

# Le couplage ontologies/probabilités pour l'intégration de connaissances et la rétro-ingénierie pour la conception de bioproduits ou d'aliments.



Un exemple d'application à la production de PHA, des biopolymères obtenus par voie fermentaire pour substituer les plastiques pétrosourcés dans les emballages alimentaires.

Magalie Weber, IE INRAE, UR BIA, Nantes

# Plan de la présentation

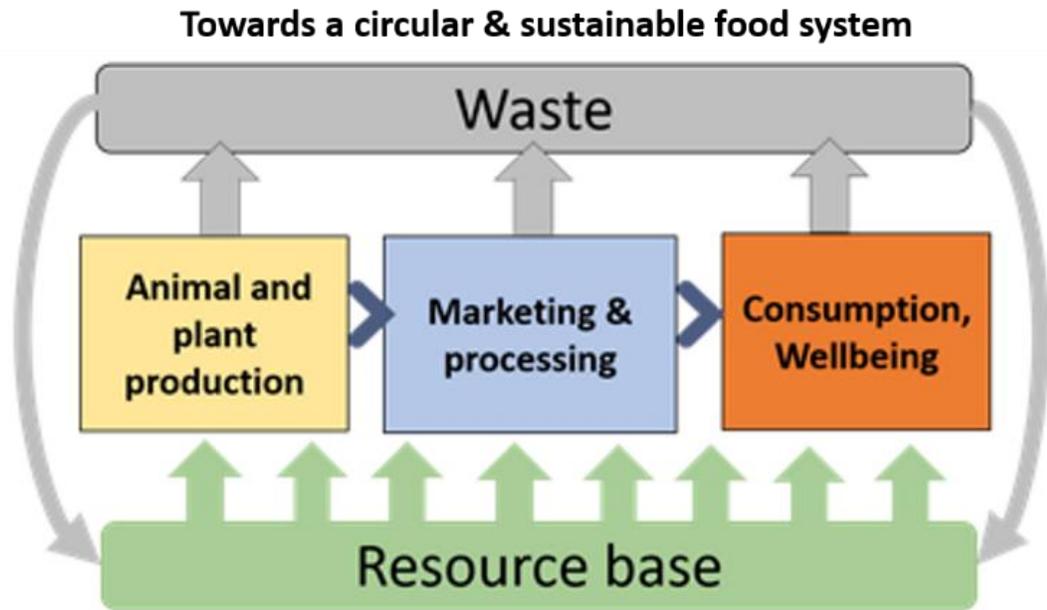
- Enjeux, verrous et solutions pour l'intégration de connaissances et la rétro-ingénierie
  - L'apport des ontologies et du web sémantique
  - Le modèle PO2 : Process & Observation Ontology
- PO2/TransformON pour les procédés de transformation de la biomasse et son extension FermentON pour les procédés de fermentation
- Le couplage ontologies/probabilités : exemple d'application à la production de PHA et de matériaux composites

# Enjeux et questions de recherche math-info

1. Intégration de données expérimentales, hétérogènes, incomplètes, imprécises
  - ✓ **Comment les modéliser?**
  - ✓ **Comment les comparer?**
  - ✓ **Comment intégrer le savoir expert?**
  
2. Raisonnement (données et connaissances)
  - ✓ **Comment raisonner dans l'incertain?**
  - ✓ **Comment découvrir de nouvelles connaissances?**
  - ✓ **Comment expliquer?**
  - ✓ **Comment évaluer la qualité/l'intégrité des données ?**

# Comment intégrer les multiples dimensions du système agro-alimentaire pour concevoir des Aliments et des Bioproduits?

Enjeux sociétaux : Bioéconomie & Alimentation avec un objectif de santé globale (*One Health*)



Environmental drivers  
Nutrition & Health  
Socio-economic drivers



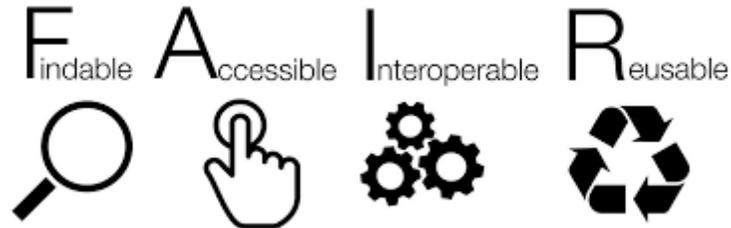
Transformation  
des bioressources

Objectifs applicatifs : valorisations alimentaires ou non alimentaires des bioressources

- Concevoir des aliments sains, bons et durables
- Concevoir des bioproduits et des matériaux biosourcés en intégrant le recyclage des résidus

