



# Réseau IN-OVIVE

21 septembre 2021



## ➤ AIC TransformON (2021-2022)

**TransformON : Construire une ontologie sur les procédés alimentaires et non alimentaires pour permettre une meilleure interopérabilité des données et des outils développés pour l'intégration de connaissances au sein du Département TRANSFORM.**



# ➤ Éléments de contexte

Le périmètre d'étude « TRANSFORM »

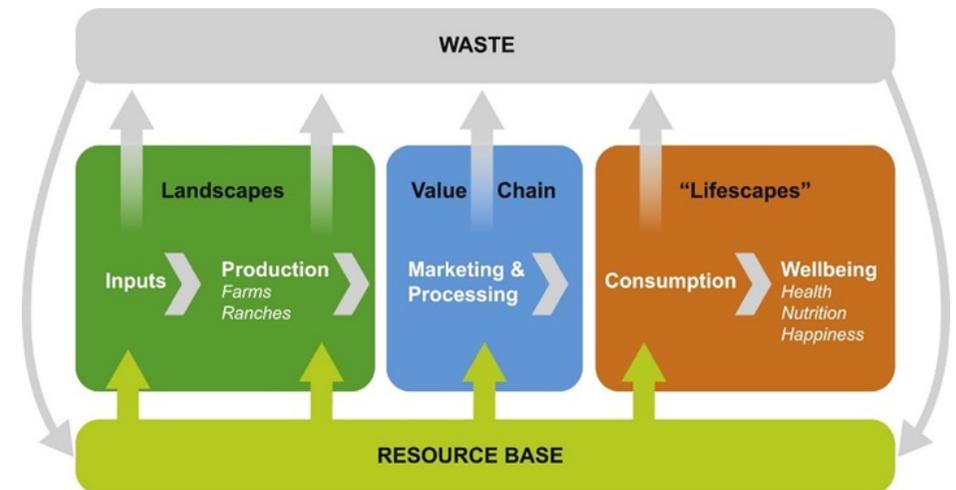
Les objectifs et les livrables du projet

L'environnement du projet et les ressources existantes



## ➤ Périmètre d'étude de TRANSFORM

- Itinéraires de construction et de déconstruction des bioressources (*intégrant les aspects sensoriels et nutritionnels, la sécurité et les techno-fonctionnalités des produits transformés, la durabilité des procédés*).
- Aliments, bioproduits et biodéchets agricoles ou alimentaires
- Liens Produit/Procédé/Emballage
- Systèmes alimentaires de la ferme à l'assiette avec rebouclage des cycles



Représentation schématique du « Food System » avec rebouclage des cycles (*from farm to fork*)

<https://doi.org/10.1016/j.agsy.2018.01.002>

## ➤ Motivations

- **TRANSFORM : besoins spécifiques /connaissances expertes**

Food system = système complexe, multidimensionnel  
Bioéconomie et Santé Globale (*one Health*)

