How the PO2/TransformON and FermentON domain ontologies can help us achieve interoperability and reuse of data on (bio)processes.

Magalie Weber, Ibanescu Liliana, Cufi Julien, Dervaux Stéphane

Summer Webinar 29-30 August 2024

How the PO2/TransformON and FermentON domain ontologies can help us achieve interoperability and reuse of data on (bio)processes

Magalie Weber, UR BIA, Nantes

Liliana Ibanescu, UMR MIA Paris-Saclay, Palaiseau

Julien Cufi, UMR IATE / Dpt TRANSFORM, Montpellier

& Stéphane Dervaux, UMR MIA Paris-Saclay, Palaiseau

TRANSFORM MICA MATHNUM

Sylvie Clerjon
IR INRAE
QUAPA Clermond

Amidou Traoré
IR INRAE
QUAPA Clermond

Arnaud Charleroy
IE INRAE
MISTEA Montpellier

Nicolas Forquet
AEDI-Reversaal Lyon-Villeurbanne

Liliana Ibanescu
MC AgroParisTech
MIA Paris-Saclay

Cédric Baudrit
CR INRAE
I2M Bordeaux

Christophe Fernandez
IE INRAE
I2M Bordeaux

Géraldine Lucchi
IR INRAE
CSGA Dijon

Michel Visali
IR INRAE
CSGA Dijon

Eric Latrille
IR INRAE
LBE Narbonne

Emilie Fernandez
CDD AI
LBE Narbonne

Claire Mayer
DR INRAE
IATE Montpellier

Hélène Falentin
IR HDR INRAE
STLO Rennes

Magalie Weber
IR INRAE
BIA Nantes

Stéphane Dervaux
IE INRAE
MIA Paris-Saclay

Food, Bioproducts
& Waste

Mathematics & Digital
Technologies

Microbiology
& the Food Chain

Human Nutrition
& Food Safety

Plant biology
& Breeding

Agroecosystems

Julien Cufi
IR INRAE
Dpt Transform

Mélanie Münch
IR INRAE
STLO Rennes

Adrien Réau
IE INRAE
IATE Montpellier

Virginie Rossard
IE INRAE
LBE Narbonne

Valérie Guillard
PR UM
IATE Montpellier

Valérie Gagnaire
DR INRAE
STLO Rennes

Anne Thierry
CR INRAE
STLO Rennes

Florence Valence
IR INRAE
CIRM BIA

