

Integration et consolidation de données, par le biais des technologies du Web **Sémantique**



Catherine ROUSSEY , Ghislain ATEMEZING



INRAe



A l'issue de cette séquence, vous aurez

- Découvert les principes et les difficultés de la gestion de données
- Les données ?
- Découvert pourquoi les informaticiens construisent des schémas de données documentés et pourquoi ils ont besoins de vous
- Echanger des données: un début de bonnes pratiques
- La science ouverte et les données FAIR
- Visualiser des exemples de ressources sémantiques
- Trouver des liens utiles
- Les ontologies du Web et les données ouvertes et liées
- Partager des données sur le Web



Définitions

L'intégration de données : consiste à unifier différents sources / systèmes d'information afin qu'elles fonctionnent comme une seule source / système.

La consolidation de données : permet de combiner des données similaires provenant de multiples sources, ce qui implique leur tri, leur nettoyage (Data Cleaning) afin d'améliorer leur cohérence

Interopérabilité des systèmes d'information

Capacité de plusieurs systèmes informatiques indépendants de :

- Communiquer : échanger des données
- Utiliser les données échangées

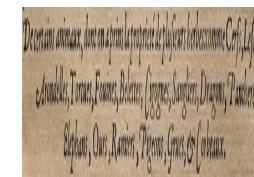


Interopérabilité technique :

Aptitude à établir un canal de communication pour échanger un message

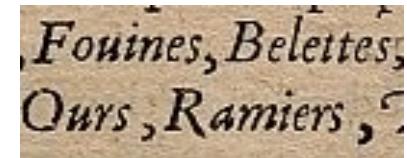
Interopérabilité syntaxique :

Aptitude à décoder un message



Interopérabilité sémantique :

Aptitude à interpréter le contenu du message



Définitions

Données, Informations

“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)

Données: un élément d'information,

percevable par l'humain,



manipulable par la machine



Informations: donnée + sens

Sens : il faut fournir à l'humain du contexte



Type → élément d'un schéma de données



Orange?

fruit

ville

couleur

Définitions

Données, Informations et Connaissances

“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)

Données: un élément d'information

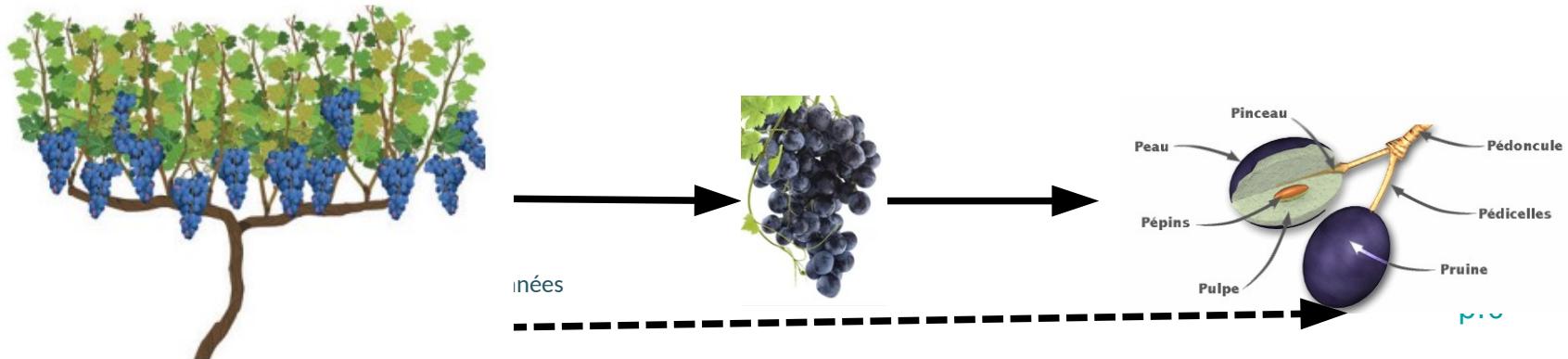
Informations: donnée + sens

Connaissances: information formalisée dans un langage de représentation des connaissances

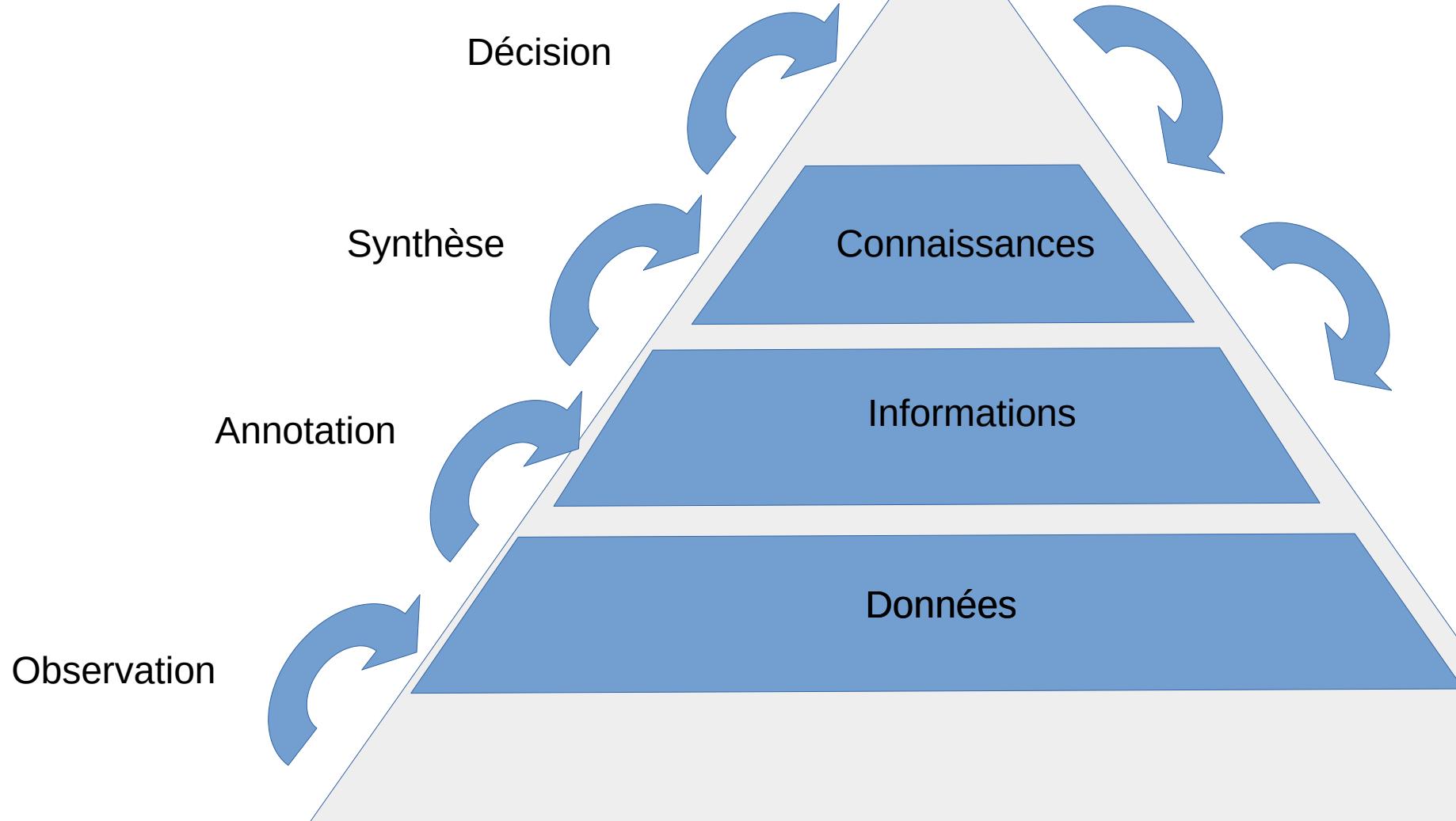


Pour inférer de nouvelles connaissances

“est composé de “ est une relation transitive



Continuum DIK



Définitions

Données, Métadonnées

Données "Les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de recherche. " (OCDE, Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics. Paris, 2007)

MetaDonnées : Au-delà des données, il y a les métadonnées qui permettent de décrire plus précisément les données. Ce sont des données sur les données. Il est recommandé de renseigner les métadonnées au fur et à mesure de l'avancée du projet.



auteur

titre

éditeur



What is MetaData for experimentation?

Metadata is data reporting

Who create the data ?

When were the data created ?

Where is it geographically ?

How were the data developed ?

- Sensor, human observation, ...

Why were the data developed ?

- Which experiment, ...

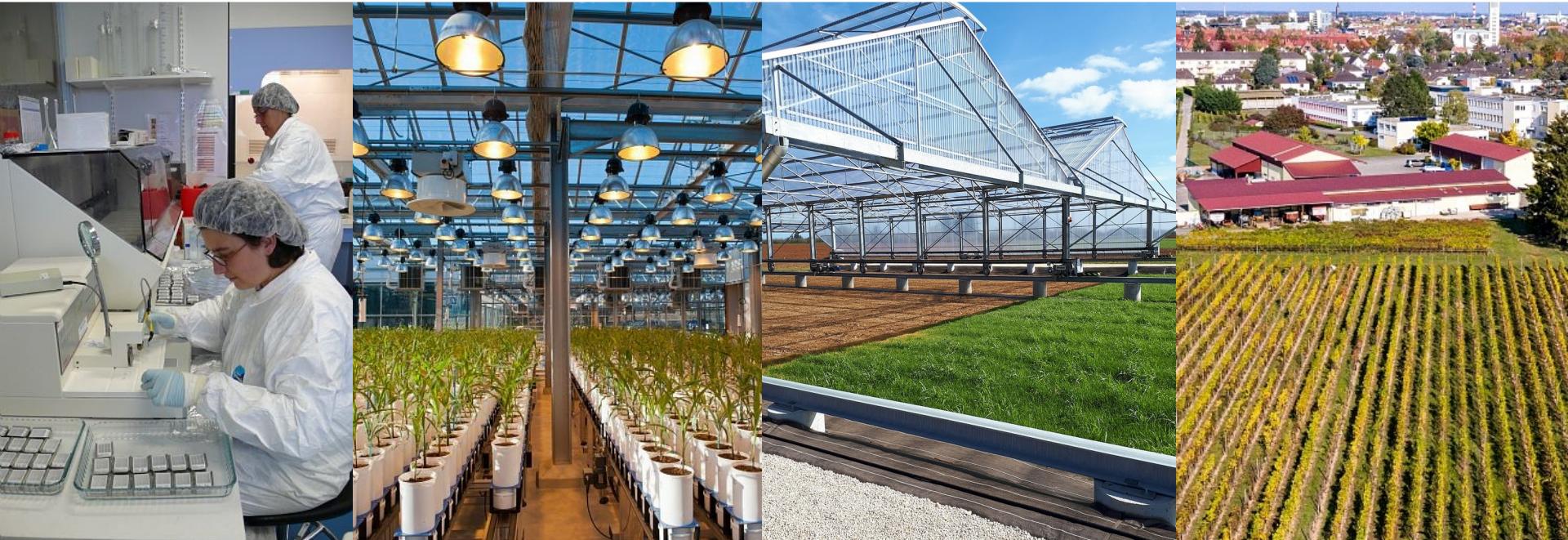
What is the content of the data ?

- Unit, type of measure,



Source de données: Les plateformes d'expérimentations

laboratoires d'analyse, serres, unités expérimentales, exploitations agricoles



Données complexes: différentes échelles

genes, cellules, organismes, écosystèmes



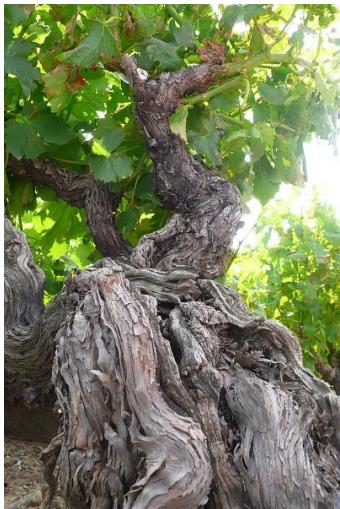
Données complexes: différents organismes

bactéries, insectes, champignons, plantes, mammifères



Données complexes: différents organes

Le bois, le bouton floral, la feuille, la grappe



Données complexes: différents stades d'évolution

Débourrement, gonflement du bourgeon, pointe verte, stade plombs de chasse, ...