- **Des projets existants en ingénierie de connaissances :**

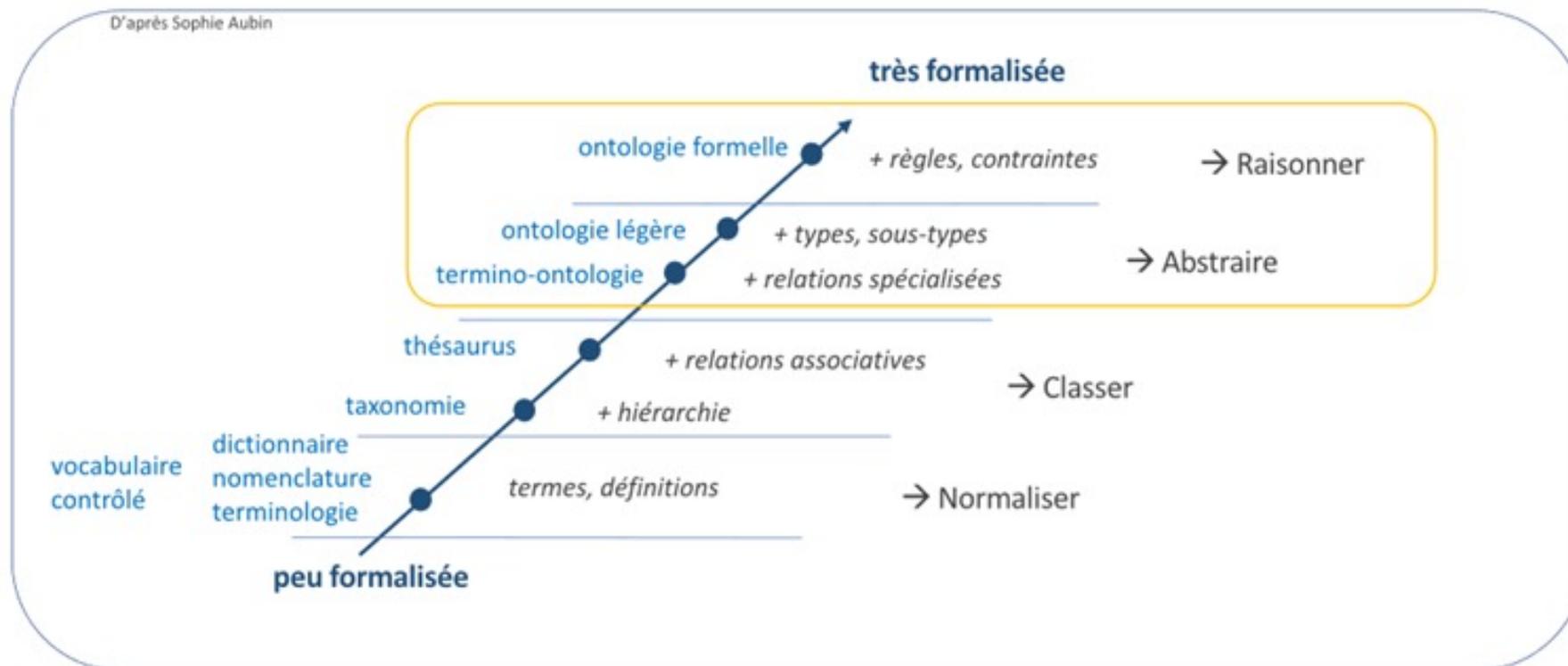
- Bases de données guidées par des ontologies : RTO (@Web, BaGaTel)
- Livres de connaissance électroniques (ex. MOOC Agreenium « Mestral »)
- Outils d'aide à la décision (ex. Ascopain, Evagrain...sur la qualité du blé)
- Listes de vocabulaires contrôlés dans divers projets

- **Développement du web sémantique (ontologies et linked data)**

Nécessité de standardiser le vocabulaire et de  
construire un *référentiel* sur le périmètre TRANSFORM



## ➤ Rappel préalable sur la terminologie ;-)



## ➤ Objectifs & livrables de l'AIC TransformON

- **Objectifs :**

- **Harmoniser les ontologies de domaine** utilisées dans les bases de données existantes (@Web, BaGaTel)
- **Proposer un référentiel termino-ontologique** basé sur le modèle PO<sup>2</sup> (Process and Observation Ontology)

- **Livrables**

- **Méthode de gestion du vocabulaire** en lien avec des référentiels internationaux et les ressources sémantiques disponibles (**dont le Thésaurus INRAE**)
- **Ontologie de domaine TransformON** sur les procédés alimentaires et non alimentaires





