format (pdf, html)
- Un bulletin agricole est un document
- Un bulletin agricole est associé à une région
- Un bulletin agricole a une date de publication
- Un bulletin agricole a pour thème une ou plusieurs cultures
- Une ressource sémantique (comme un thésaurus) définit un ensemble d'entité nommée (exemple les cultures)
- Une annotation indique qu'une mention d'une entité nommée a été identifiée dans le texte d'un document
- Une annotation peut être produite par un humain ou par un logiciel
- Une annotation joue un rôle dans une application dédiée (validation correction, identification, complément d'information)
- Un document est annoté par divers processus d'annotation



Sélection des ontologies

DÉFINITION DES MODELES

- Time W3C (Hobbs & Pan, 2006)
- GeoSparql, OGC (Battle & Kolas, 2012)
- Prov Ontology (prov) W3C (Lebo et al., 2013)
- Web Annotation Data Model (oa) W3C (Sanderson et al., 2013)
- Simple Knowledge Organisation System (skos) W3C

Web Annotation Ontology

prefix oa: <<http://www.w3.org/ns/oa#>>

Vocabulaire pour décrire le modèle de données d'annotation du web : il s'agit d'offrir un cadre universel pour annoter des ressources du web (page web ou image, article de blog, ...).

C'est aussi une recommandation du w3c.

Documentation sur <https://www.w3.org/ns/oa>

The image shows a screenshot of a French agricultural bulletin titled "Bulletin de santé du végétal GRANDES CULTURES Hauts-de-France". The bulletin is issue N° 20, dated 2 juillet 2019. It contains information about the health of crops in the Hauts-de-France region. Handwritten blue ink annotations are present throughout the document, including arrows pointing to specific sections and words, and circles around certain terms. The annotations include: "Publication" pointing to the date, "Localisation" pointing to the region, "GC: Edition" pointing to the title, "Cultures" pointing to the crop section, "BETTERAVES" pointing to the beet section, "agresseur/évolution du risque" pointing to the beet section, "Météo" pointing to the weather section, "Etat des cultures" pointing to the crop status section, "agresseur" pointing to the pest section, and "risque" pointing to the risk section. The bulletin text includes: "BETTERAVES : Présence de parasitisme en diminution. Début des observations des maladies du feuillage : situation saine.", "MAÏS : Vol de pyrale en cours.", "Betteraves : 71 parcelles observées", "Ravageurs : La présence d'asticots de pégomyes est en diminution suite aux fortes chaleurs de la semaine dernière. Les asticots se sont desséchés. Elle est signalée dans 20% des parcelles observées. Aucune parcelle n'atteint le seuil de nuisibilité de 50% de betteraves avec présence de galeries et de larves.", "Seuil de nuisibilité Pégomyes : A la couverture du sol : 50% de plantes avec galeries et présence d'asticots dans certaines galeries.", and "Dans 35 % des parcelles, des œufs de pégomyes ont été observés. A la suite des températures élevées, les avortements régulent les populations." There is also a small image of a green leaf at the bottom right.