# Collecter, gérer et mettre à disposition des données

Dans un contexte de science ouverte



**Données Faciles à trouver, Accessibles, Interopérables, Réutilisables**

## « Data-driven research »

- Nécessité de **formaliser la connaissance** pour la rendre « explicite » et « partageable » : pas seulement les données mais aussi les informations sur les données



**Importance majeure de l'utilisation des ontologies**

**Ontologie = une représentation structurée et formalisée du vocabulaire d'un domaine**

- un consensus entre les experts du domaine
- basée sur un formalisme logique
- exploitable par la machine

# Technologies du Web sémantique (le web des données)

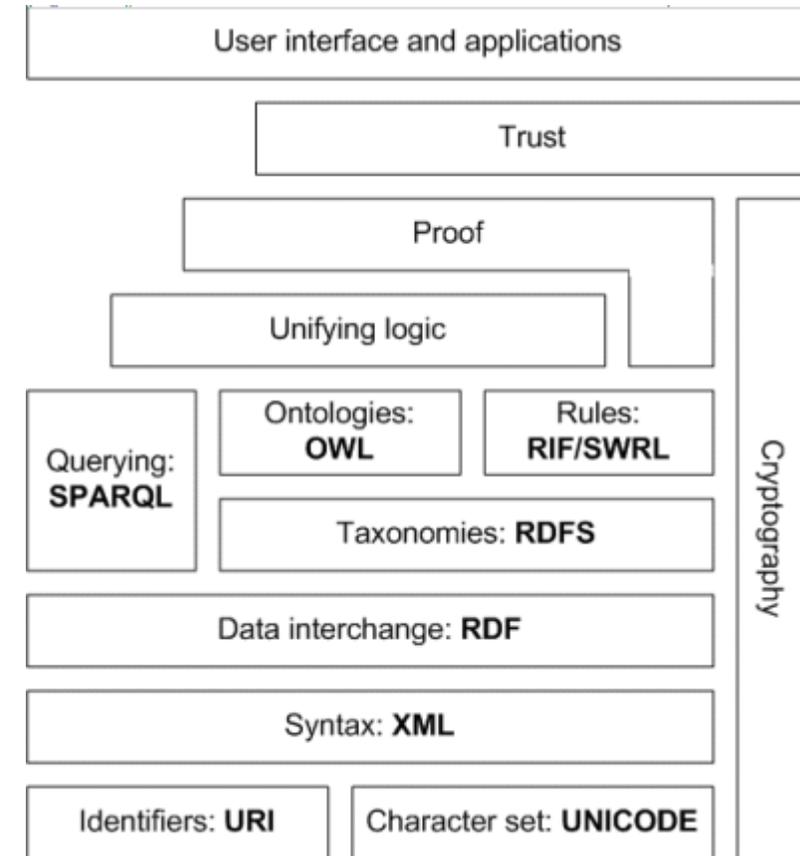
Famille de langages standards définis au niveau international (consortium W3C).

Fonctionnalités essentielles:

- **STRUCTURER:** Triplets RDF (Sujet/Propriété/Objet) identifiés avec des URI (Uniform Resource Identifiers).
- **STANDARDISER:** Existence de schémas pour décrire les classes et les propriétés des ontologies (RDFS)
- **ENRICHIR, CONTRÔLER:** Mécanismes de raisonnement et inférences logiques (SWRL, OWL, SHACL)
- **INTERROGER et INTEROPERER:** Interrogation d'une fédération de bases de données RDF avec SPARQL



Un modèle qui permet de lier et structurer l'information sur le Web sous la forme de graphes orientés (triplets RDF).



# Ontologie PO2: Process & Observation Ontology

(Ibanescu et al. 2016)

## Un **modèle** pour représenter

- des processus de transformation
- des compositions d'entrée et sortie
- des matériels et des méthodes
- des mesures ou observations
- des échelles d'observation

## Un noyau (**core model**) en OWL

qui réutilise des ontologies standards

- [SOSA/SSN](#)
- [Time Ontology](#)
- [BFO](#)

## 7 classes à spécialiser:

- PO2:Process
- PO2:Step
- PO2:Component
- PO2:Attribute PO2:Material
- PO2:Method
- PO2:Scale

## Spécialisé en **domaines d'application**:

- fabrication des gels laitiers et de fromages [[MTR 2016](#), [IDJ 2019](#)]
  - poudres végétales [[DIB 2020](#)]
  - matériaux biocomposites [[MTR 2021](#), [ESWA 2022](#)]
  - microfiltration du lait [[IJAEIS 2022](#)]
- .../...