Données complexes: différentes transformations

La vigne, les raisins, le moult, le vin



Météo à Montoldre - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

Neo4j - The World's Leading ... Neo4j - The World's Leading ... Specifications/OpenSearch/1... Météo à Montoldre meteo.clermont.cemagref.fr Latest Headlines

Météo actuelle à Montoldre, Allier(03)

Temperatur	16.5°C
Humidité	93%
Point de rosé	15.4°C
Vent	E at 0.0 km/hr
Pression atmosphérique	760.0 mm & Falling Slowly
Pluie du jour	2.5 mm
Pluie moyenne	0.0 mm/hr
Orage Total	22.9 mm
Pluie du mois	35.1 mm
Pluie de l'année	864.1 mm
Température ressentie	16.5°C

Température extérieure

Outside Temp (°C)

Humidité

Outside Hum (%)

Index UV

UV Index

Date	Time	Temp	Out	Wind	Wind	Rain	Solar
			Hum	Speed	Dir	Bar	
08/01/19	23:30	3.3	89.0	6.4	WNW	1025.1	0.25 1.5 0
08/01/19	23:40	3.3	89.0	6.4	WNW	1025.0	0.25 1.5 0
08/01/19	23:50	3.2	89.0	6.4	WNW	1025.0	0.25 1.5 0

10/10/14 15:54

Température ressentie

Wind Chill (°C)

Direction du vent

Wind Direction

Vitesse du vent

Wind Speed

GSN Server :: GSN

HOME DATA MAP FULLMAP

Auto-refresh every : 1min refresh close all

wan4 03/08/2009 11:10:00 +0200

Real-Time Addressing Structure Description Download

record 7920.0
relative_humidity 74.44
air_temperature 7.326
wind_speed_northerly_component 8.13
wind_speed_westerly_component -10.28
wind_speed_vertical_component -10.04
snow_height -198.3

wannengrat_unterhalb_felsen 03/06/2010 02:06:00 +0200

Real-Time Addressing Structure Description Download

wind_direction null
wind_speed null
solar_rad null
snow_water_content_1 null
snow_water_content_2 null
snow_temp_1 null
snow_temp_2 null
air_humid null
air_temp null
air_temp_ir null
snow_surface_temp_ir null

Crédit JP Calbimonte

Measurement Location: wan4, Wannengrat

Deployment/Fieldsite Information [hide]

- Deployment: Wannengrat (link: Home)
- Measurement Location: wan4
- Coordinates: 46.803716666° N, 9.778030555° E
- Altitude: 2417 m
- Swiss Northings
- Swiss Eastings
- Experiments: Measuring and Modelling the Snow Distribution on Wannengrat (about)
- Edit information Click here
- Tasks
 - Register an observation
 - Register a sensor deployment
 - Bulk Load information

Location

Map Satellite Hybrid Terrain Overlays

Image

Model	S/N	DB name	Link
200-	2100	wind_speed_northerly_component	4da160f
27005		wind_speed_westerly_component	
UVW		wind_speed_vertical_component	
SR50	C5635	snow_height	A948534
Hygroclip	45614	air_temperature	E57d1e4
S3	323	relative_humidity	

Jeux de données d'expérimentation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Region	MP	Orgnanisme	Commune	Lat	Lon	Varietes	Année	Date	APP	Taitement	BBCH	H	L	PS_BR	OPTIDOSE	MA
2	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	MZ	56	NA	NA	4	NA	Metirame de Zir
3	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	S	56	NA	NA	4	NA	Soufre
4	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	HP	56	NA	NA	4	NA	Hydrogenocarbonate de
5	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	BP	56	NA	NA	4	NA	Bicarbonate de pota
6	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	DP	56	NA	NA	4	NA	Disodium Phospho
7	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	PP	56	NA	NA	4	NA	Phosphonate de pot
8	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	HEO	56	NA	NA	4	NA	Huile essentielle d'oran
9	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	Curatio	56	NA	NA	4	NA	Polysulfure de Cal
0	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	PP ou DP	56	NA	NA	4	NA	Phosphonate de pot
1	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	SPP	56	NA	NA	4	NA	Soufre
2	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	SPP	56	NA	NA	4	NA	Phosphonate de pot
3	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	SBP	56	NA	NA	4	NA	Soufre
4	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	18/05/2022	1	SBP	56	NA	NA	4	NA	Bicarbonate de pota
5	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	MZ	65	NA	NA	4	NA	Metirame de Zir
6	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	S	65	NA	NA	4	NA	Soufre
7	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	HP	65	NA	NA	4	NA	Hydrogenocarbonate de
8	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	BP	65	NA	NA	4	NA	Bicarbonate de pota
9	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	DP	65	NA	NA	4	NA	Disodium Phospho
0	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	PP	65	NA	NA	4	NA	Phosphonate de pot
1	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	HEO	65	NA	NA	4	NA	Huile essentielle d'oran
2	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	Curatio	65	NA	NA	4	NA	Polysulfure de Cal
3	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	PP ou DP	65	NA	NA	4	NA	Phosphonate de pot
4	BFC	HVE	IFV	Liergues	45,9762	4,6525	Gamay 509	2022	25/05/2022	2	SPP	65	NA	NA	4	NA	Soufre

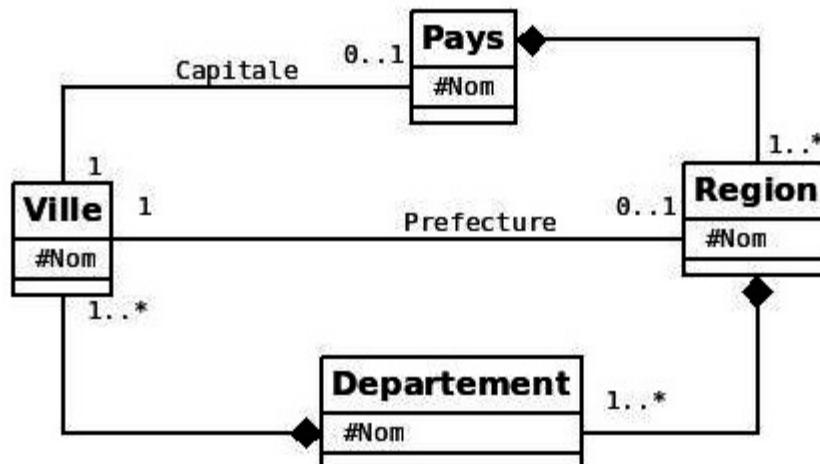
Définitions

Schéma de données

représentation abstraite et schématique d'une organisation des données. Les données sont regroupées par **entités** (des types d'objects ou classes) partageant des caractéristiques communes (les **attributs**). Les **relations** entre entités et leurs contraintes sont aussi représentées.

Suivant le language informatique, et l'étape de modelisation du système de gestion de données, il existe plusieurs types de schémas de données: modèle conceptuel, modèle logique, modèle physique.

Un schéma de données est représenté dans un language graphique (UML, modèle Entité-Association, ...) pour faciliter le partage entre concepteurs, développeurs et utilisateurs.



Schema UML
modèle conceptuel des villes françaises
Credit Stephane Crozat
<https://stph.scenari-community.org/bdd/mod3/co/rel4e004.html>

Identification des entités

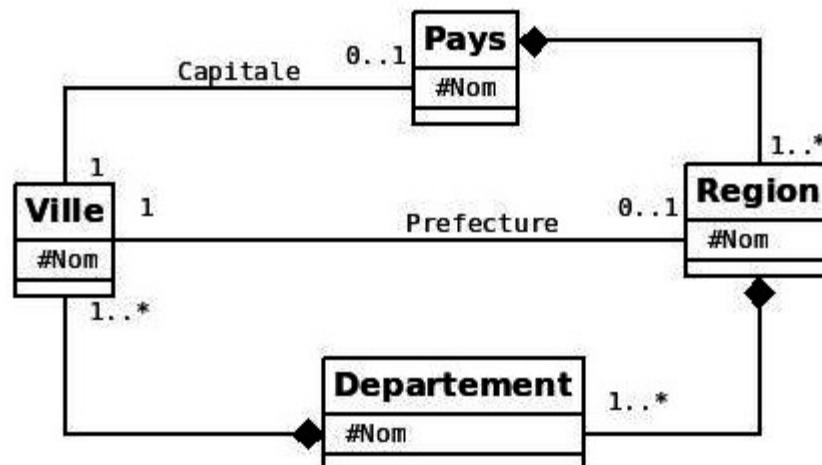
Schéma de données

Département : Nom, Numéro

Commune : Nom, Numero postal

Personne : Nom, Prénom, N° de sécurité sociale

Personnel de recherche : idHAL, Orcid



Schema UML
modèle conceptuel des villes françaises
Credit Stephane Crozat
<https://stph.scenari-community.org/bdd/mod3/co/rel4e004.html>

Entity Identification in the Web

URL Uniform Resource Locator
<https://www.inrae.fr/>



URI Uniform Resource Identifier

Sur wikidata
<https://www.wikidata.org/wiki/Q70571774>
<wikidata:Q70571774>

Sur Research Organization Registry (ROR)
<https://ror.org/003vg9w96>
<rор:003vg9w96>

Semantic Resources:

Thésaurus AGROVOC



The screenshot shows a search interface for the Thésaurus AGROVOC. At the top, there is a navigation bar with 'Content language' set to 'English', a search input field containing 'buds', and a 'Search' button. Below the search bar, it says '1 results for "buds"'.

buds (en)

- ↳ plant reproductive organs (en)
- ↳ adventitious buds (en), flower buds (en), leaf buds (en)
- ∅ eyes (buds)
- 🌐 بَرَاعِم (ar), 芽 (zh), 芽眼 (zh), pupeny (cs), očka (pupeny) (cs), bourgeon (fr), Knospe (de), Auge (botanisch) (de), কলিকাঁ (hi), আঁক (কলিকা) (hi), rügy (hu), szem (rügy) (hu), Gemme (it), Occhi (gemme) (it), 芽 (ja), 幼芽 (ja), 눈 (ko), ຕາຕັນໄມ້ (lo), Kudup (ms), knopper (nb), حوانه‌ها (fa), جشنگ (fa), Pąk (pl), Pąk (oczko) (pl), gomo (planta) (pt), muguri (ro), ochiuri (muguri) (ro), почки растений (ru), глазки (почки) (ru), пупольци (sr), púčiky (sk), Yema (planta) (es), vipukizi (sw), ໃສ້ງສູງທຳນົວ ມາລະນະ (te), ກຈູກ (ກອຽກາຍ) (te), ຕາຕັນໄມ້ (th), ດາ (ດາຕັນໄມ້) (th), tomurcuk (tr), göz (tomurcuk) (tr), бруньки (uk), очки (брюньки) (uk)

http://aims.fao.org/aos/agrovoc/c_1135

Entity Identifier

Semantic Resources:

Thésaurus INRAE

Thésaurus INRAE

Langue des données français ▾

Chercher

A-Z Hiérarchie Collections Nouveautés

- architecture de la plante
- barrière épithéliale
- barrière mucosale
- caractéristique anatomique
- complexe vagal dorsal
- déjection animale
- densité osseuse (déprécié)
- élément d'anatomie animale
- élément d'anatomie végétale
 - actinorhize
 - apex (botanique)
 - ascocarpe
 - ascospore
 - aubier
 - bille de pied
 - bois de réaction
 - **bourgeon**
 - bourgeon apical
 - bourgeon axillaire
 - bourgeon floral
 - bourgeon végétatif
 - branchaison
 - branche
 - canopée
 - carpelle
 - cep
 - tige de réaction

05. SCIENCES BIOLOGIQUES > 05.04 ANATOMIE ET FLUIDES BIOLOGIQUES > élément d'anatomie végétale > bourgeon

TERME PRÉFÉRENTIEL

bourgeon 

CONCEPT(S) GÉNÉRIQUE(S)

élément d'anatomie végétale

CONCEPT(S) SPÉCIFIQUE(S)

bourgeon apical
bourgeon axillaire
bourgeon floral
bourgeon végétatif

APPARTIENT AU(X) GROUPE(S)

coll ALIGNES A AGROVOC
coll BOIS ET FORêt
coll CONCEPTS ALIGNES

MICROTHESAURUS

BIO anatomie et fluides biologiques
Thésaurus INRAE

TRADUCTION(S)

bud anglais
plant bud

URI

http://opendata.inrae.fr/thesaurusINRAE/c_5987 

TÉLÉCHARGER CE CONCEPT :