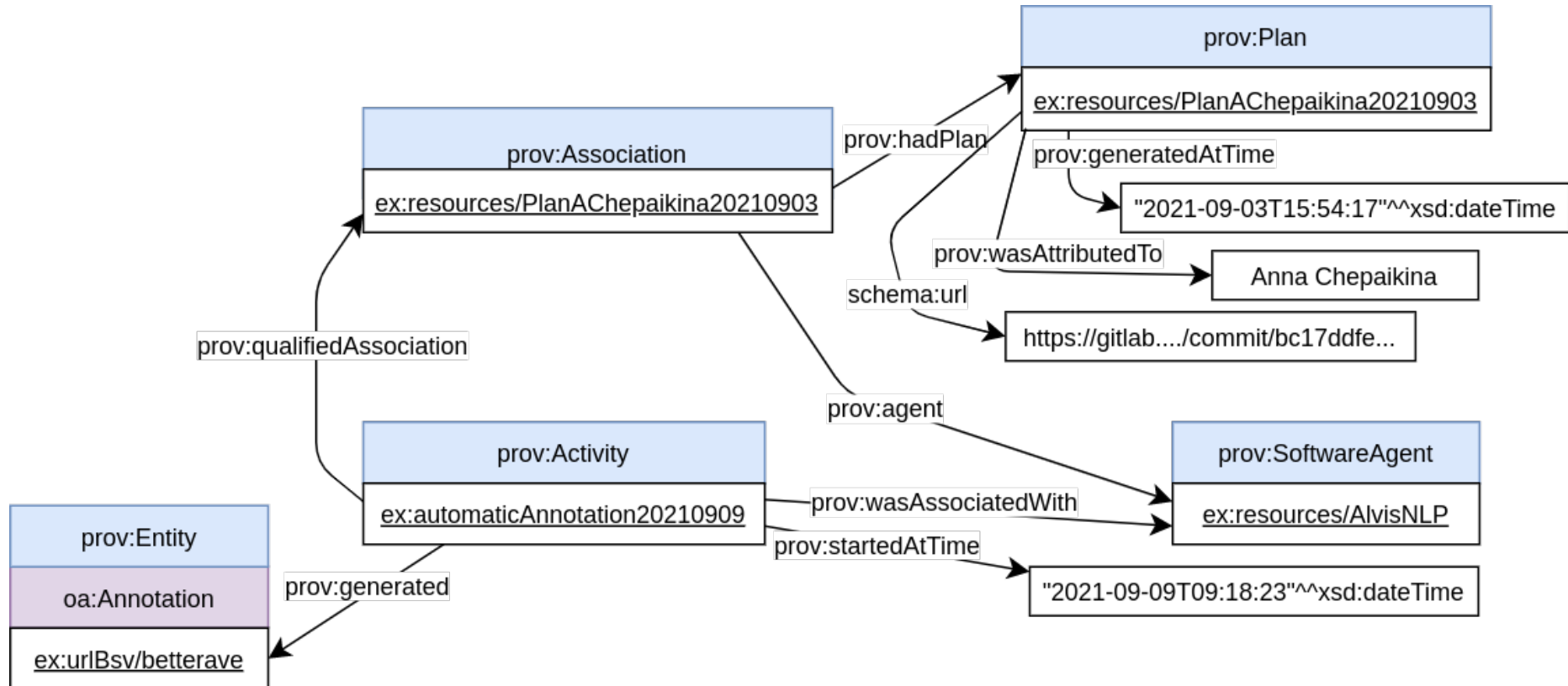


INRAE

Mars 2022 / TSCF

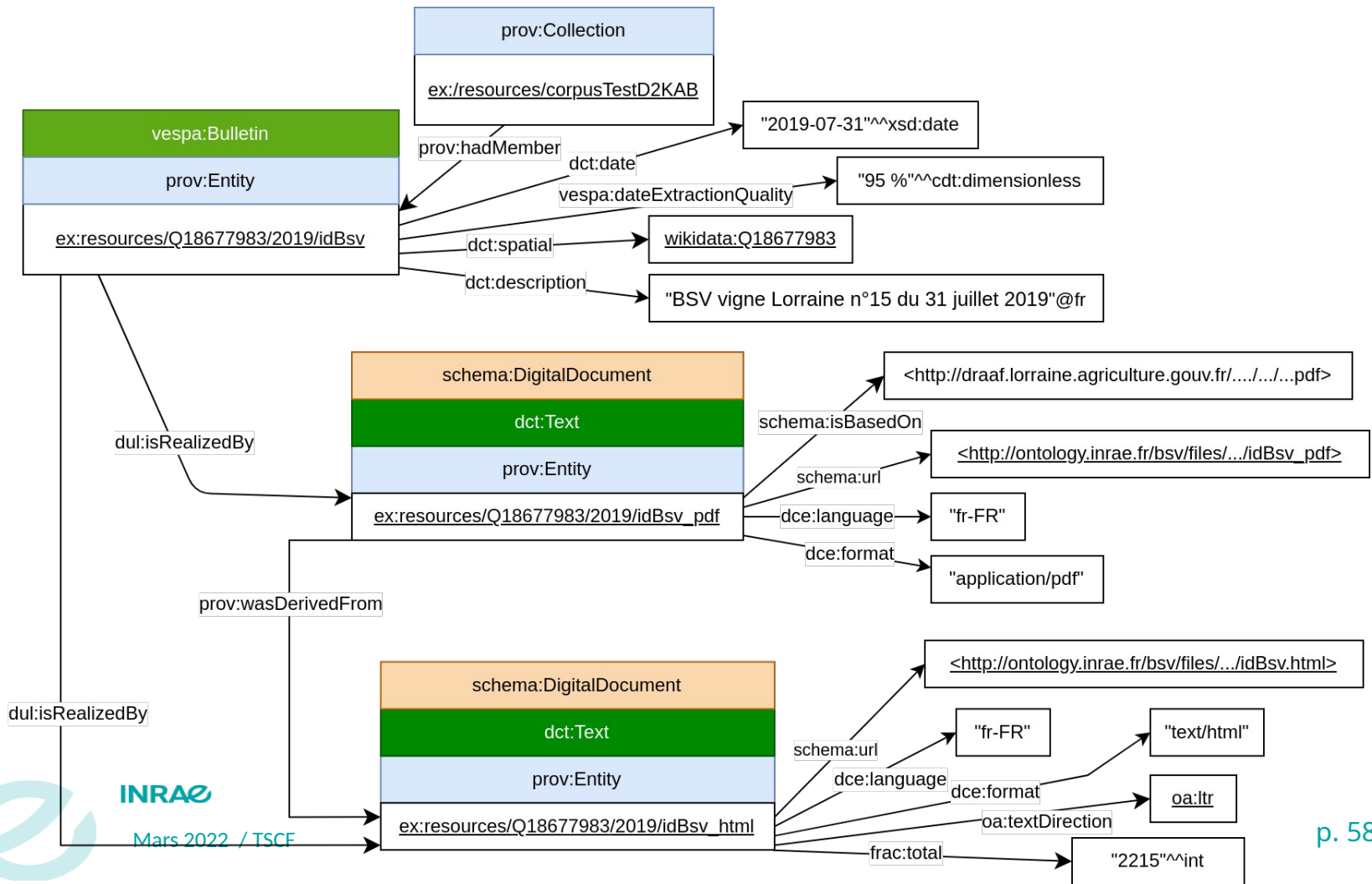
Structure des données

Activité d'annotation