# ➤ Thésaurus INRAE : le référentiel institutionnel « pivot » au format SKOS

Thésaurus IRSTEA  
5 965 concepts

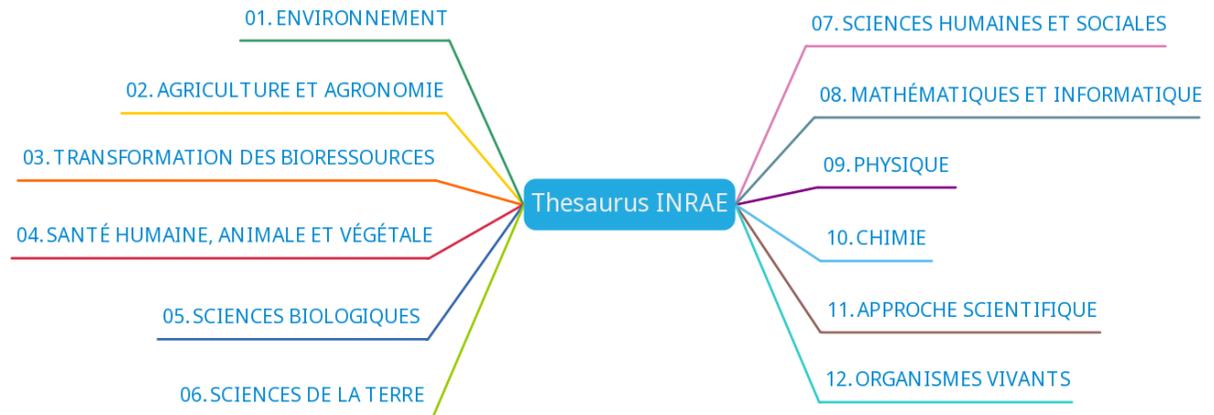
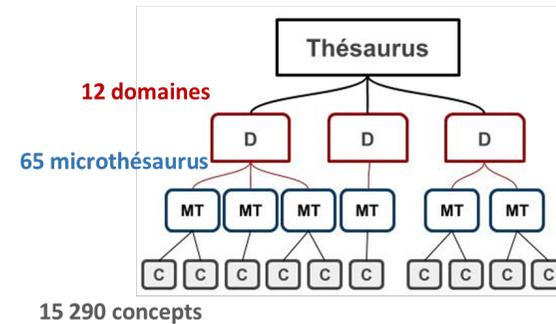


28 242 concepts à :  
- dédoublonner  
- corriger  
- classer  
- retravailler ...

Thésaurus INRAE  
15 290 concepts



Vocabulaire INRA  
22 277 concepts



## ➤ Complémentarité TransformON / Thésaurus INRAE

- **Granularité** (structuration + fine pour TransformON)

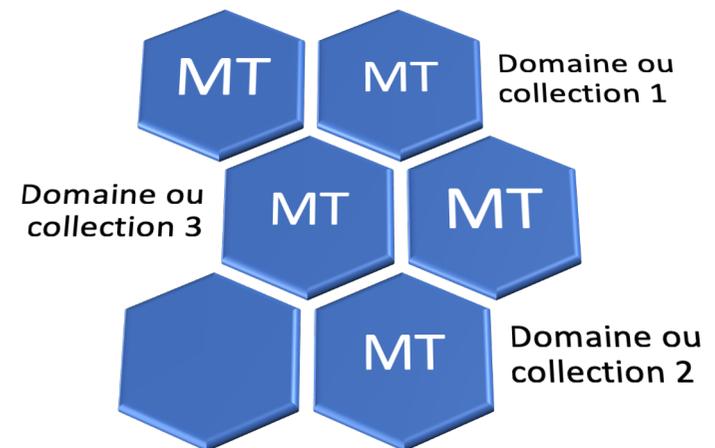
-> **concepts pivots** (= de même niveau)

-> alignements skos:exactMatch

-> spécialisation des concepts pivots

- **Modularité** (affichage modulable selon les besoins)

-> **Microthésaurus et collections SKOS**  
(MT = conceptScheme SKOS)