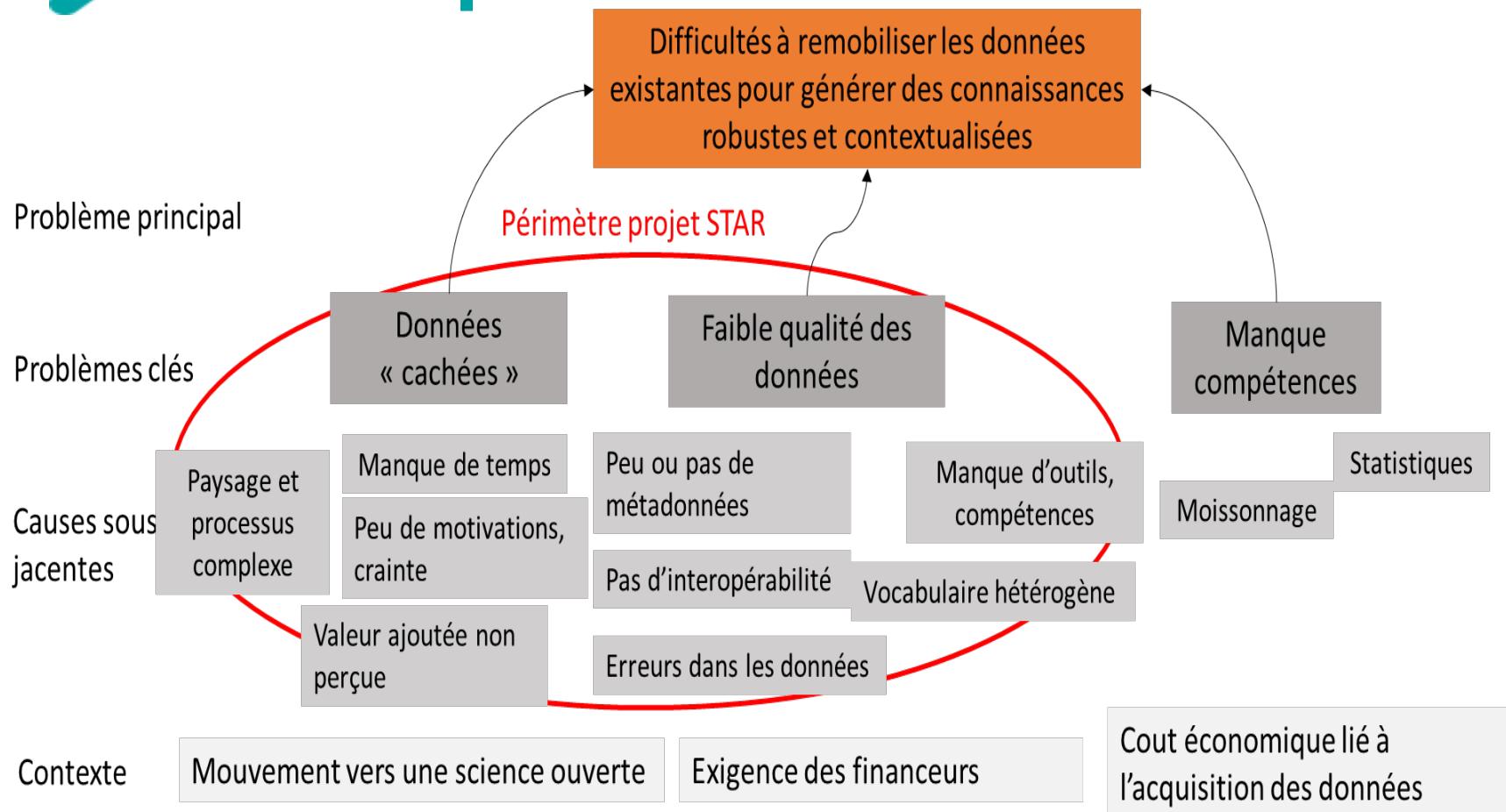
RDF/XML TURTLE JSON-LD

Dernière modification le 01/07/2025

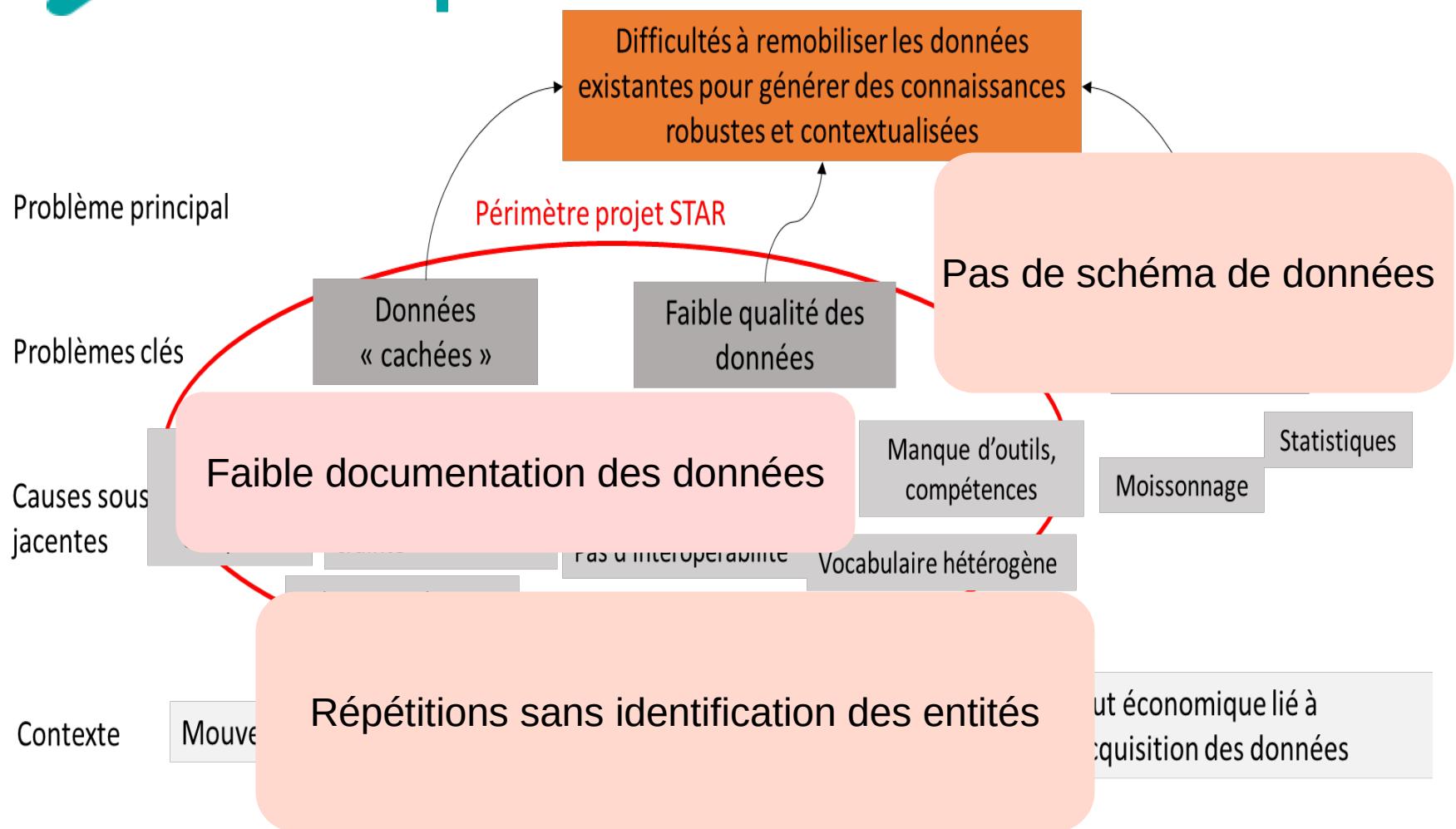


Entity Identifier

> Arbre à problème



Arbre à problème



Données FAIR

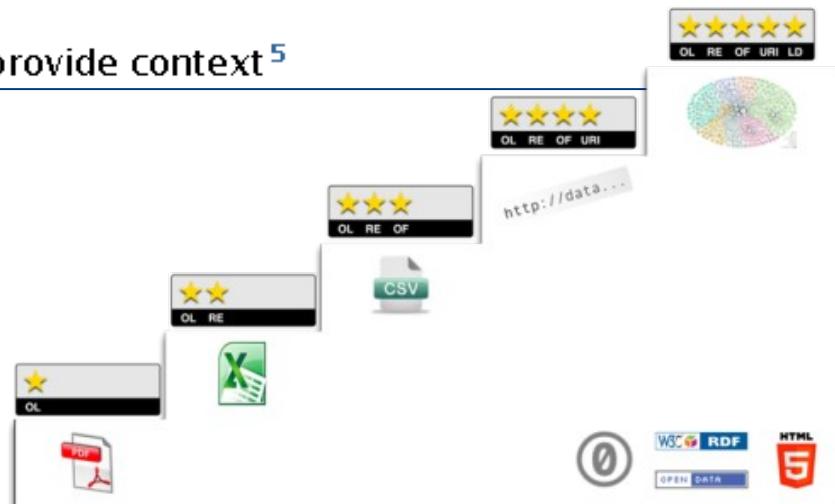


Credit FAIR in a nutshell. Image: ARDC 2018 - CC-BY 4.0.
<https://librarycarpentry.org/Top-10-FAIR/2018/12/01/aus-government-data/>

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



Dataset Example



Human Readable with editor tool like
word e.g.



0BRPechRouge2014.txt

Beausoleil
Site Pech Rouge
Position R4-P5
Variety Carignan
The plot is supervised by Jean
Year 2014
Date : 2025/06/06
Development stage observed 19

2013_Inrae_0BR.txt

Verson Sud
Pech-Rouge
LOCATION = [5,6]
CULTIVAR Carignan
jean.bernard@ifv.fr
year : 2013
stg : BBCH 56
date 07 07 2025

Dataset Example



Machine accessible

Key/value a syntax for structured format

Tool : spreadsheet, JSON, XML



0BRPechRouge2014.txt

Plot : Beausoleil
Site : Pech Rouge
Position : R4-P5
Variety : Carignan
Person: Jean Bernard
Year : 2014
Date : 2025/06/06
Stage : EL 19

2013_Inrae_0BR.txt

PLOT name = Verson Sud
FARM name = Pech-Rouge
LOCATION rank = [5,6]
CULTIVAR name = Carignan
MANAGER family name = BERNARD
MANAGER first name = Jean
PLANTATION year = 2013
OBSERVATION stg = BBCH 56
OBSERVATION date = 07 07 2025

Dataset Example



Machine accessible

Key/value a syntax for structured format

Schema : entities and attributes



0BRPechRouge2014.txt

Site

Name :Pech Rouge

Farm Plot

Name: Beausoleil
Variety : Carignan
Year : 2014

Person

Name : Jean Bernard

Cep

Position : R4-P5

Observation

Date : 2025/06/06
Stage : EL 19

2013_Inrae_0BR.txt

FARM

name = Pech-Rouge

PLOT

name = Verson Sud
cultivar name = Carignan
planted year = 2013

PLANT

location = [5,6]

MANAGER

family name = BERNARD
first name = Jean

OBSERVATION

stg = BBCH 56
date = 07 07 2025

Dataset Example



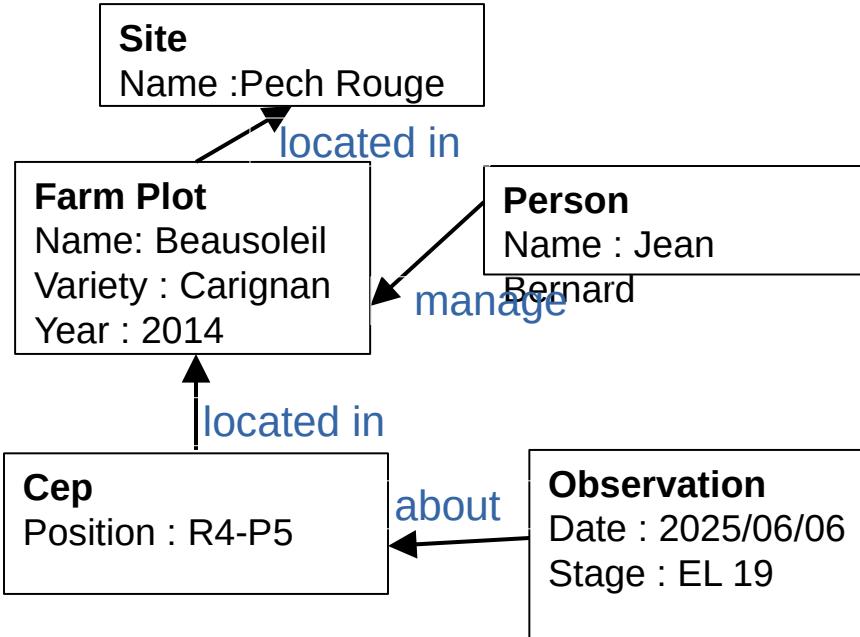
Machine accessible

Key/value a syntax for structured format

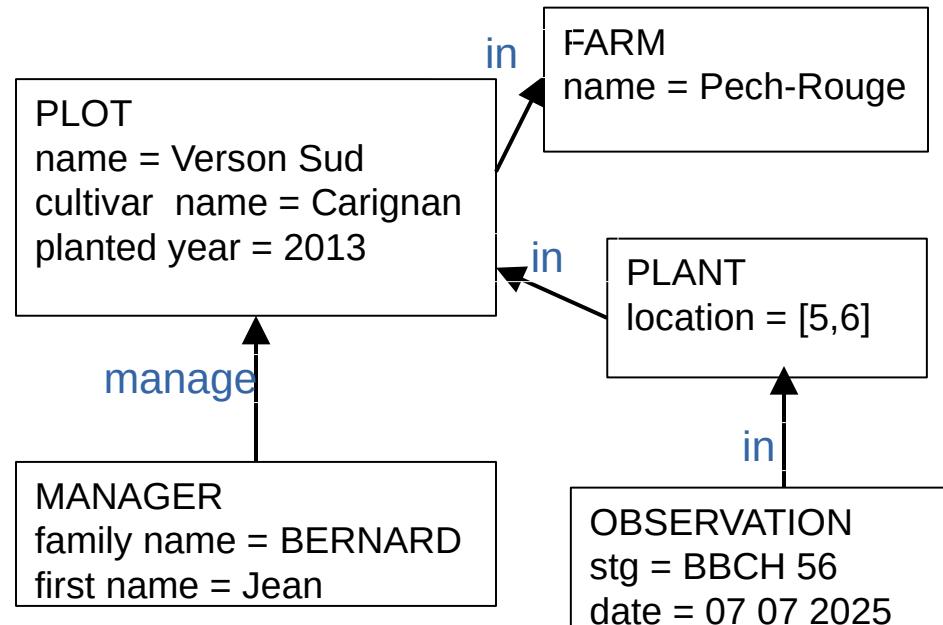
Schema : entities and attributes and relations



0BRPechRouge2014.txt



2013_Inrae_0BR.txt



INRAE

Intégration et consolidation de données

Bi Sem Web 2025 / C Roussey

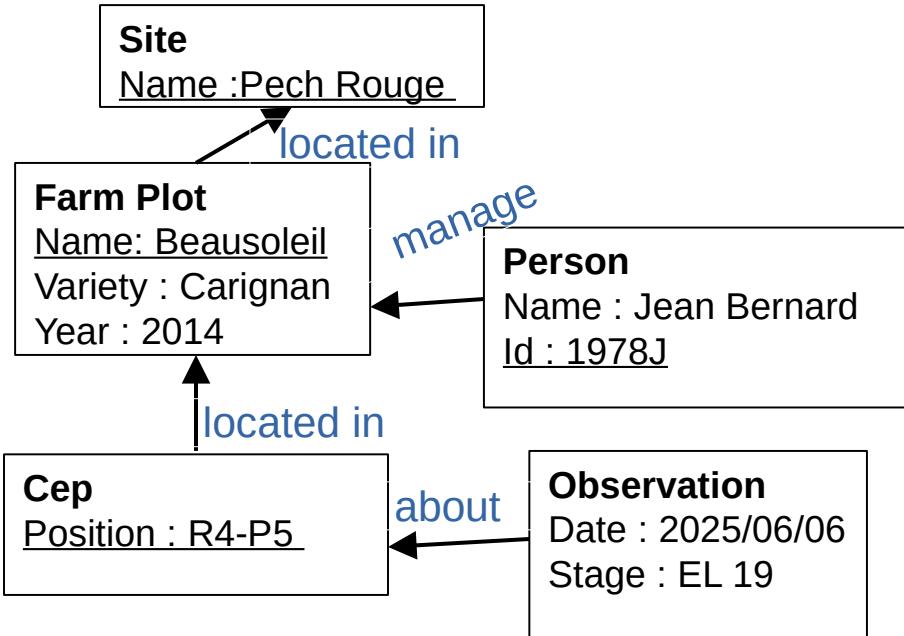
Dataset Example



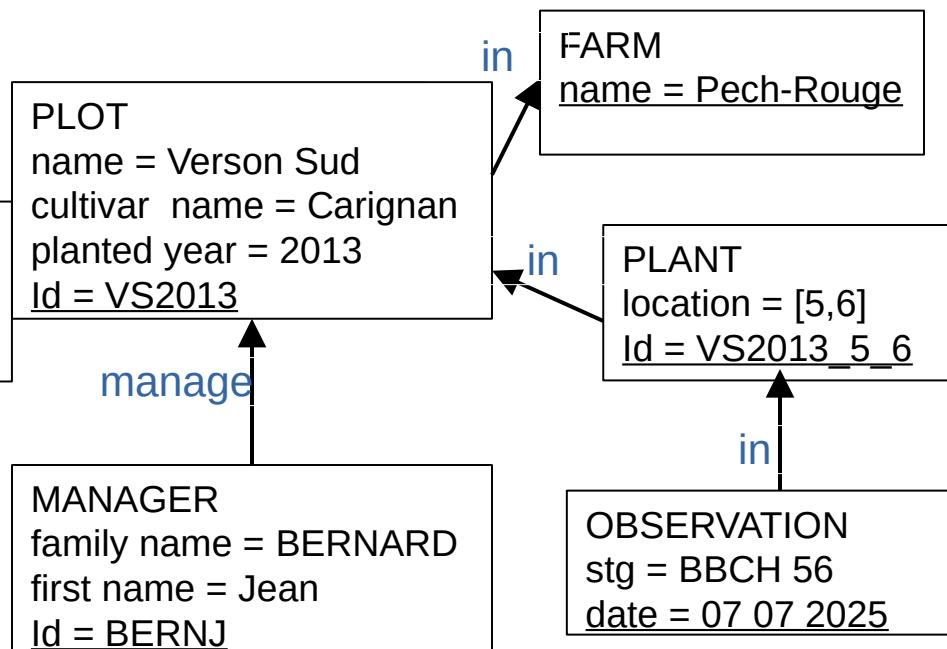
Machine accessible
Key/value a syntax for structured format
Schema : entities and attributes and relations
identification