# Ontologie de domaine PO2/TransformON

(Weber et al. 2023)

## Process Part

### PO2 / Component

#### living organism

- ▶ algae (as living organism)
- ▶ animal (as living organism)
- ▶ bacteria (as living organism)
- ▶ fungi (as living organism)
- ▶ lichen (as living organism)
- ▶ plant (as living organism)

#### substance

- ▶ biochemical constituent
- ▶ feed
- ▶ food
- ▶ non-food substance
- ▶ water (generic)

### PO2 / Process

#### physiological process

- ▶ human physiological process
- ▶ microbial physiological process
- ▶ plant physiological process

#### planned process

- ▶ characterization process
- ▶ transformation process

### PO2 / Step

#### characterization step

#### physiological process step

#### transformation step

- ▶ cleaning
- ▶ handling
- ▶ harvesting
- ▶ packaging
- ▶ pre-processing
- ▶ processing
- ▶ slaughtering
- ▶ storage
- ▶ transport

## Result Part

### PO2 / Attribute

#### calculation outcome

- ▶ experimental data attribute
- ▶ LCIA
- ▶ nutritional score

#### inherent quality

- ▶ identification attribute
- ▶ label or labelling claim
- ▶ physical state
- ▶ status of food name

#### measurement attribute

- ▶ biological attribute
- ▶ mensuration
- ▶ physico-chemical attribute
- ▶ quantity
- ▶ temporality

## Observation Part

### PO2 / Material

- ▶ measuring instrument
- ▶ processing equipment

### PO2 / Method

- ▶ analytical method
- ▶ assessment method
- ▶ computation and modelling
- ▶ semi-empiric instrumental method
- ▶ standard operating procedure

### PO2 / Scale

- ▶ measurement scale
- ▶ process scale

 Dimensions

UCUM codes

Add dimension from unit

Dimension : none

Show prefix code

Show units code

Only UCUM Code can be used in the text field.

e.g. [degF]/min -> °F/min  
ums-1.d-1 -> µm/s/d

If the unit is not available in UCUM, you can annotate your unit with {}  
e.g. (CBU)/g -> CBU/g

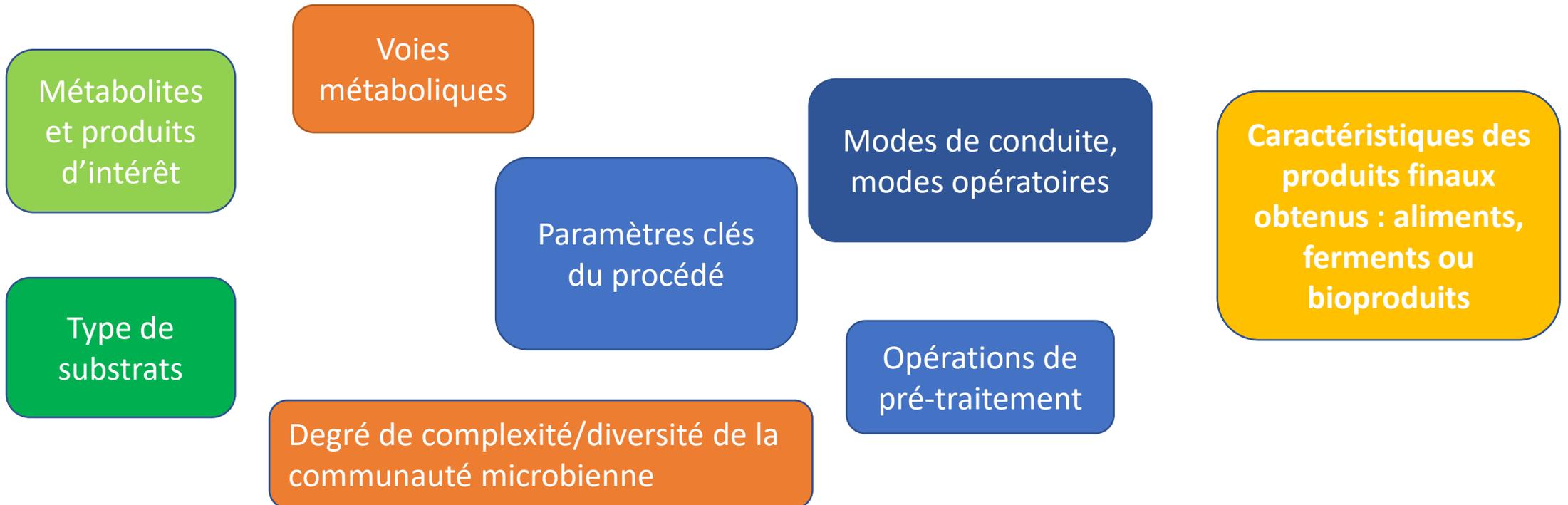
Hide Details

OK

Cancel

# FermentON : extension de l'ontologie pour les bioprocédés

**Périmètre :** concepts clefs pour décrire **les procédés de fermentation (= à l'aide d'agents microbiens)** et plus largement, les procédés biologiques (= à l'aide d'agents microbiens et/ou d'enzymes)