# ➤ Travail en cours à partir des modèles PO<sup>2</sup> et @Web

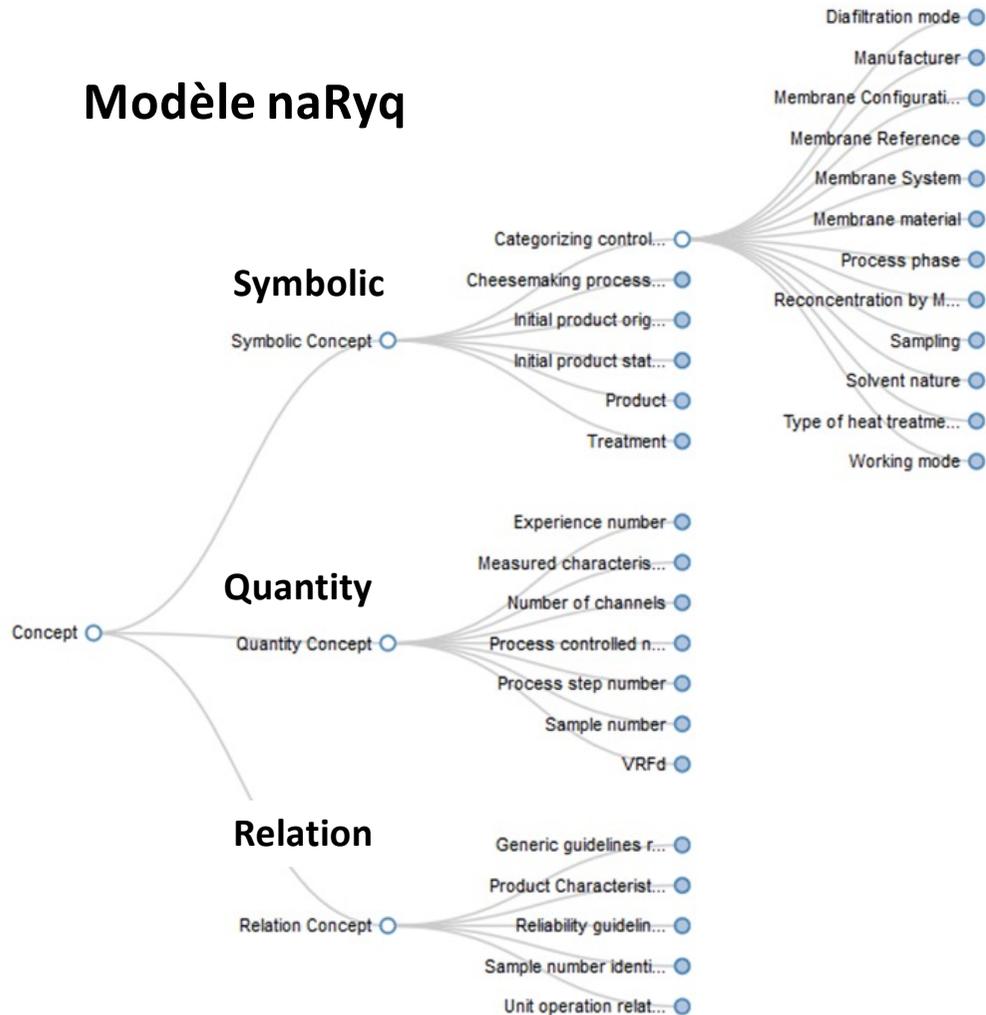
Défis et verrous à lever pour l'interopérabilité et l'intégration des données existantes





# ➤ Interopérabilité entre @Web et PO2

## Modèle naRyq



À quelles entités du modèle PO2 correspondent les concepts Symbolic et Quantity du modèle naRyQ ?