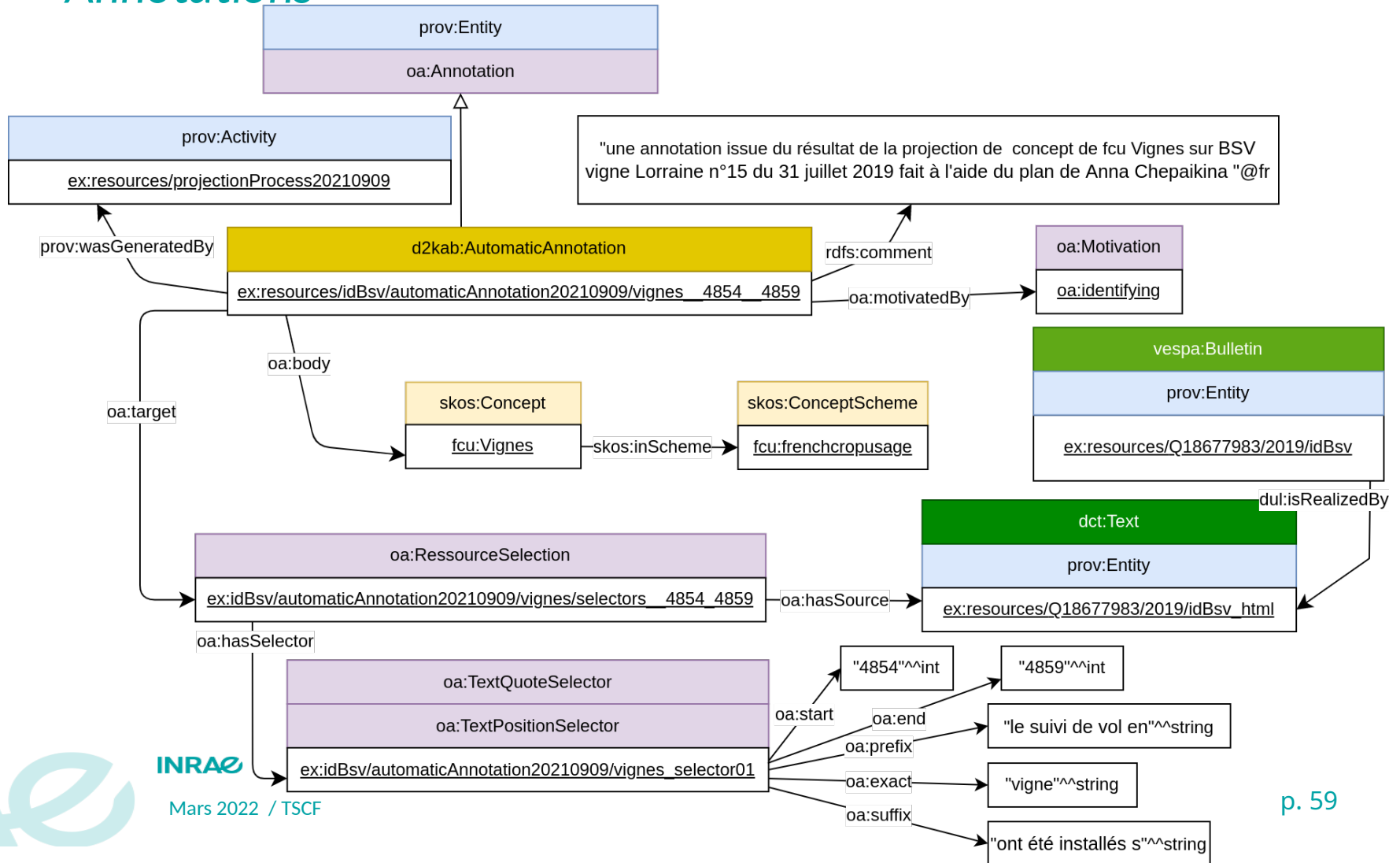
Structure des données

Représentation des BSV



Structure des données

Annotations





Validation: example

RC2: Combien de bulletins contiennent les corpus ?

```
SELECT (COUNT(?bsv) AS ?nb) ?nom ?corpus
WHERE {
  ?corpus a prov:Collection.
  ?corpus rdfs:label ?nom.
  ?corpus prov:hadMember ?bsv.
}
GROUP BY ?nom ?corpus
```