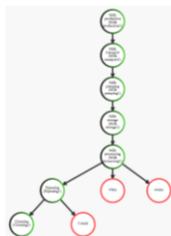
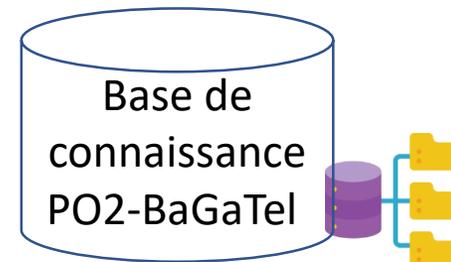


Exemples de questions de compétence auxquelles l'ontologie devra pouvoir répondre :

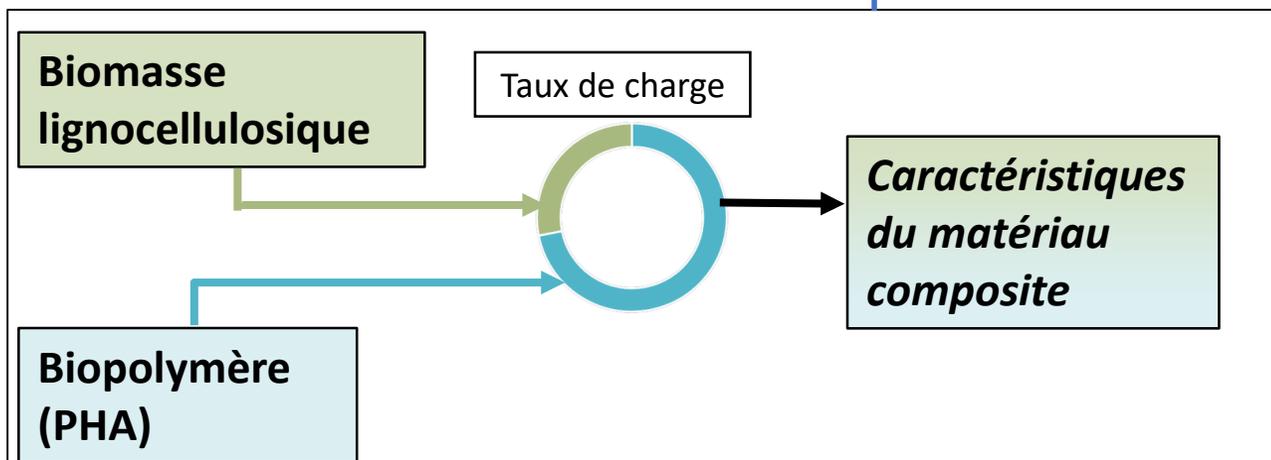
- **Quelle combinaison substrat/agent/produit visé et matériel disponible pour choisir la conduite de culture ?**
- **Quels sont les produits résultat de l'activité métabolique (produits intermédiaires et/ou métabolites) ? Avec quelles enzymes ?**

# Le couplage ontologie/réseaux bayésiens

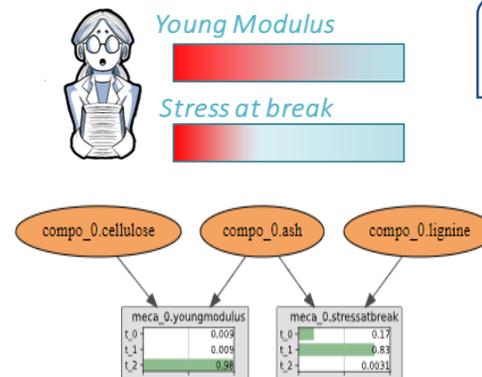
Données expérimentales structurées selon le modèle PO2

Itinéraires de fabrication du matériau composite



Trouver le taux de charge de poudre de biomasse optimum pour réduire le coût de production tout en gardant de bonnes propriétés fonctionnelles du matériau composite