PO2 / component

PO2 / observation

PO2 / step

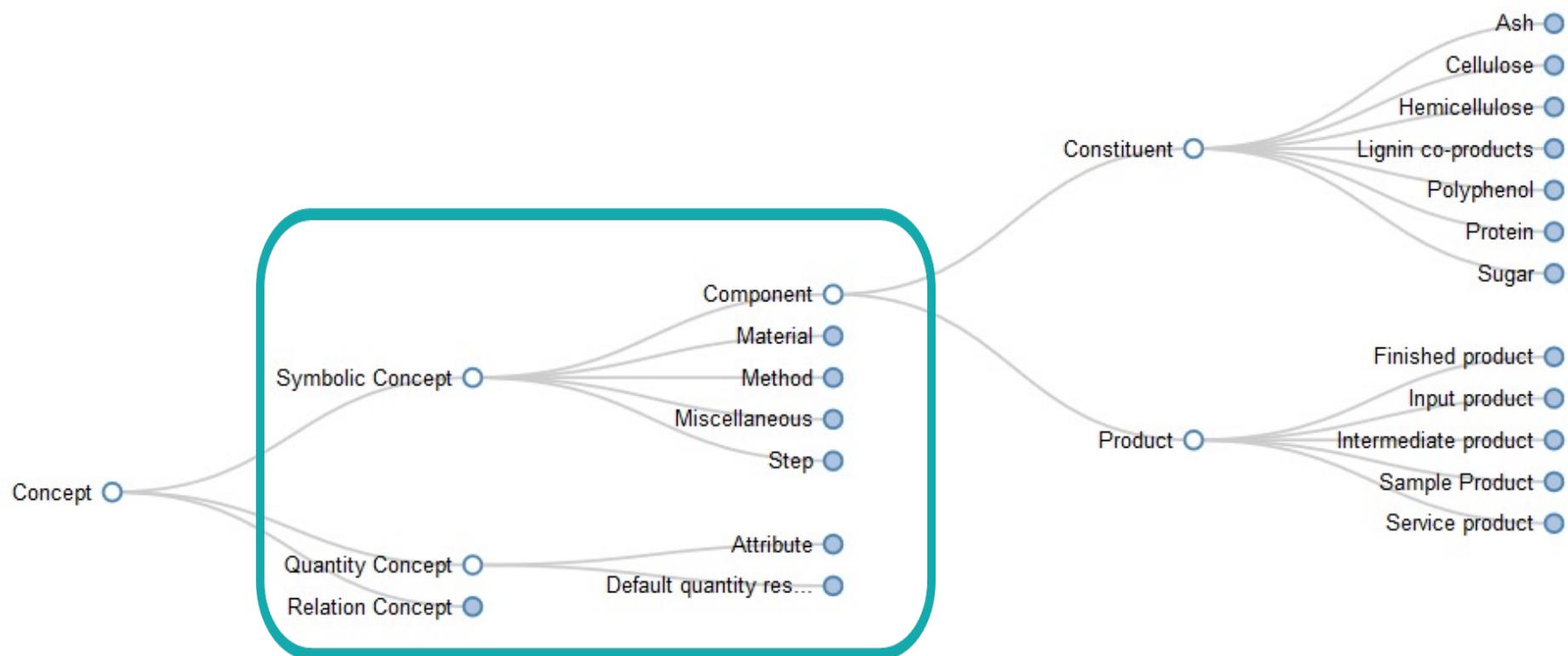
PO2 / material

PO2 / attribute

PO2 / method

PO2 / scale

## ➤ Exemple de l'ontologie de domaine Biorefinery



## ➤ Démarche pour l'intégration des vocabulaires

### Top-Down

**Créer une hiérarchie de référence avec les concepts de « haut niveau »**

- en spécialisant les « branches » de l'ontologie noyau
- en couvrant l'ensemble du domaine d'intérêt pour TRANSFORM :
  - Périmètre FOOD + BIOPRODUCTS + WASTE
  - En lien avec Amont/Aval
- en s'alignant sur des référentiels internationaux existants (ontologies, thésauri et classifications)

### Bottom-up

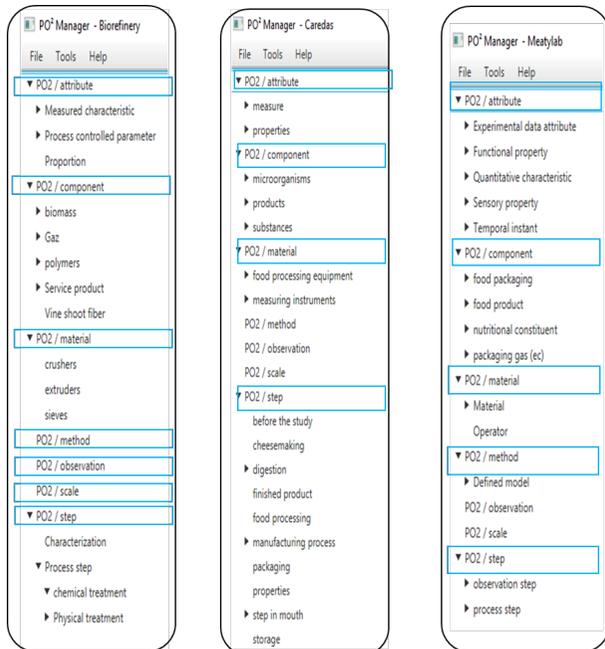
**Identifier les entités et les relations associés aux concepts « bas niveaux »**