Caroline Pénicaud
CR HDR INRAE
SAYFOOD Saclay

Emeline Goussé
IE INRAE
STLO Rennes

Leslie Lhomond
IE INRAE
IATE Montpellier

Hélène Coussy
MC HDR UM
IATE Montpellier

Reine Barbar
MC Institut Agro
IATE Montpellier

Camilo Corrales-Munoz
IR INRAE -TBI Toulouse

Alrick Oudot
CDD IE INRAE
IATE Montpellier

Patrice Buche
IR HDR INRAE
IATE Montpellie

Luc Menut
AI INRAE
IATE Montpellier

Hervé Guillemin
IR INRAE
UMR PAM
Dijon/Poligny

Evelyne Aguera
IE INRAE
UE PechRouge

Florence Jacques
IE INRAE
UE PechRouge

Gaelle Tanguy
IE INRAE
STLO Rennes

Marc Perez
IE INRAE
SPO Montpellier

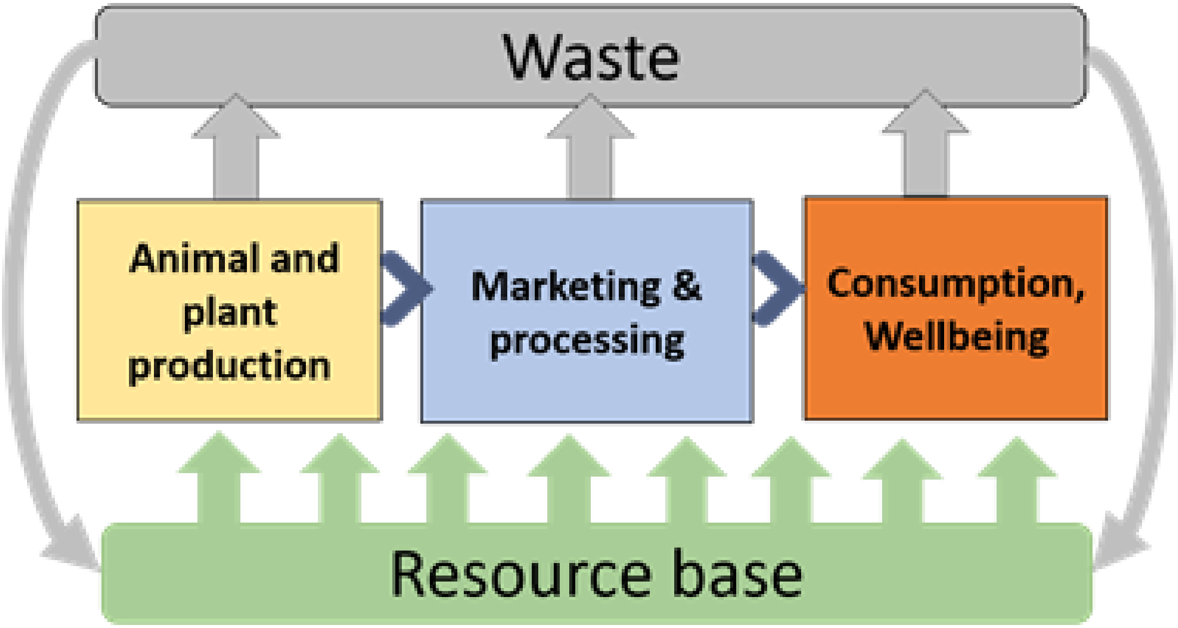
Geneviève Gésan-Guiziou
DR INRAE
STLO Rennes

Numerous collaborators…

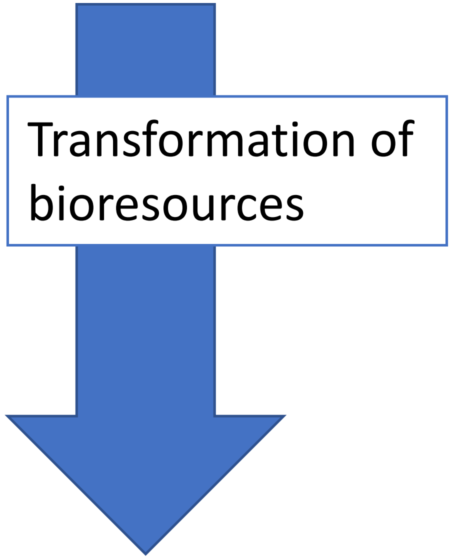
Challenges & solution

A tour on PO2/TransformON and FermentON

Food, Bioproducts and Waste engineering in the context of bioeconomy and global agri-food systems



Socio-economic drivers
Environmental drivers
Nutrition & Health



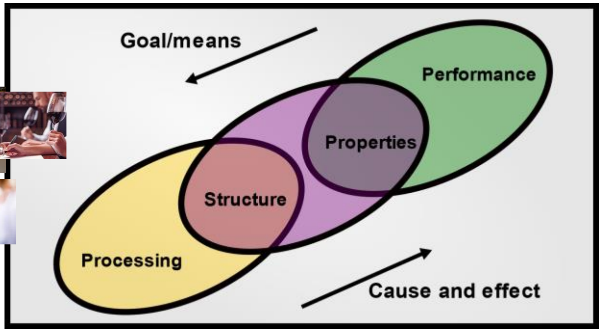
Towards a circular & sustainable agri-food system

Research challenges /ontology purposes:

- Designing food quality, from agricultural production to the impact of diets on health and the environment
- Designing biomass construction/deconstruction processes for the production of bioproducts and the use of waste (or residues) generated by human activities.

Knowledge discovery in Food and Bioproducts engineering is challenging

Complex research questions on both process performances & product usages...



Multidisciplinary fields & disciplines : Chemistry, Physics, Biology, Microbiology, Process engineering, Analytical chemistry, Sensometrics, ….