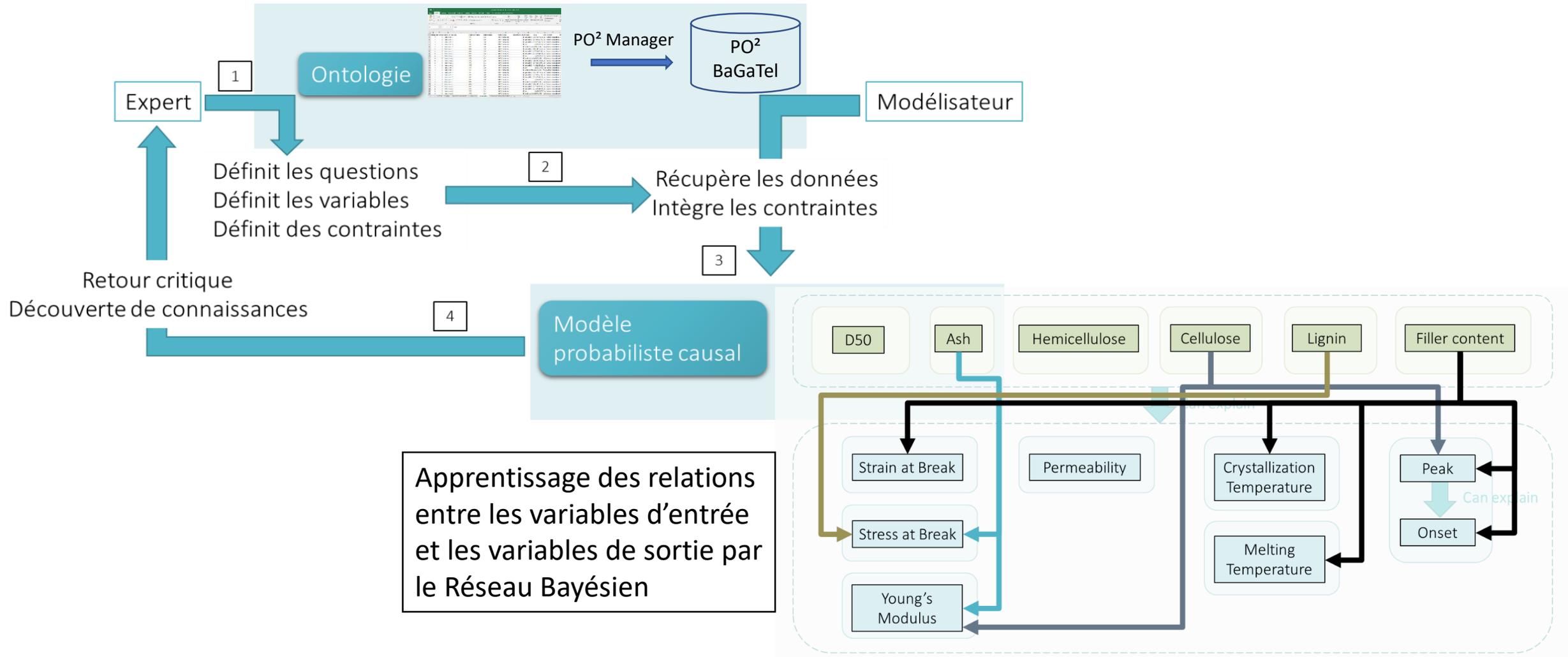


Scenario  $P=0.82$

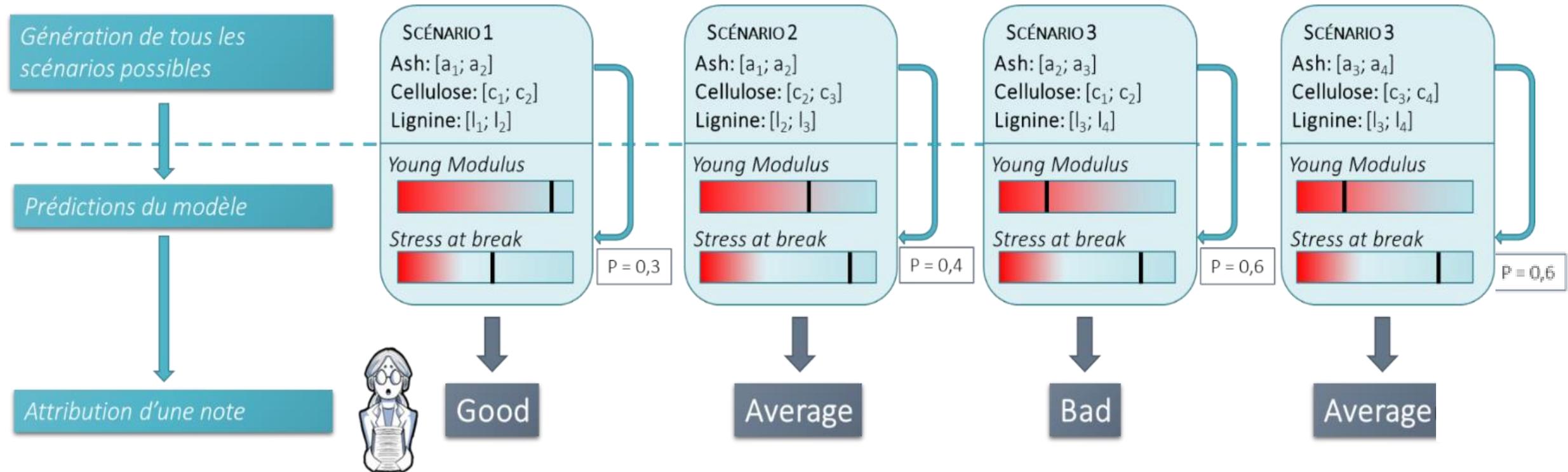
?biomasse	Rice Husk
?lignin	25,69
?cendre	14,35
?cellulose	31,9