- en s'appuyant sur des jeux de données et des publications (fouille de texte et de données)
- en identifiant les termes spécifiques ou communs aux différents sous-domaines d'intérêt
- en proposant des regroupements (catégorisation) et des définitions avec l'aide des experts (validation)



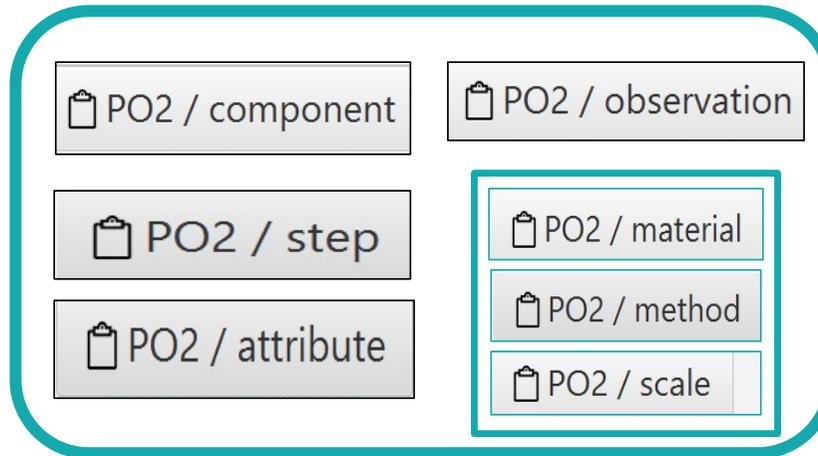
## Démarche « Bottom-up »

Quels sont les concepts utilisés pour annoter les données des différents projets existants dans les bases @Web et PO2 ?



## Démarche « Top-down »

Quelles sont les ressources réutilisables ?



Sélection de référentiels :

- **FoodON et FoodEX2** de l'EFSA pour spécialiser la partie Transformation Process (Food et Feed/organic Waste)
- Autres ontologies compatibles OBO Fundry pour la partie Transformation Process Observation (travail sur les descripteurs des M&M, les attributs et observations) : **CDNO/CheBI, OBI...**
- Autres ressources INRAE à explorer en lien avec la description des Habitats et Phénotypes : **Ontobiotope, Wheat, ATOL...**
- **Thésaurus AGROVOC et INRAE** pour l'ajout de définitions FR/EN

### **Volumétrie des données à intégrer :**

370 000 observations (~40 projets) dans PO2 RDF annotées par 3 ontologies de domaine : *biorefinery, dairy (Carédas), meat(Meatyl@b)*

+ 332 000 données dans @Web annotées par 3 ontologies de domaine : *biorefinery, TransMat, milk microfiltration*



# ➤ Nouvelle hiérarchie en cours de construction: outil collaboratif LUXID Webstudio



1 projet créé pour chaque « branche » pour intégrer et retravailler les concepts issus des différentes ontologies de domaine existantes dans un thésaurus (partie terminologique)



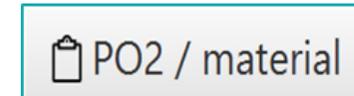
- ▼ component
  - > feed
  - > food
  - > food constituent
  - > food contact material
  - > live organism
  - > non-food matrice
  - > water (generic)



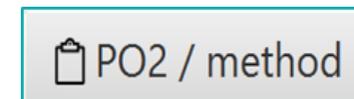
- ▼ step
  - Characterization Process step
  - ▼ Material Transformation step
    - cleaning step
    - packaging step
    - transformation step
    - transport step
  - ▼ Physiological step
    - step in mouth