0BRPechRouge2014.txt



2013_Inrae_0BR.txt



INRAE

Intégration et consolidation de données

Bi Sem Web 2025 / C Roussey

Dataset Example



Machine accessible

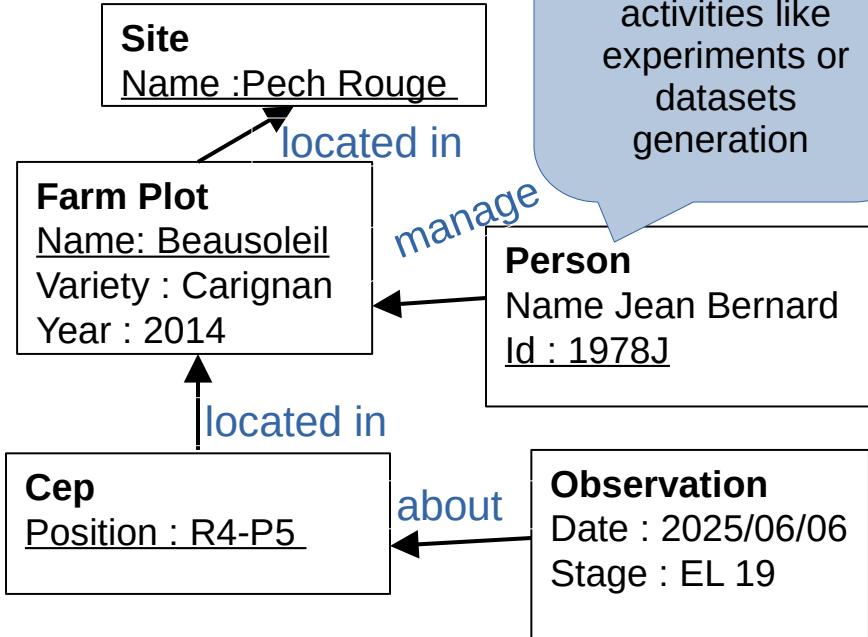
Key/value a syntax for structured format

Schema : entities and attributes and relations



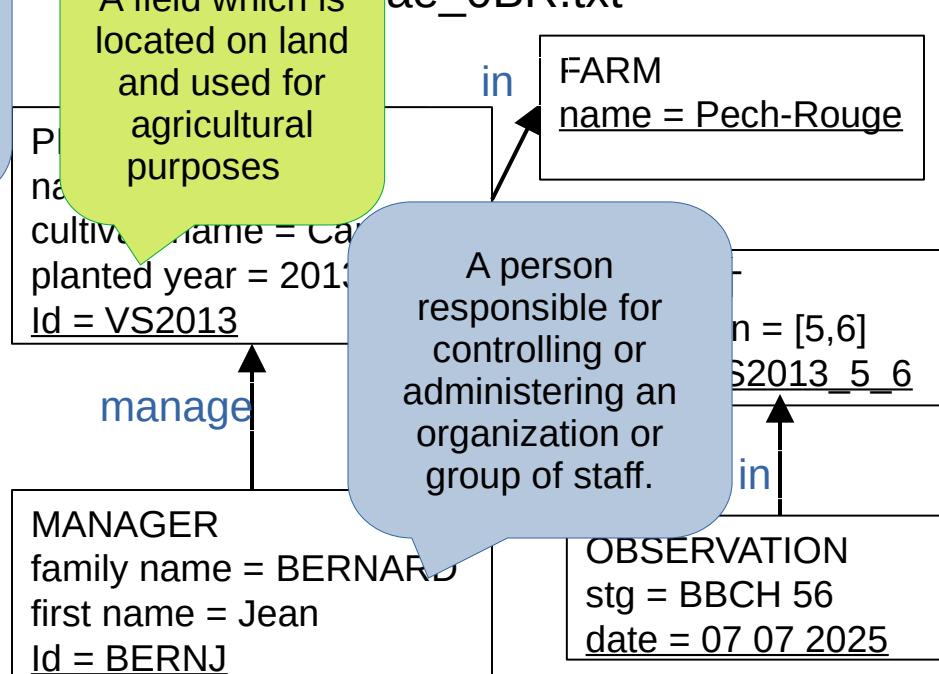
Identification

0BRPechRouge2014.txt



entities → data dictionary

ae_0BR.txt



INRAE

Intégration et consolidation de données

Bi Sem Web 2025 / C Roussey

Dataset Example



Machine accessible

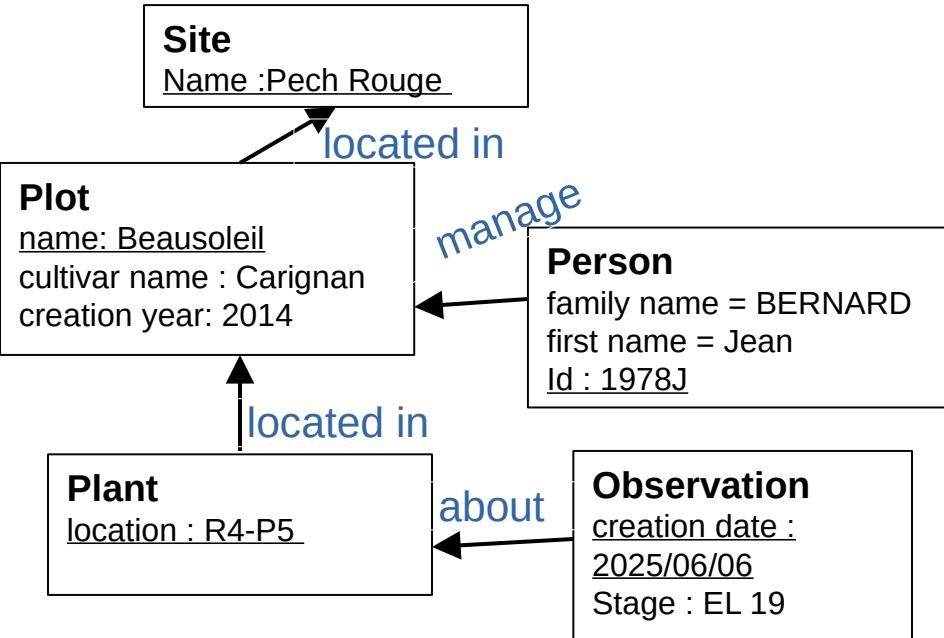
Key/value a syntax for structured format

Schema : entities, attributes, relations, id, def

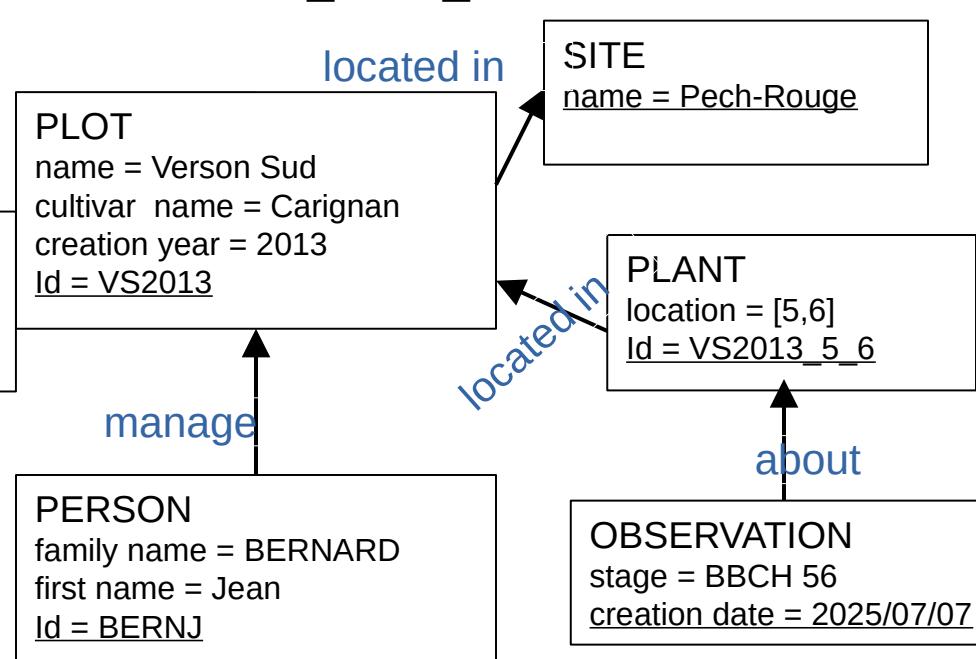
Shared vocabulary



0BRPechRouge2014.txt



2013_Inrae_0BR.txt



Dataset Example



Machine accessible

Key/value a syntax for structured format

Schema : entities, attributes, relations, id, def

Shared vocabulary on the Web



0BRPechRouge2014.txt

<bot:Site>

<dc:name> :Pech Rouge

<inra:Plot>

<dc:name>: Beausoleil

<inra:cultivarName> :

Carignan

<dc:creationDate>: 2014

<foaf:Person>

<foaf:familyName> =
BERNARD
<foaf:firstName> = Jean
<dc:Identifier> : 1978J

<inra:Plant>

<inra:location> : R4-P5

<ssn:Observation>

<dc:creationDate> :
2025/06/06
<inra:growthStage> : EL 19

2013_Inrae_0BR.txt

<bot:Site>

<dc:name> = Pech-
Rouge

<inra:Plot>

located in
<dc:name> = Verson Sud
<inra:cultivarName> =
Carignan
<dc:creationDate> = 2013
<dc:Identifier> = VS2013

manage

<foaf:Person>

<foaf:familyName> =
BERNARD
<foaf:firstName> = Jean
<dc:Identifier> = BERNJ

located in

about

<ssn:Observation>

<inra:growthStage> = BBC
56
<dc:creationDate> =
2025/07/07

Ressources sémantiques: Schémas de données partagés sur le Web

Un ensemble d'entités (des types), d'attributs et de relations utilisées pour l'intégration de données focalisé sur un besoin précis issu d'un organisme de normalisation (W3C).

DCterms Dublin Core Metadata Initiative : Dublin Core ou DC

- Une communauté, créée à Dublin au commencement du Web, venant en particulier du monde des bibliothèques (Library), qui a produit un vocabulaire pour cataloguer les pages Web.
- Un schéma composé d'un ensemble d'attributs ou de relations
- documentation: <https://www.dublincore.org/specifications/dublin-core/dc Terms/>

FOAF Friend Of A Friend

- Un petit schéma de données pour décrire les personnes, leurs attributs et leurs relations
- aligné avec des éléments de DC
- documentation: http://xmlns.com/foaf/spec/#term_Person
- réutilisé dans d'autres schémas



Dataset Example



Machine accessible

Key/value a syntax for structured format

Schema : entities, attributes, relations, id, def

Shared vocabulary on the Web



0BRPechRouge2014.txt

<bot:Site>

<dc:name> : Pech Rouge

<inra:Plot>

<dc:name>: Beausoleil

<inra:cultivarName> :

Carignan

<dc:creationDate>: 2014

<foaf:Person>

<foaf:familyName> =

BERNARD

<foaf:firstName> = Jean

<dc:Identifier> :

<orcid:1978J>

2013_Inrae_0BR.txt

<bot:Site>

<dc:name> = Pech-Rouge

<inra:Plant>

<inra:location> = [5,6]

<dc:Identifier> = VS2013

<inra:Plant>

<inra:location> : R4-P5

<ssn:Observation>

<dc:creationDate> :

2025/06/06

<inra:growthStage> :

<ppdo:EL19>

<foaf:Person>

<foaf:familyName> = BERNARD

<foaf:firstName> = Jean

<dc:Identifier> = <orcid:1978J>

<ssn:Observation>

<inra:growthStage> =

<ppdo:BBCH56>

<dc:creationDate> =

2025/07/07

Ressources sémantiques: Thésaurus publié sur le Web

Un vocabulaire organisé partagé sur le Web: Un réseau de termes normalisés pour décrire un sous domaine précis. Les thésaurus suivent tous la même structuration minimale. Il existe d'autre forme de vocabulaire organisé.