# Raisonner à partir des ontologies à l'aide de modèle probabiliste

## Modélisation d'un système d'aide à la décision



# Quelle est la meilleure composition ?



→ Le réseau Bayésien permet de **simuler tous les cas possibles de combinaisons** entre les valeurs des variables d'entrée auxquels on associe les **préférences** des experts pour les variables de sortie

# Exemple d'inférence

Scénario **Good**

$P = 0,82$

Lignine: [19.4 ; 26.42 ]

Cendre: [6.68 ; 24.67]

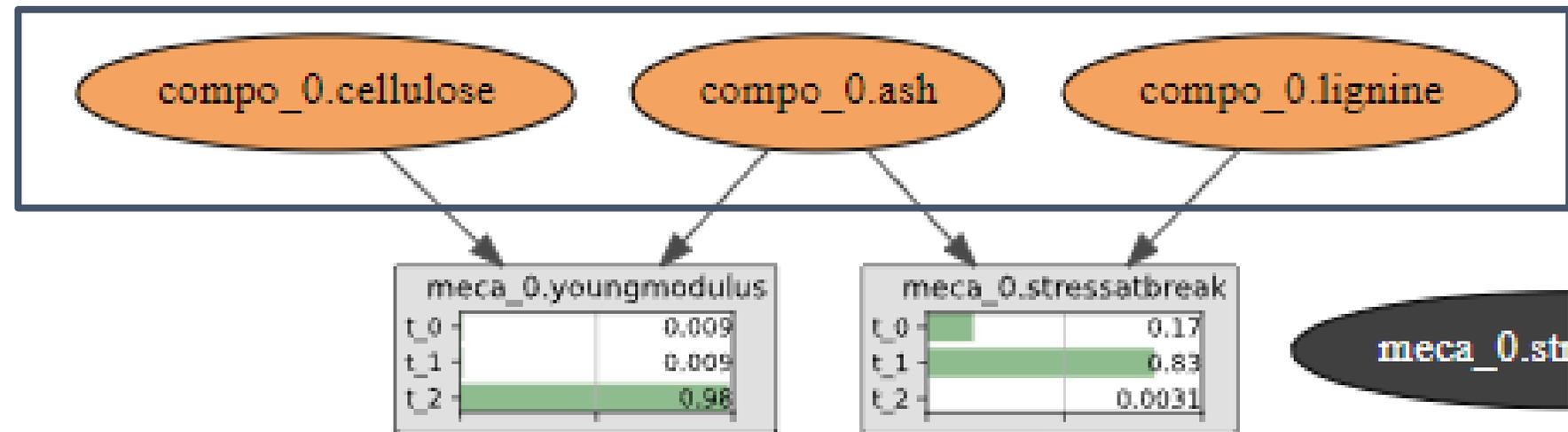
cellulose: [25.59 ; 33.05]

BIOMASSE

CORRESPONDANTE

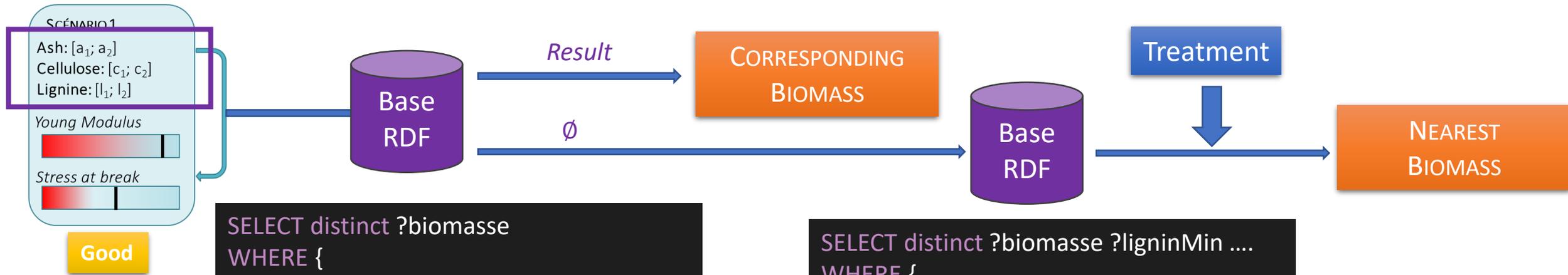
?biomasse	Rice Husk
?lignin	25,69
?cendre	14,35
?cellulose	31,9