- ▼ attribute
  - > Measured characteristic
  - > Process controlled parameter



- ▼ Material
  - Characterisation device (sensor)
  - Processing device (sampler)



- ▼ Method
  - > Analytical technique
  - Model
  - > Semi-empiric instrumental method
  - Sensory analysis

-> **Construction en cours de la hiérarchie commune** couvrant l'ensemble du périmètre défini :  
~ 2000 concepts pour la branche component, « alignés » sur FoodEX2

-> **Réintégration ultérieure des concepts dans l'ontologie de domaine via l'outil PO2 Manager**

## ➤ Conclusion et perspectives

Défis et verrous à lever pour l'interopérabilité et l'intégration des données existantes



## Ressources existantes

@ web  
Biorefinery @Web  
Microfiltration Lait  
TransMat  
En cours : Wheat Evagrain



URIs @web

PO<sup>2</sup>  
Biorefinery PO<sup>2</sup>  
Meatyl@b  
Caredas dairy gels



URIs PO<sup>2</sup>

## Assemblage des ressources existantes

PO2 Vocabulary & Data Manager

exactMatch

exactMatch



URIs TransformON



Gestion des URIs internes et déréférencement

## Alignements vers des ressources externes



Ressources institutionnelles :  
Thésaurus INRAE ou ontologies INRAE

Autres référentiels externes : AGROVOC,  
FoodON, etc.

Data INRAE



Accès applicatif  
Web services,  
endpoint Sparql

Triple store INRAE / DipSO



## ➤ Prochaines étapes...

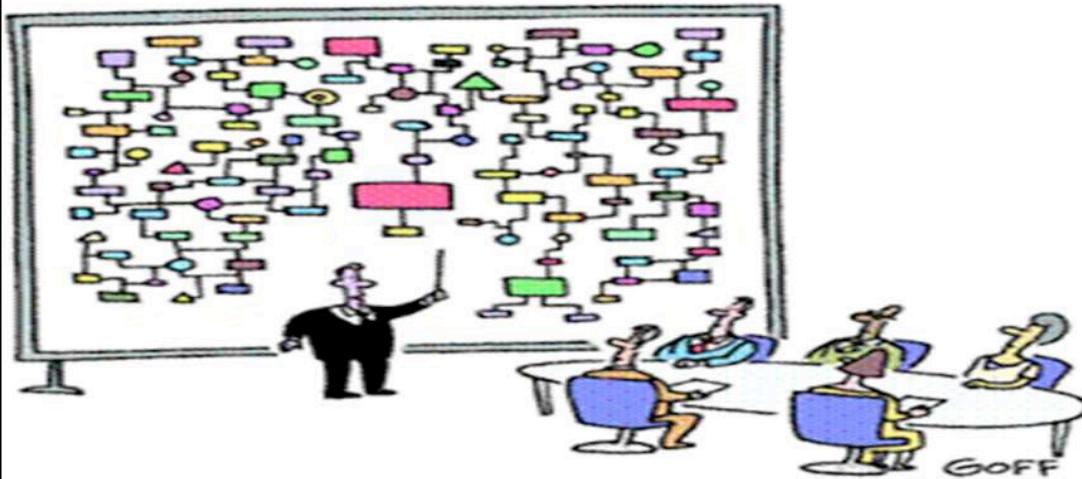
- **Evaluer l'ontologie**

- **Comment elle structure les connaissances du domaine :**
- Est-ce que tous les concepts existants peuvent se rattacher à TransformON ?
- Comment intégrer les nouveaux concepts
- Comment organiser la modularité

- **Maintenir et enrichir l'ontologie, mettre en place la gouvernance**

- Travail de curation et d'alignement sur le long terme
- Méthodes d'apprentissage pour acquérir de nouvelles données et exploiter les données (fouille de texte, analyses statistiques...)





**« And that's why we need a computer. »**

Merci pour votre attention.

Vos questions ou remarques sont les bienvenues !