Heterogeneous experimental data,
multi-source, multi-format, multi-scale, sometimes incomplete, imprecise…

How to reason under uncertainty ??

Need for
STRUCTURING data

STANDARDISING metadata

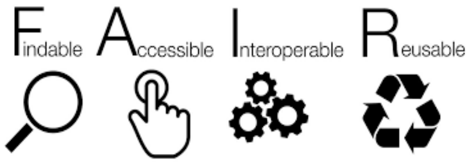
QUERYING &
LINKING data (W3C)

SHARE ------> **Re-USE**

STATISTICAL MODELLING & MACHINE LEARNING

Feedback with rules & constraints

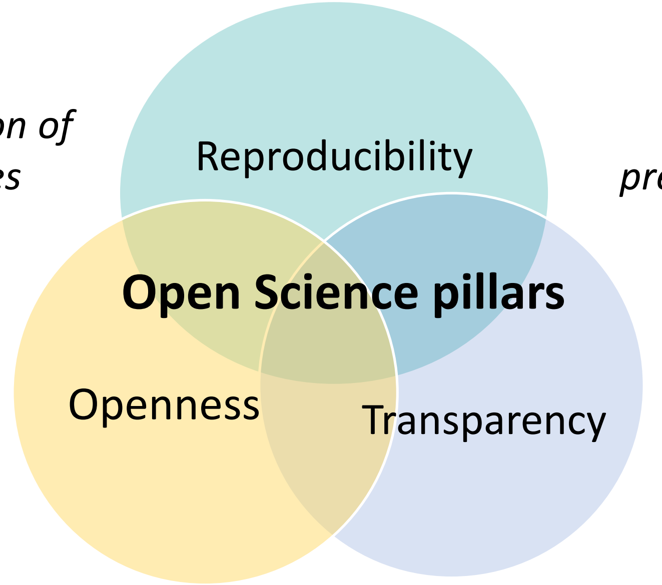
KNOWLEDGE DISCOVERY



W3C Semantic Web

How to collect, manage and provide FAIR data (by-design)

Effective collaboration of research communities



Long-term preservation of data

Build Machine-Learning models

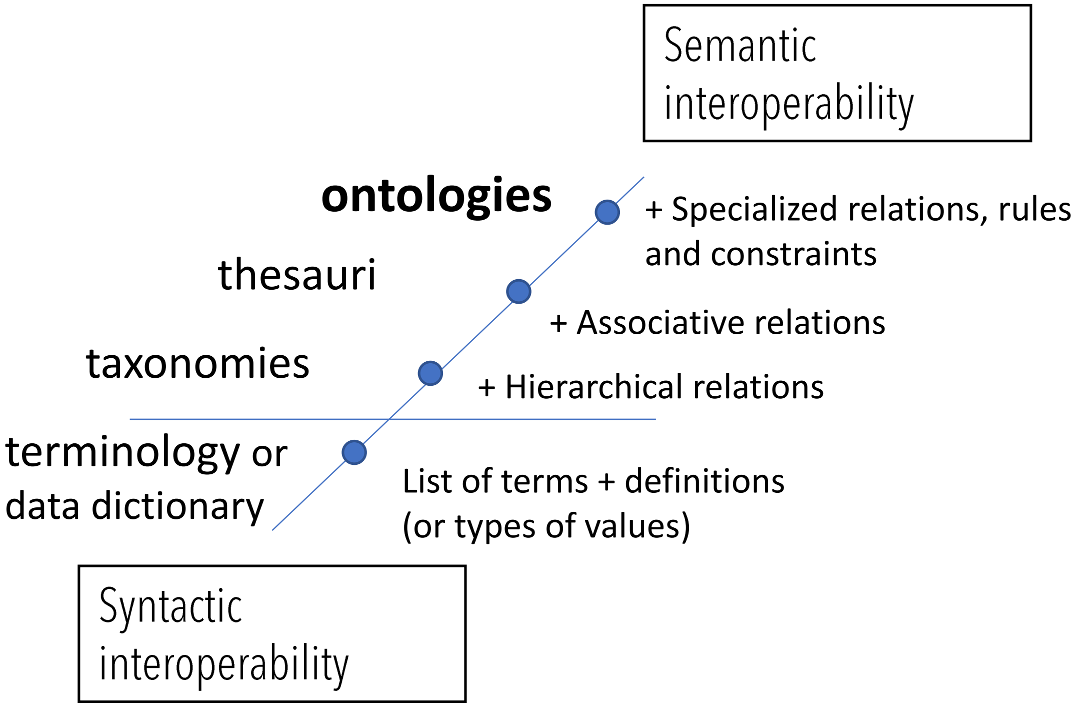
In the context of « Data-driven research » and Open Science, we need to formalize knowledge to make it "explicit" and "shareable": ***not** just data* but also **information about data (*metadata*)** to share the context and provenance of data