Enveloppe de riz



→ Sélection d'un scénario correspondant aux préférences de sortie et obtention des caractéristiques de la biomasse correspondante

## Recherche des biomasses correspondant au scénario visé dans des bases de données RDF



```
SELECT distinct ?biomasse
WHERE {
  # ...
  # Constraints on the quantities
  filter(?ligninMin >= "a1"^^xsd:float &&
?ligninMax <= "a2"^^xsd:float)
  filter(?ashMin >= "c1"^^xsd:float &&
?ashMax <= "c2"^^xsd:float)
  filter(?cellMin >= "l1"^^xsd:float &&
?cellMax <= "l2"^^xsd:float)
}
```

```
SELECT distinct ?biomasse ?ligninMin ...
WHERE {
  # ...
  # No constraint, we select everything
}
```

# Conclusion et perspectives

- Méthodes d'IA basées sur la **représentation de connaissances et le raisonnement dans l'incertain** dotées de **capacité d'explication**
- Développements méthodologiques en Intelligence Artificielle pour résoudre des **besoins applicatifs à fort enjeu sociétal**
- Ce travail est réalisé dans un réseau de **collaborations pluridisciplinaires**
- **Transposition possible** à l'ensemble des transformations alimentaires ou non
- **Travaux en cours et perspectives:**
  - liage de données par l'interrogation de bases de données fédérées
  - contrôle de la qualité des données grâce aux règles d'intégrité apprises par le modèle
  - intégration dans des systèmes multi-agents

# Les principaux contributeurs



Patrice Buche  
IR INRAE  
IATE Montpellier



Cédric Baudrit  
CR INRAE  
I2M Bordeaux



Mélanie Münch  
IR INRAE/ Univ.  
Bordeaux



Stéphane Dervaux  
IE INRAE  
MIA Paris-Saclay



Liliana Ibanescu  
MC AgroParisTech  
MIA Paris-Saclay



Magalie Weber  
IE INRAE  
BIA Nantes



Juliette Dibie-Barthelemy

Prof AgroParisTech  
(anciennement UMR MIA Paris)