Les échelles phénologiques

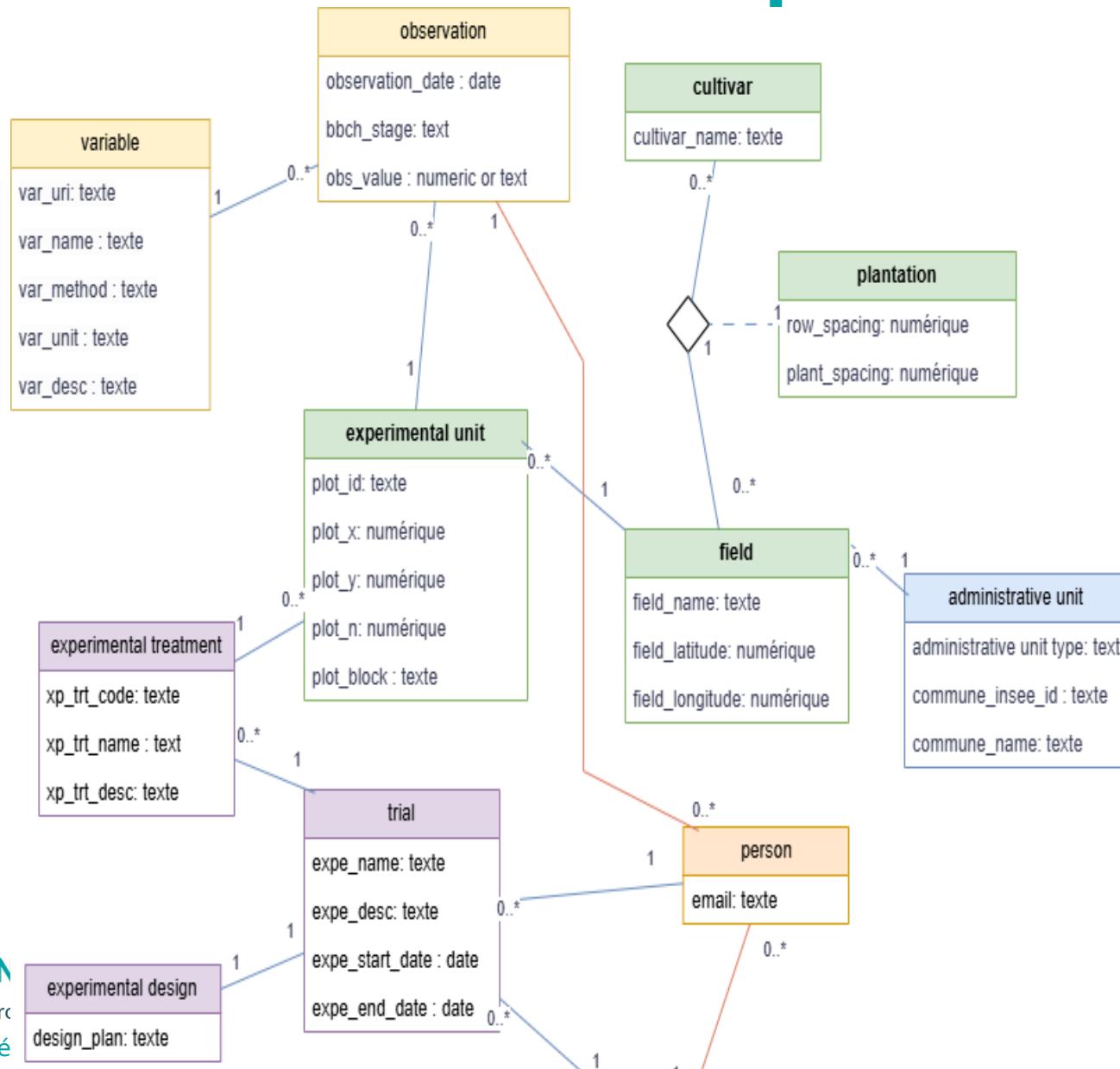
- ontologie BBCH-based Plant Phenological Description Ontology (PPDO)
- URI: <https://opendata.inrae.fr/ppd-def#>
- git : <https://forge.inrae.fr/bsv/phenologicalstages>
- Agroportal: <https://agroportal.lirmm.fr/ontologies/PPDO>

Un graphe contenant les échelles alignées des mêmes cultures

- exemple pour la vigne 5 échelles (BBCH, EL, Baggioini, IFV, Epicure) ont été alignées
- URI <https://opendata.inrae.fr/ppd-res/grapevine#>
- Agroportal: <https://agroportal.lirmm.fr/ontologies/PPD-GV>
- SPARQL endpoint <https://rdf.codex.cati.inrae.fr/ppd-res/grapevine/sparql>



Jeux de données d'expérimentation



Jeux de données d'expérimentation

1 fichier excel

+ 4 feuilles par entités :
Essai, Parcelle,
Modalités, Placette

+ 1 feuille dictionnaire
de données

+ des feuilles de
données d'observation

A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	plot_id	xp_trt_code	plot_block	plot_x	plot_y	plot_n							
2	C2	2 C		1	1	7							
3	TNT3	TNT	C	1	2	7							
4	C5	5 C		1	3	7							
5	C9	9 C		1	4	7							
6	C7	7 C		1	5	6							
7	C10	10 C		1	6	7							
8	C4	4 C		1	7	7							
9	C8	8 C		1	8	7							
10	C6	6 C		1	9	7							
11	C3	3 C		1	10	7							
12	C1	1 C		1	11	7							
13	D2	2 D		1	12	7							
14	TNT4	TNT	D	1	13	5							
15	D6	6 D		1	14	6							
16	D10	10 D		1	15	6							
17	D1	1 D		1	16	7							
18	D8	8 D		1	17	7							
19	D9	9 D		1	18	7							
20	D5	5 D		1	19	7							
21	D3	3 D		1	20	7							
22	D4	4 D		1	21	7							
23	D7	7 D		1	22	7							
24	A1	1 A		2	1	7							
25	A8	8 A		2	2	6							
26	A6	6 A		2	3	7							
27	A3	3 A		2	4	6							
28	TNT1	TNT	A	2	5	5							
29	A2	2 A		2	6	7							
30	A9	9 A		2	7	7							
31	A5	5 A		2	8	6							

1	cuvre 2 semaines après symptômes
2	cuvre à partir de symptômes
3	cuvre
4	Ramélo
5	Limoxide
6	Belvino
7	Equisetum / salix
8	Esseva
9	Mimétic
10	Tanins châtaignier

Typologie des ressources sémantiques par usage

classes, relation de subsomption entre classes, propriétés objets et de type de données, ... données

Structurer

Organiser en classe les objets numériques selon leurs caractéristiques

schéma de données

classes, relation de subsomption entre classes

Indexer / classer

retrouver les objets numériques selon une thématique

taxonomie scientifique

réseau de termes, distance entre termes

Normaliser

S'affranchir de la variabilité de la langue, proposer une recherche à facettes

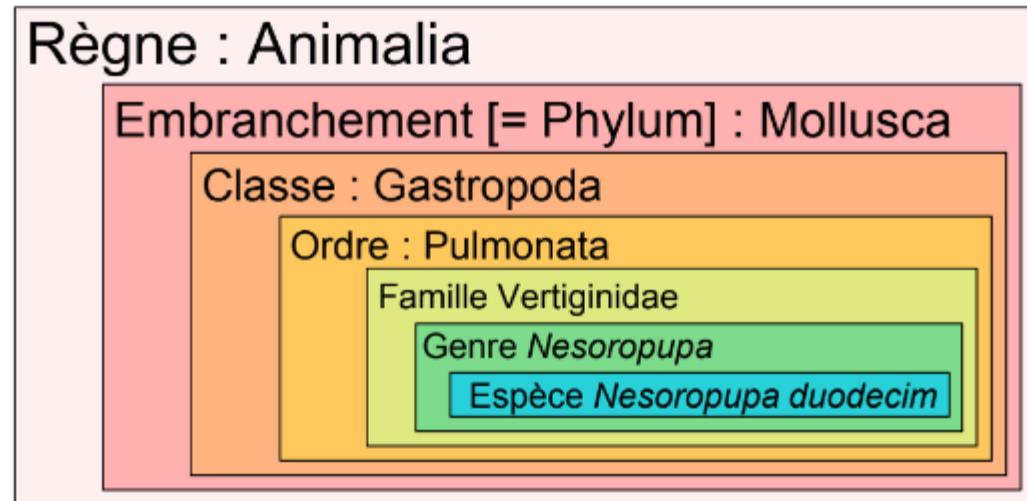
thesaurus

base de données terminologiques

terminologie

Les taxonomies: des hiérarchies de classes

Classification des organismes vivants



Nomenclature: comment nommer les boites → codes de nomenclature

Taxonomie: faire des boites pour regrouper les organismes partageant des points communs

credit Olivier Gargominy MNHN

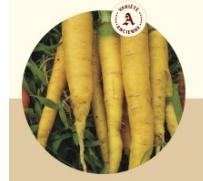
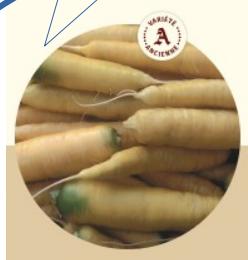
Taxons et noms scientifiques

Synonyme de l'espèce

Daucus Vulgaris Lam. 1779

Nom de référence de l'espèce

Daucus Carota L. 1753



INRAE

Source des images credit association Kokopelli

Ressources sémantiques : les Taxonomies

des classifications d'entités observables

SWEET ontologies

- Official repository for Semantic Web for Earth and Environmental Terminology (SWEET) Ontologies.
- des classes représentants des phénomènes par thématique
- git: <https://github.com/ESIPFed/sweet/blob/master/README.md>
- URI : forme d'agriculture <http://sweetontology.net/humanAgriculture>
- URI : phenomene de precipitation <http://sweetontology.net/phenAtmoPrecipitation>

Livestock Ontologies (ATOL, EOL, AHOL)

- des classes représentants des traits (des types de mesures) des animaux d'élevage
- documentation <https://www.atol-ontology.com/en/atol-2/>
- Agroportal: <https://agroportal.lirmm.fr/ontologies/ATOL?p=summary>



Ressources sémantiques : les unités



Ontology of units of Measure (OM)

- des URI représentants des unités et des quantités
- Git: <https://github.com/HajoRijgersberg/OM?tab=readme-ov-file>
-

Unit Ontology (UO)

- des URI représentants des unités
- Git: <https://github.com/bio-ontology-research-group/unit-ontology>

Où trouver des ressources sémantiques?

Organismes de normalisation

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

Entrepôts

- 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/>

Répertoires d'articles scientifiques

- Google scholar <https://scholar.google.com/>

INRAE

Intégration et consolidation de données

Bi Sem Web 2025 / C Roussey

Ontologies et Ingénierie des connaissances



INRAE

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 = modèle conceptuel de données sous forme d'un diagramme UML ≠ ontologie

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

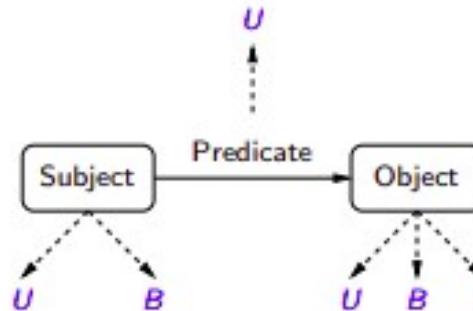
Ontologie du web : les schémas de données formalisés dans un des langages des technologies du Web sémantique