Ontologies are a key solution for interoperability & reusability (I & R)

Why are ontologies so useful?

Semantic interoperability is **unambiguous**, leaving few room for error or misinterpretation.

Semantically interoperable systems share the **same conceptual understanding of what the underlying data conveys,** thanks to ontologies.



There is a broad spectrum of ontologies :

- **Upper-level : foundational ontologies**
  - *BFO, DOLCE, …*
- **Mid-level : core ontologies, task ontologies**
  - *SOSA/SSN, PROV-O, Time ontology…*
- **Application level : domain ontologies , termino-ontologies**
  - *Vocabulary specific to the field (labels+ definitions)*

➢ *The domain ontologies, PO2/TransformON and FermentON are based on the PO2 core model*

Syntactic interoperability provides a **consistent data format** that can be interpreted by different systems.

Ontologies allow us to reach *Semantic interoperability*

The Process and Observation Ontology (PO2)

<u>A generic model in OWL</u>,

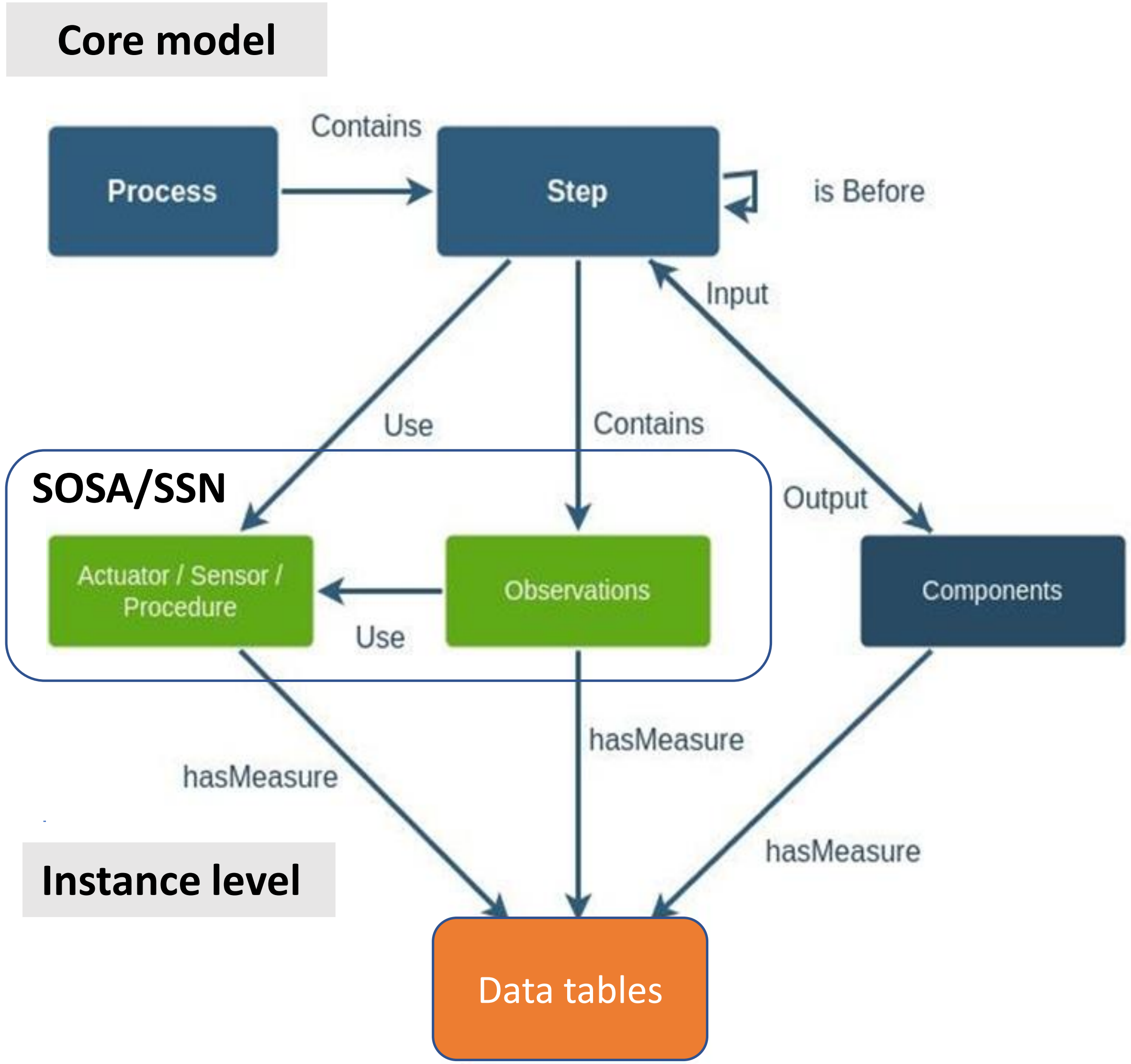
reusing SOSA/SSN concepts

together with other ontologies (BFO, IAO, Time ontology, QUDT), metadata schemas (DCAT, Schema.org), and the Unified Code for Units of Measure (UCUM)