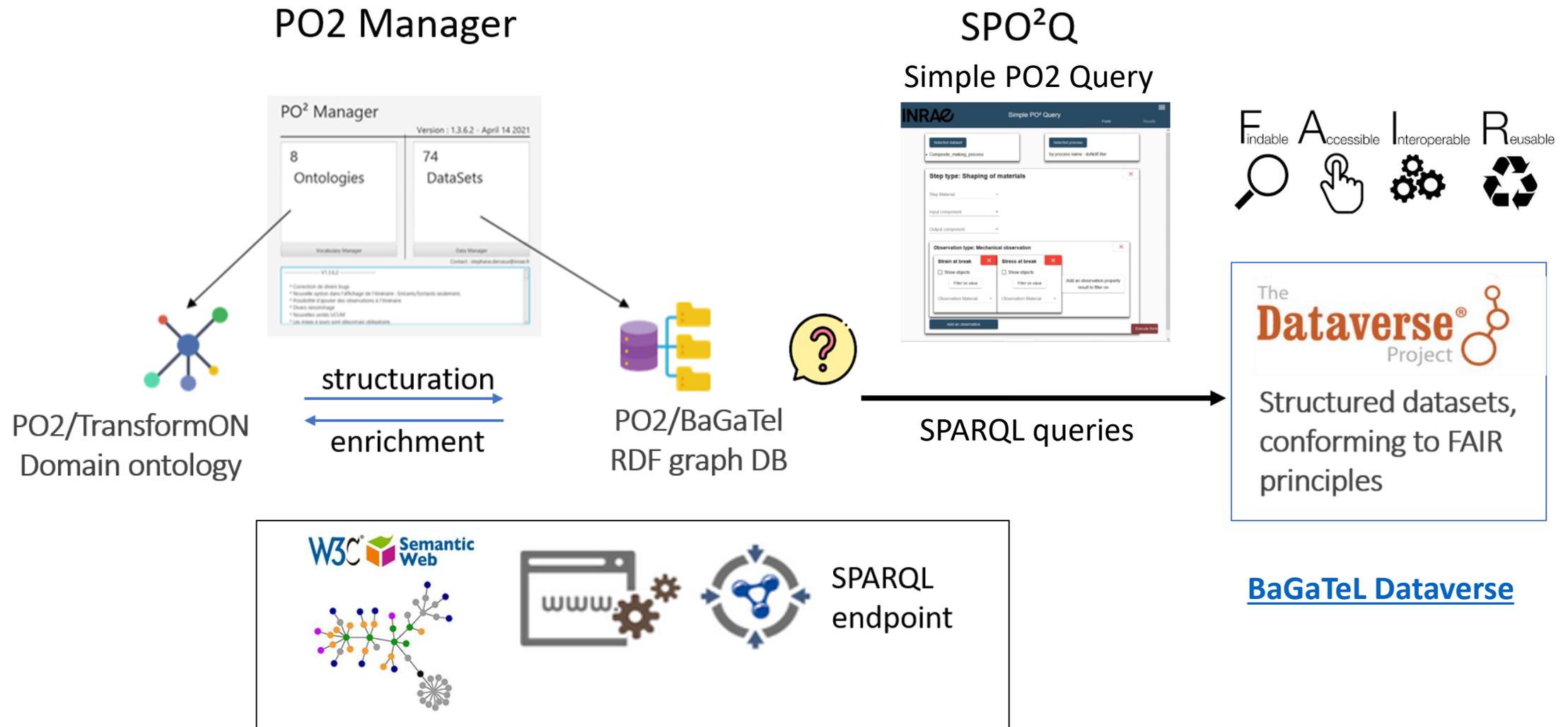
Cristina Manfredotti

MC AgroParisTech  
MIA Paris-Saclay

Pierre-Henri Wuillemin

MC LIP6  
CNRS-Sorbonne Université

# Un écosystème de données et d'outils basé sur le web sémantique



# Verrous méthodologiques: comment réutiliser et exploiter nos données ?



**Données d'intérêt**

**Hétérogènes:** multi-source, multiformat...

**Multidisciplinaires :** microbiologie, biochimie, nutrition, physico-chimie, génie des procédés...

**Complexes :** perceptions sensorielles, impacts environnementaux, performance des usages...

STRUCTURER

STANDARDISER

INTERROGER

**REUTILISER**

**OUVRIR**

Avec d'autres outils  
*Interopérabilité*

PRODUIRE DE NOUVELLES  
CONNAISSANCES,  
AIDER A LA DECISION, ...

Réutilisation des règles apprises pour  
contrôler la qualité des données

**COUPLAGE DE MODÈLES,  
SYSTEMES MULTI-AGENTS....**

# Références bibliographiques

- Weber et al. (2023). **PO2/TransformON, an ontology for data integration on food, feed, bioproducts and biowaste engineering**, *npj Sci Food* 7, 47. <https://doi.org/10.1038/s41538-023-00221-2>
- Baudrit et al. (2022). **Decision support tool for the agri-food sector using data annotated by ontology and Bayesian network: a proof of concept applied to milk microfiltration**. *International Journal of Agricultural and Environmental Information Systems*, IGI Global, 2022, 13 (1). [⟨10.4018/IJAEIS.309136⟩](https://doi.org/10.4018/IJAEIS.309136). [⟨hal-03738973⟩](https://hal.archives-ouvertes.fr/hal-03738973)
- Münch et al. (2022a). **Combining ontology and probabilistic models for the design of bio-based product transformation processes**. *Expert Systems with Applications*, Elsevier, 2022, 203, pp.117406. [⟨10.1016/j.eswa.2022.117406⟩](https://doi.org/10.1016/j.eswa.2022.117406).
- Münch et al. (2022b) **Formalising contextual expert knowledge for causal discovery in linked knowledge graphs about transformation processes: application to processing of bio-composites for food packaging**. *International Journal of Metadata, Semantics and Ontologies*, 2022, 16 (1), pp.1-15. [⟨10.1504/IJMSO.2022.131129⟩](https://doi.org/10.1504/IJMSO.2022.131129).
- Münch et al. (2021) **A process reverse engineering approach using Process and Observation Ontology and Probabilistic Relational Models: application to processing of bio-composites for food packaging**. *Communications in Computer and Information Science*, vol 1537. Springer, Cham. [https://doi.org/10.1007/978-3-030-98876-0\\_1](https://doi.org/10.1007/978-3-030-98876-0_1) [Communication MTSR 2021] [⟨hal-03474067⟩](https://hal.archives-ouvertes.fr/hal-03474067)
- Ibanescu et al. (2016). **PO2 - A Process and Observation Ontology in Food Science. Application to Dairy Gels**. In: *Metadata and Semantics Research. Communications in Computer and Information Science*, vol 672. Springer [https://doi.org/10.1007/978-3-319-49157-8\\_13](https://doi.org/10.1007/978-3-319-49157-8_13) [Communication MTSR2016]