Resource Description Framework RDF

Credit Semantic Web Data/RDF/SPARQL

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

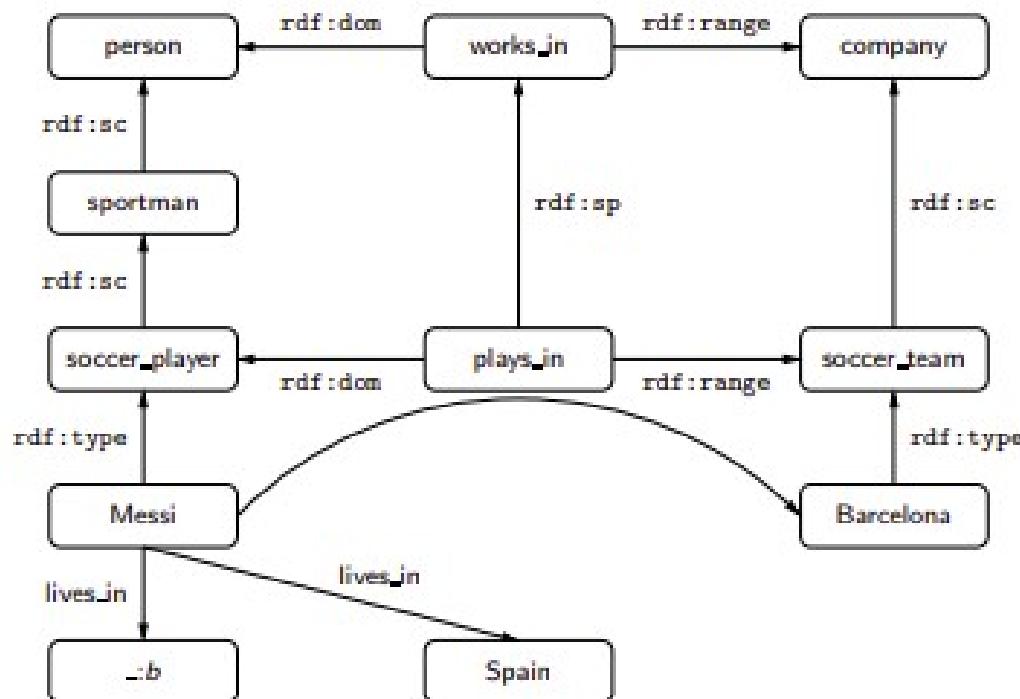
Triplet RDF



U = set of **Uris**

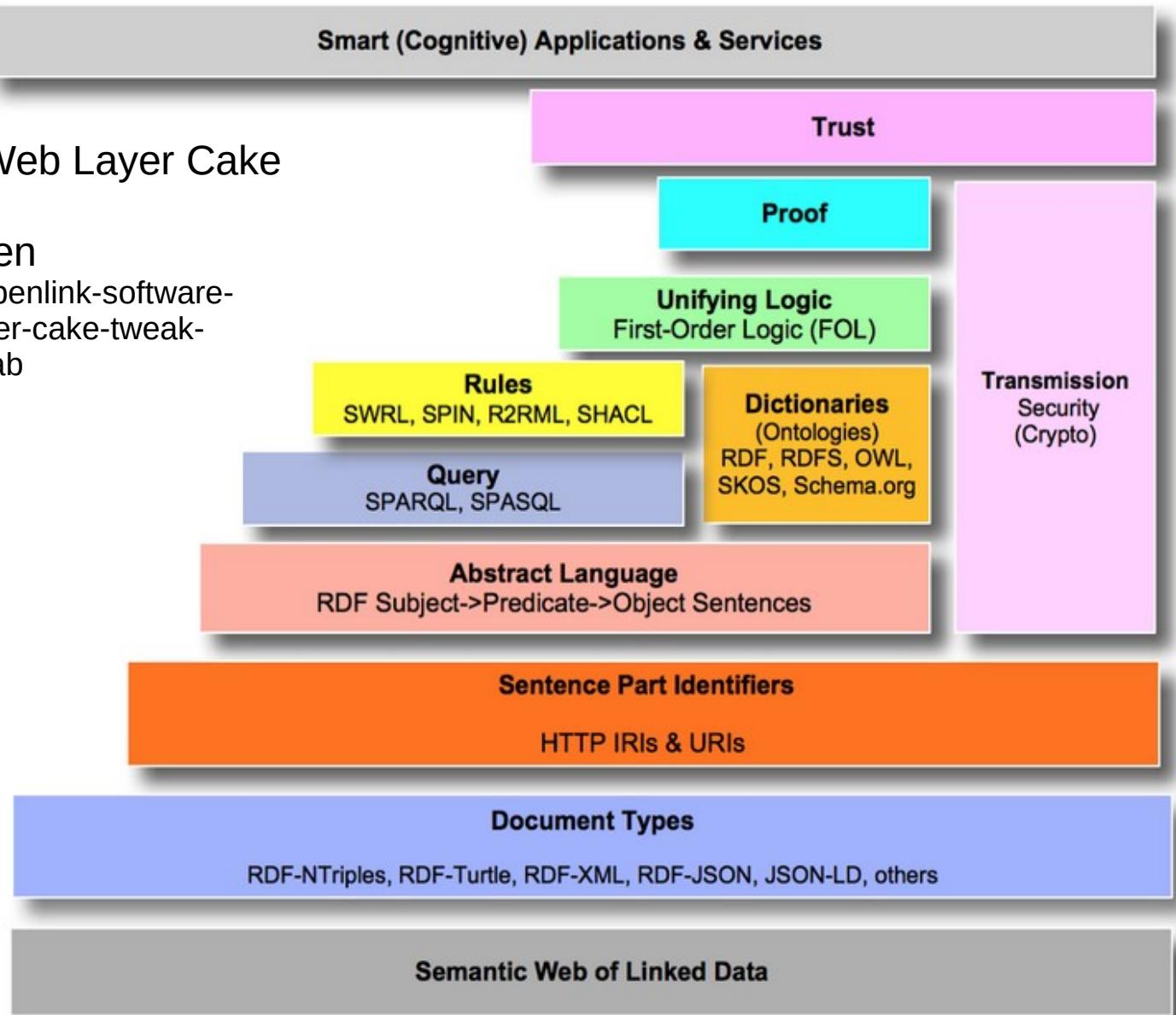
B = set of **Blank nodes**

L = set of **Literals**



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>



Typologie des ressources sémantiques par usage

classes, relation de subsomption entre classes, contraintes, ...

théories logiques

Raisonner

valider une théorie logique et déduire de nouveaux faits

axiomes logiques

classes, relation de subsomption entre classes, propriétés objets et de type de données, ...

données

Structurer

Organiser en classe les objets numériques selon leurs caractéristiques

schéma de données

classes, relation de subsomption entre classes

Indexer / classer

retrouver les objets numériques selon une thématique

taxonomie scientifique

réseau de termes, distance entre termes

terminologie

Normaliser

S'affranchir de la variabilité de la langue, proposer une recherche à facettes

thésaurus

base de données terminologiques

Typologie des ressources sémantiques par complexité croissante

classes, relation de subsumption entre classes, contraintes, ...

SPARQL

SWRL

OWL

ontologie axiomatisée

classes, relation de subsumption entre classes, propriétés objets et de type de données, ...

SHACL

SPARQL

RDFS

OWL

schéma de données

classes, relation de subsumption entre classes

SPARQL

SKOS

RDFS

taxonomie scientifique

réseau de termes, distance entre termes

SPARQL

SKOS

théssaurus

INRAE

Intégration et consolidation de données
Bi Sem Web 2025 / C Roussey

base de données terminologiques

de complexité croissante

Typologie des ressources sémantiques par usage

classes, relation de subsomption entre classes, contraintes, ...

Raisonner

valider une théorie logique et déduire de nouveaux faits

classes, relation de subsomption entre classes, contraintes, ...

Structurer

Organiser en classe les objets numériques selon leurs caractéristiques

classes, relation de subsomption entre classes, contraintes, ...

Indexer / classer

retrouver les objets numériques selon une thématique

réseau de termes

Normaliser

S'affranchir des termes pour proposer un

un réseau de termes

Une ontologie = au moins une nouvelle classe créée

taxonomie scientifique

logiques formelles

ées

terminologie

théssaurus

base de données terminologiques

Ontologie du Web W3C

et graphe de connaissances

Dans le monde du Web sémantique, une ontologie est l'ensemble des classes et des relations utilisés pour décrire un domaine d'intérêt. Les mots ontologie et vocabulaire sont utilisés conjointement. Le mot ontologie est employé quand la hiérarchie de classes et de relations est assez complexe 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

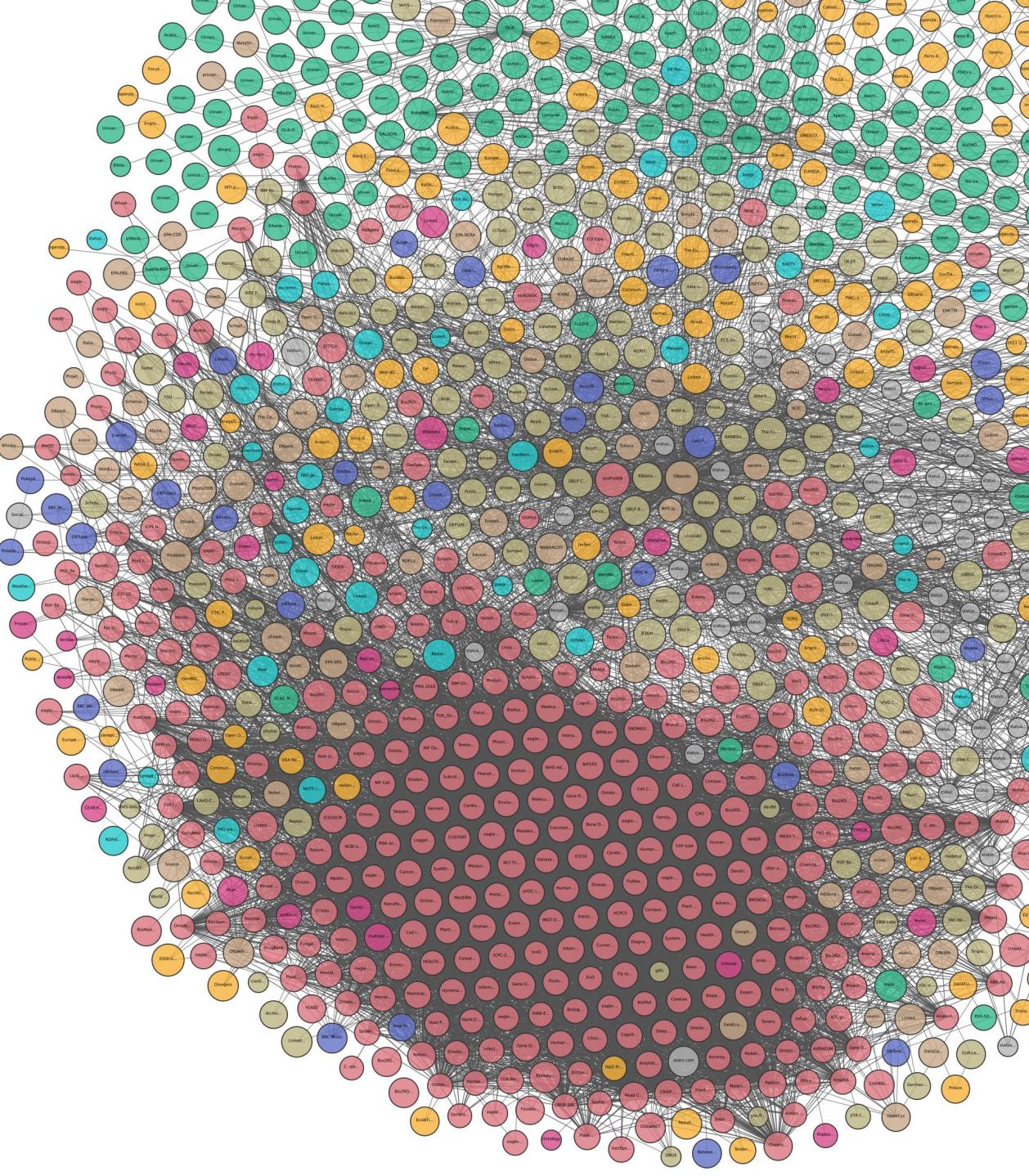
Linked Open Data (LOD)