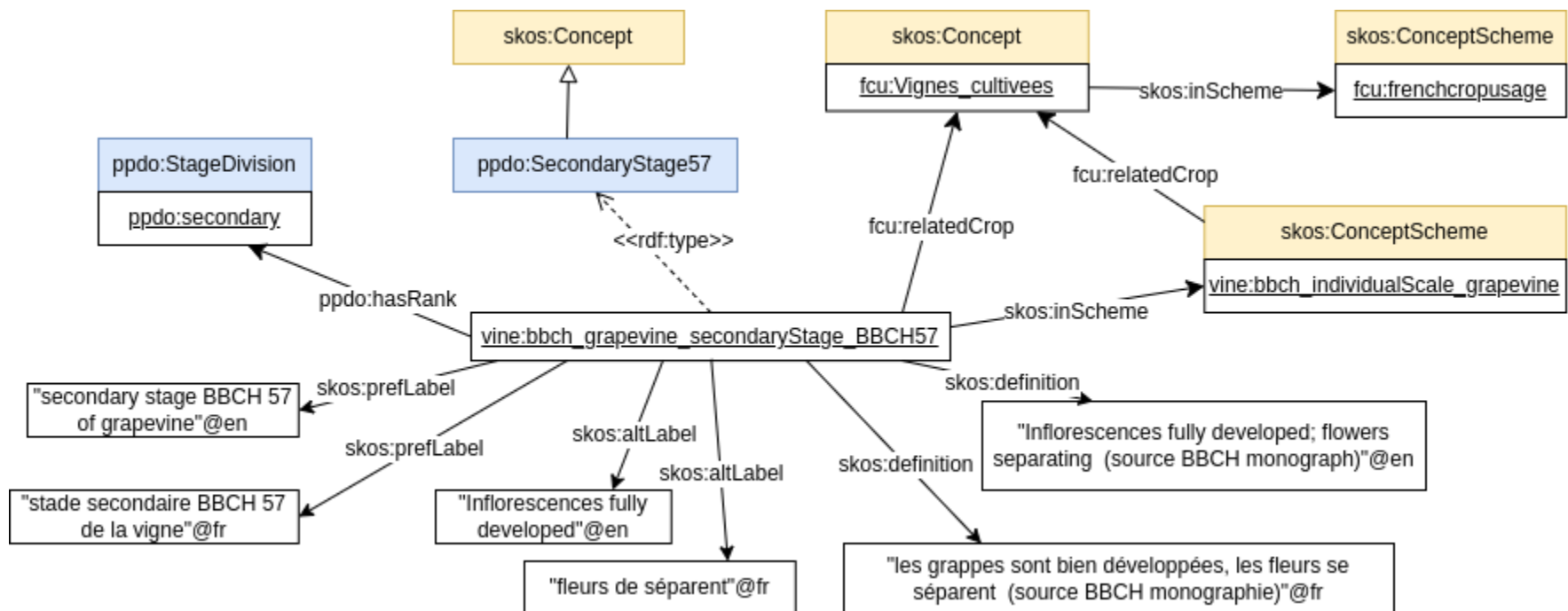
Projet ANR D2KAB

Annotation de BD + Annotation de texte

Interrogation multisources: BD d'épidémie versus bulletins d'alerte

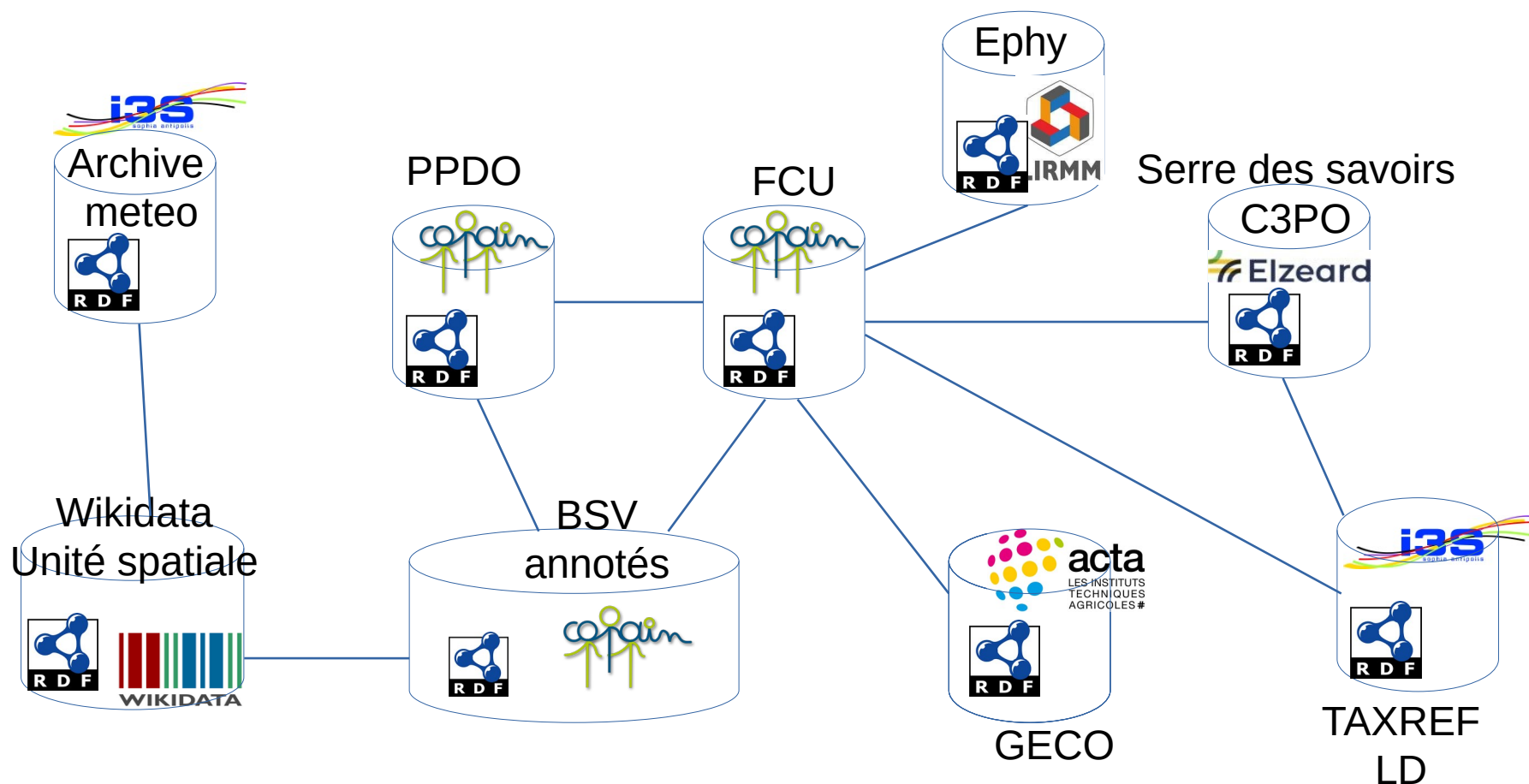
Mieux intégrer les modèles SKOS dans les réseaux d'ontologies

SKOS pour publier des données de référence comme les stades phénologiques



Graphes distribués

Interroger des graphes de connaissances distribués sur le web ET liés les uns aux autres



Conclusion

- Les ontologies du web sont des schémas de données FAIR qui contiennent des patrons de conception (des bonnes pratiques de modélisation)
- Les ontologies ont un cout: elles sont difficiles à comprendre et à réutiliser correctement
 - Quel informaticien écrit la documentation de son logiciel ou de sa base de données?
 - Qui lit les notices d'utilisation des appareils ou des médicaments?
- Les technologies du Web Sémantique changent les paradigmes de modelisation: base de graphes, Property First Class Citizen, Open World Assumption, ...



Travaux Pratiques

Découverte et interrogation du thésaurus FrenchCropUsage à distance

- Aller sur la page <http://ontology.irstea.fr/pmwiki.php/Site/FrenchCropUsage>
- Trouver le SPARQL endpoint <http://ontology.inrae.fr/frenchcropusage/sparql>
- Executer les requêtes SPARQL de la section “exemple de requêtes”

Découverte et interrogation conjointe des annotations des BSV et de FCU à distance

- Aller sur la page <http://ontology.irstea.fr/pmwiki.php/Site/BSV>
- Trouver le SPARQL endpoint <http://ontology.irstea.fr/bsv/sparql>
- Executer les requêtes SPARQL de la page <http://ontology.irstea.fr/pmwiki.php/Site/BSVSparqlSamples>

Installation d'un SPARQL endpoint en local et interrogation