Credit lod-cloud.net

Web de données Liées

Video europeana qu'est ce que le web des données

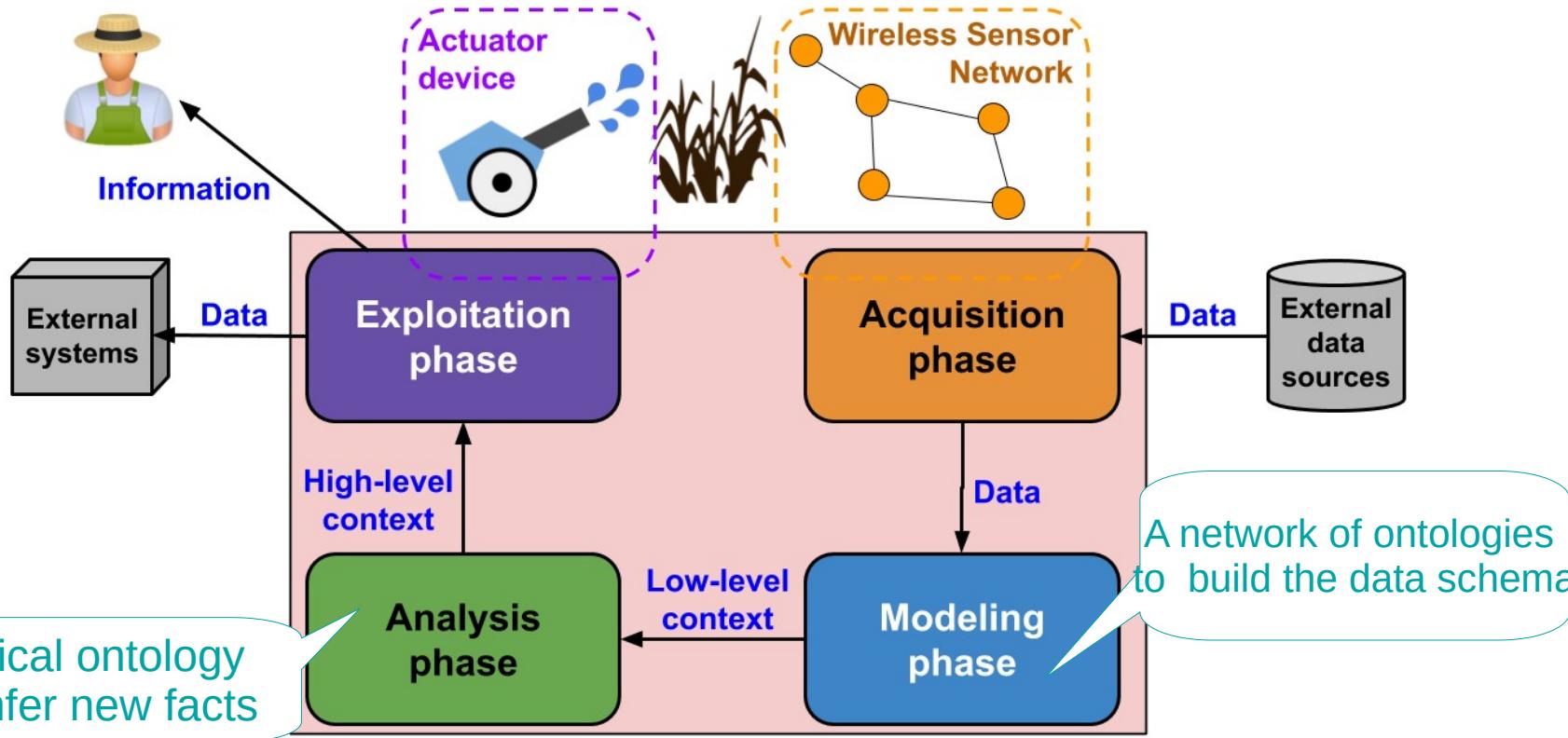
<https://vimeo.com/49231111>



Information System

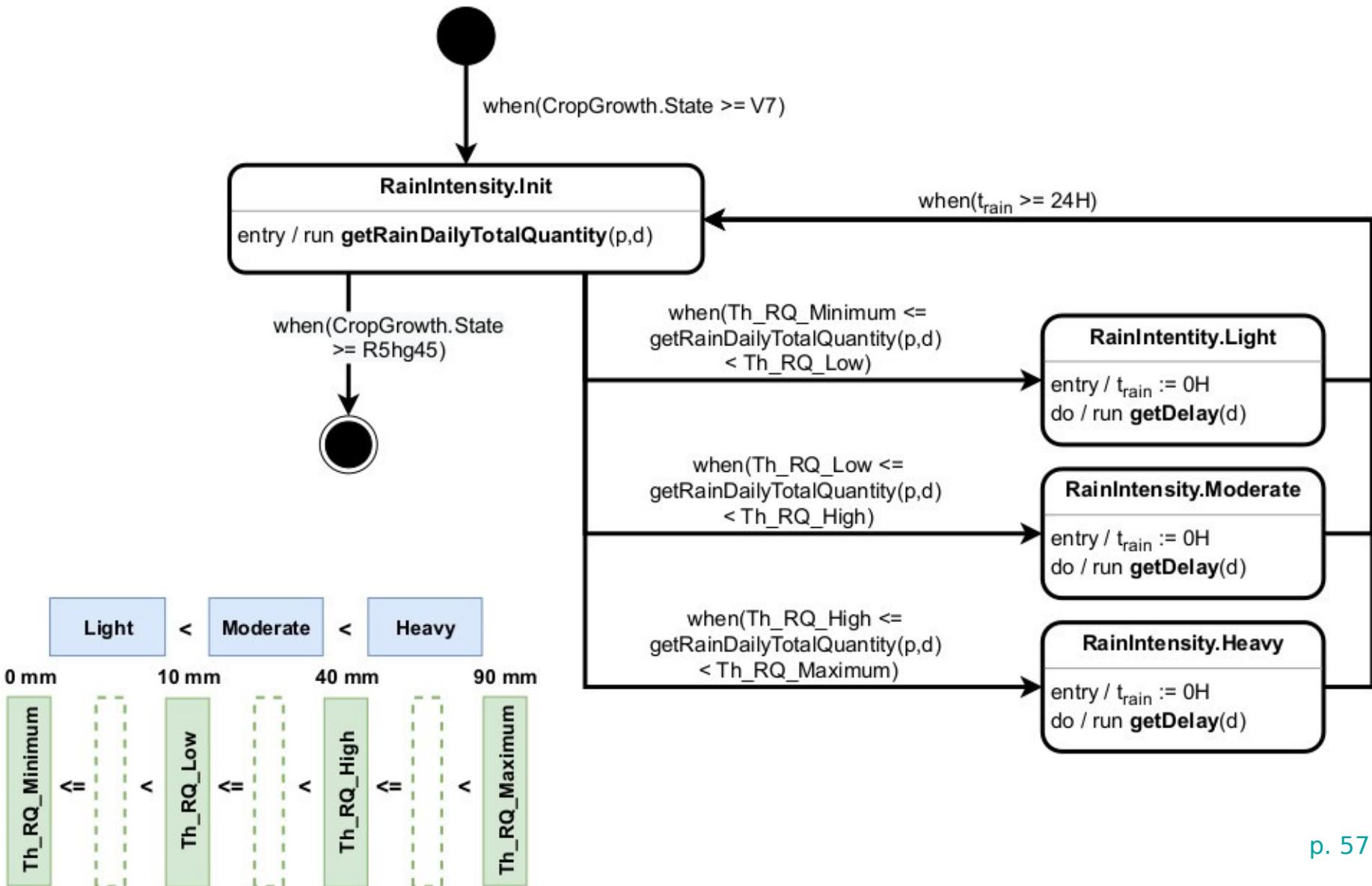
Several ontologies can be used in one system: PROV, SSN, ...

ontologies have different goals: data integration, enrichment, validation, ...

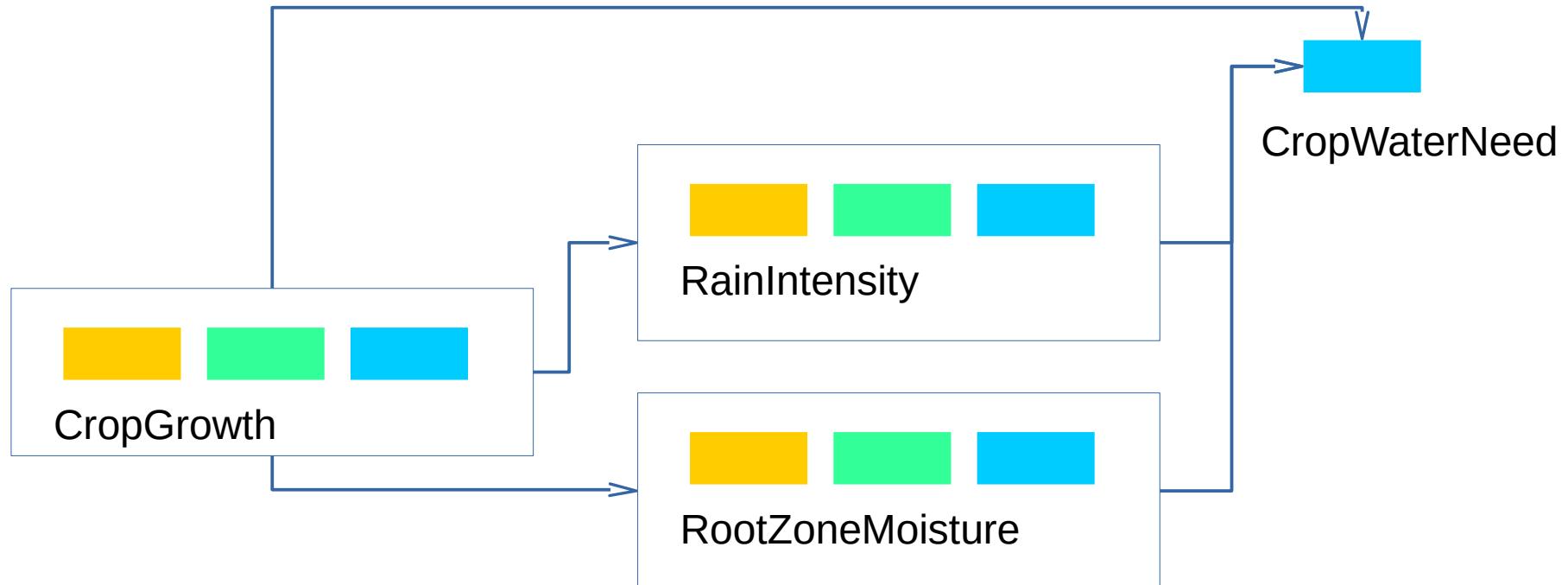


Credit Q-D. NGUYEN, C. ROUSSEY, M. POVEDA-VILLALÓN, C. DE VAULX , J-P. CHANET.
Development Experience of a Context-Aware System for Smart Irrigation Using CASO and IRRIG
Ontologies. Applied Science 2020, 10(5), 1803; <https://doi.org/10.3390/app10051803>

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



Les chaînes de traitements



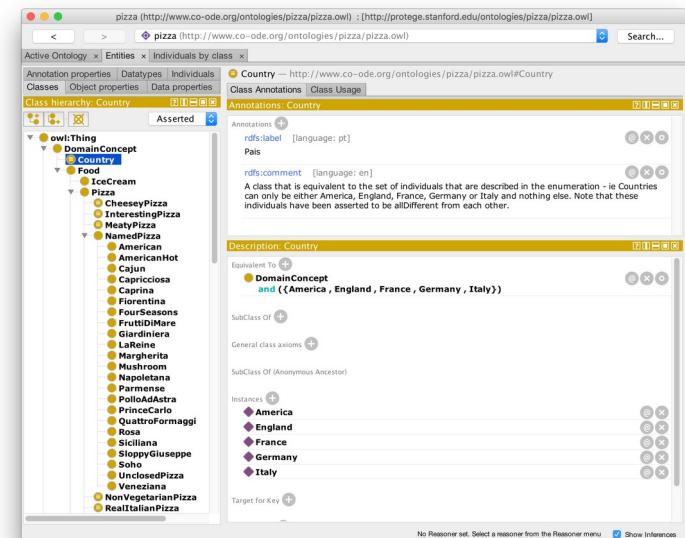
Ingénierie des ontologies

La méthode Ontology Development 101

avec l'outil Protégé et le tutorial PizzaOWL



- 1) Déterminer le périmètre de l'ontologie
 - i. avec des questions de compétences
- 2) Chercher des ressources à réutiliser
- 3) Lister les termes du domaine
- 4) Définir les classes
- 5) Définir les propriétés
- 6) Définir les contraintes
- 7) Définir les instances



Crédit

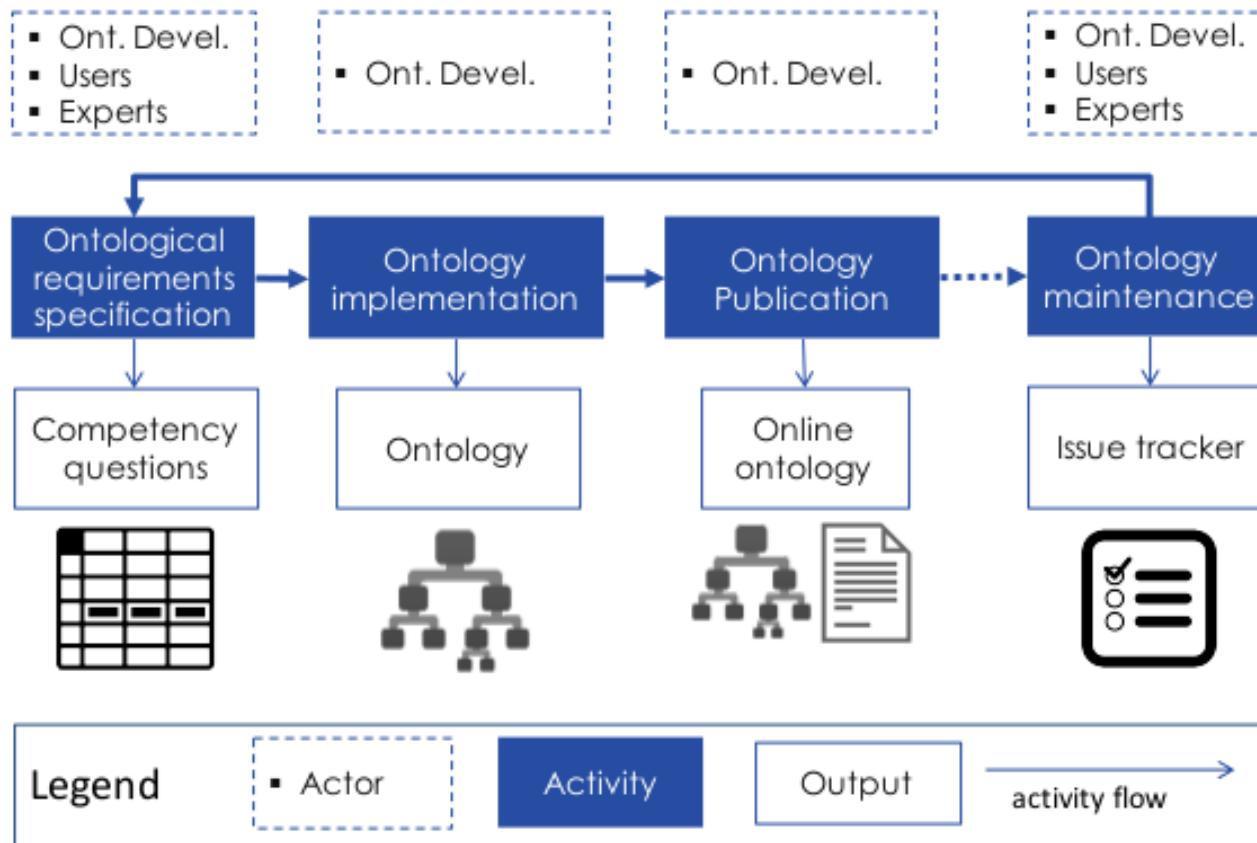
<https://protegeproject.github.io/protege/getting-started/>

Noy N., McGuinness D. (2001). Ontology Development 101: A Guide to Creating Your First Ontology.

Intégration et consolidation de données

Ingénierie des ontologies

La méthode Linked Open Terms (LOT)



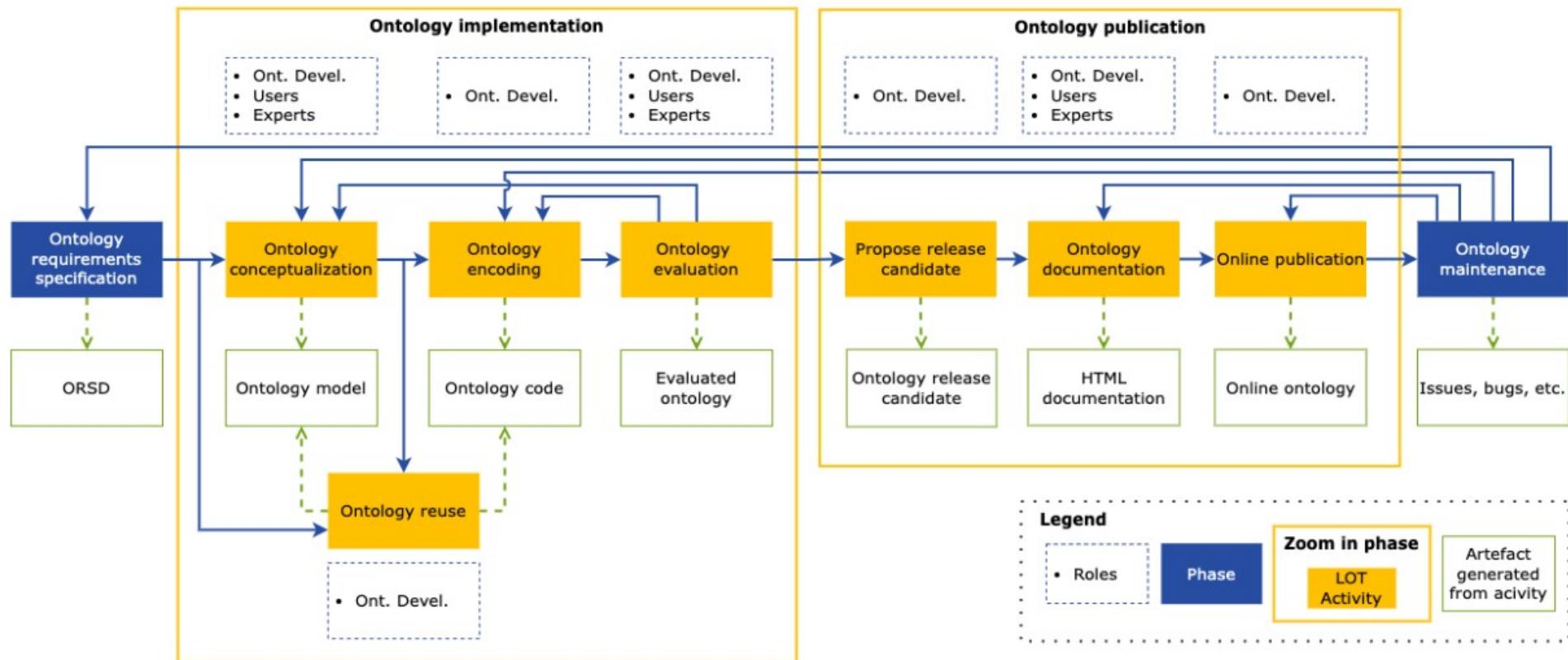
Crédit Espinoza-Arias, P., Poveda-Villalón, M. & Corcho, O.

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>

Ingénierie des ontologies

La méthode Linked Open Terms (LOT)



Ontology Design Pattern

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

ODP sont présents :

- dans les ontologies fondationnelles Basic Formal Ontology (BFO) ou Descriptive Ontology for Linguistic and Cognitive Engineering (DolCe)
- Dans les ontologies cœur de domaine (core domain) comme Semantic Sensor Network (SSN) ou Sensor, Observation, Sample, and Actuator (SOSA) ou PROV ontology

Un site web dédié et des conférences associées <https://odpa.github.io/workshop-on-ontology-design-and-patterns/2024/>

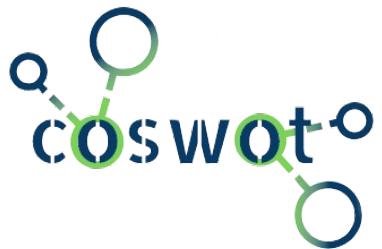
Ils peuvent prendre différents noms et forme:

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

Publication de données sur le Web : les données liées



INRAE



Bâtiment Intelligent



Credit Isaac Fatokun, Arun Raveendran Nair Sheela, Thamer Mecharnia, Maxime Lefrançois, Victor Charpenay, et al.. Modular Knowledge integration for Smart Building Digital Twins. LDAC'23: 11th Linked Data in Architecture and Construction Workshop, Jun 2023, Matera, Italy. {emse-04086955}



Bâtiment Espace Fauriel des Mines Saint-Etienne

La représentation virtuelle



Questions de Compétences

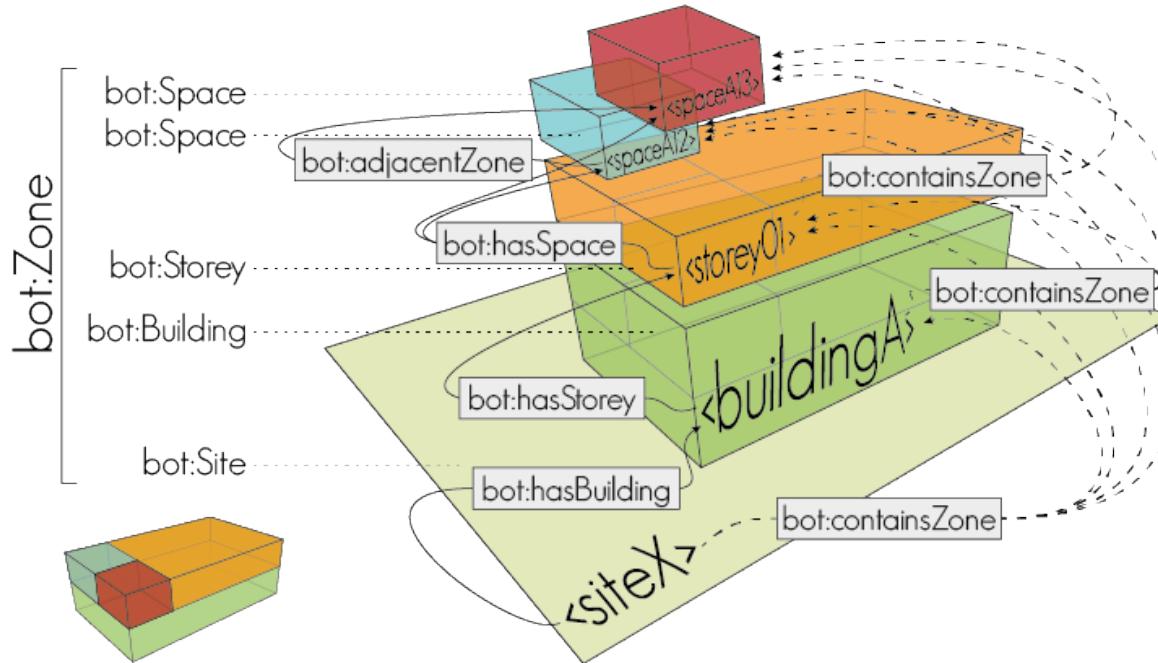
- **Où est situé un bâtiment ?** Le bâtiment Fayol est situé à Saint Etienne.
- **Quels sont les étages qui compose un bâtiment ?** Le bâtiment Fayol est composé de l'étage N°1, N°2, N°3 et 4N°.
- **Quelles sont les salles qui composent un étage ?** L'étage N°4 se compose de la salle 429, de la salle 416 etc.
- **Quel est le type d'une salle donnée ?** La salle 429 est de type bureau
- **Quelles sont les types de mesures disponibles dans une salle données ?** La salle 429 dispose de mesure de température de l'air, etc.
- **Quelles sont les unités associées aux mesures ?** La température est mesurée en degré celsius.
- ...

Le réseau d'ontologies

Building Topology Ontology (BOT) <https://w3c-lbd-cg.github.io/bot/>

Semantic Sensor Network SSN <https://www.w3.org/TR/vocab-ssn/>

Constraint Semantic Web of Thing CoSWoT <https://ci.mines-stetienne.fr/coswot/core/>



Credit Mads Holten Rasmussen, Maxime Lefrançois, Georg Ferdinand Schneider, Pieter Pauwels.
BOT: the Building Topology Ontology of the W3C Linked Building Data Group. Semantic Web – Interoperability, Usability, Applicability, 2021, 12 (1), pp.143-161. (10.3233/SW-200385). (emse-02890191)

Example d'instances

bot:Site <https://ci.mines-stetienne.fr/emse#>

bot:Building <https://ci.mines-stetienne.fr/emse/fayol#>

bot:Storey <https://ci.mines-stetienne.fr/emse/fayol/4ET#>

bot:Space <https://ci.mines-stetienne.fr/emse/fayol/4ET/416#>

bot:Door <https://ci.mines-stetienne.fr/emse/fayol/4ET/door/1VMOIJ3knDeODu7af36Nwh#>

bot:Device

<https://ci.mines-stetienne.fr/emse/fayol/4ET/416/device/1bDMdL0k55X8oOMH5VKzRL#>

bot:Furniture

https://ci.mines-stetienne.fr/emse/fayol/4ET/416/BureauD'Angle/1VMOIJ3knDeODu7af36LS_#

bot:Wall

https://ci.mines-stetienne.fr/emse/fayol/4ET/wall/1Qfe3f5n95yfrQ8nITW3_#

bot:Window

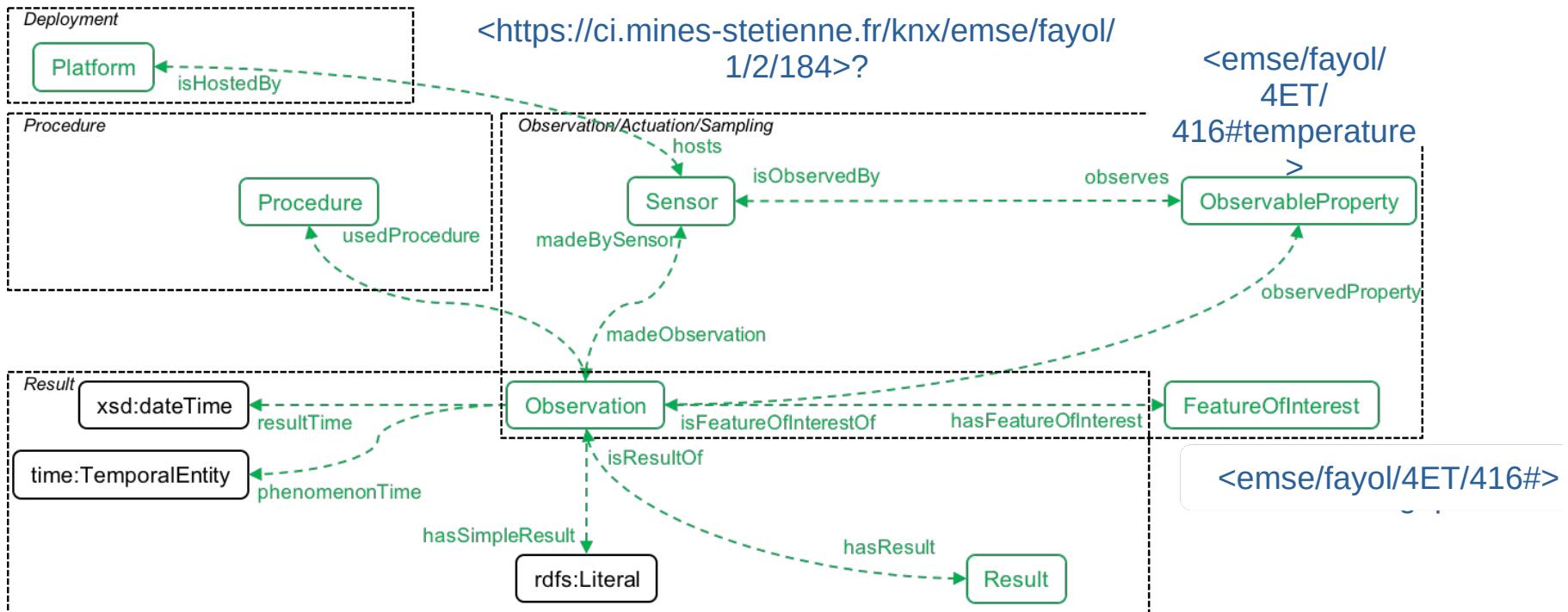
<https://ci.mines-stetienne.fr/emse/fayol/4ET/window/1VMOIJ3knDeODu7af36GPu#>

<emse/fayol/4ET/416#>

```
<emse/fayol/4ET#> bot:hasSpace <emse/fayol/4ET/416#> .  
<emse/fayol/4ET/416#> a schema:Dataset ;  
    schema:about <emse/fayol/4ET/416#> .  
  
<emse/fayol/4ET/Floor#> bot:hasSpace <emse/fayol/4ET/416#> .  
  
<emse/fayol/3ET/314#> bot:adjacentTo <emse/fayol/4ET/416#> .  
  
<emse/fayol/3ET/316#> bot:adjacentTo <emse/fayol/4ET/416#> .  
  
<emse/fayol/3ET/318#> bot:adjacentTo <emse/fayol/4ET/416#> .  
  
<emse/fayol/3ET/waiting-area#> bot:adjacentTo <emse/fayol/4ET/416#> .  
  
<emse/fayol/4ET/416#carbonDioxydeConcentration> a coswot:CarbonDioxydeConcentrationProperty .  
  
<emse/fayol/4ET/416#relativeHumidity> a coswot:RelativeHumidityProperty .  
  
<emse/fayol/4ET/416#temperature> a coswot:TemperatureProperty .  
  
<emse/fayol/5ET/514#> bot:adjacentTo <emse/fayol/4ET/416#> .  
  
<emse/fayol/4ET/414#> bot:adjacentTo <emse/fayol/4ET/416#> ;  
    coswot:canWalkTo <emse/fayol/4ET/416#> .  
  
<emse/fayol/4ET/418#> bot:adjacentTo <emse/fayol/4ET/416#> ;  
    coswot:canWalkTo <emse/fayol/4ET/416#> .  
  
<emse/fayol/4ET/waiting-area#> bot:adjacentTo <emse/fayol/4ET/416#> ;  
    coswot:canWalkTo <emse/fayol/4ET/416#> .
```

Semantic Sensor Network (SSN)

Sensor Observation Sample Actuator (SOSA)



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

Méthode de lifting de données de référence



Méthode « Linked Open Terms » (LOT)

- Questions de compétence (QC)
- Recherche de vocabulaires ou d'ontologies existantes
- **Recherche de définitions et de sources de référence**
- **Choix du niveau de formalisation OWL ↔ RDFS ↔ SKOS**
- Diagramme UML / CHOWLK
- **Outils d'extraction d'information (au besoin)**
- Production de fichiers csv avec des experts +
- **Production de définitions qui explique 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 par RDG**



INRAe

Intégration et consolidation de données

Bi Sem Web 2025 / C Roussey

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,
...

A l'issue de cette séquence, vous aurez

- Découvert les principes et les difficultés de la gestion de données
- Les données ?
- Découvert pourquoi les informaticiens construisent des schémas de données documentés et pourquoi ils ont besoins de vous
- Echanger des données: un début de bonnes pratiques
- La science ouverte et les données FAIR
- Visualiser des exemples de ressources sémantiques
- Trouver des liens utiles
- Les ontologies du Web et les données ouvertes et liées
- Partager des données sur le Web





Merci pour votre attention



p. 74



INRAE

Intégration et Consolidation de données
Bi Sem Web 2025 / C Roussey

Travaux Pratiques le bâtiment Fayol

Calculer de l'évolution des températures dans le bâtiment

- Un graphe de connaissances qui décrit le bâtiment
- Découverte du graphe et des différentes ontologies associées
- Interrogation du graphe pour calculer des indicateurs



INRAE

Intégration et consolidation de données

Bi Sem Web 2025 / C Roussey