<u>designed to describe :</u>

- Processes as a sequence of steps
- Input and output components
- Observations with materials and methods
- Results described in data tables at the instance level

- **Quantitative and qualitative values in the data tables conform to the I-ADOPT Variable Design Patterns** (https://catalogue.fair-impact.eu/resources/i-adopt)



Ibanescu et al. 2016

A unified vocabulary (TransformON) specializing the PO2 core concepts

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

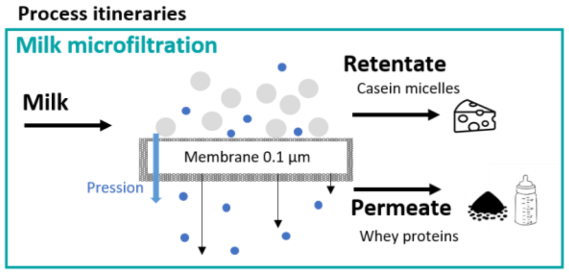
- PO2 / Method
  - analytical method
  - assessment method
  - computation and modelling
  - semi-empiric instrumental method
  - standard operating procedure

- PO2 / Scale
  - measurement scale
  - process scale

Weber et al. 2023

A variety of use cases and research objectives have been addressed

Objective: Predicting the performance of the milk microfiltration process

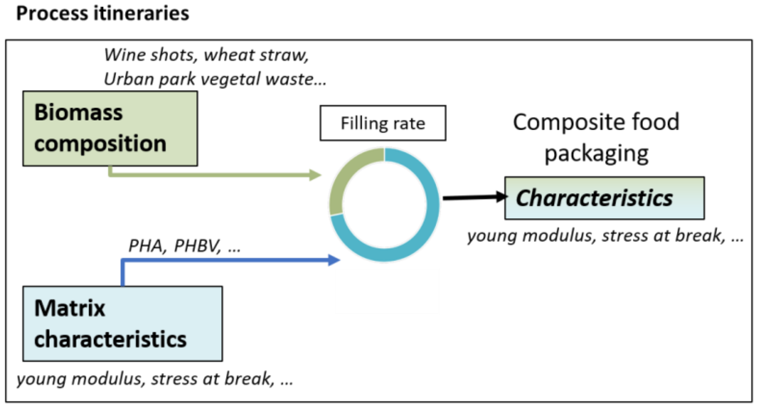


Baudrit et al. 2022

Problem to be solved:

- **Very heterogeneous and incomplete data**
- **Existing models limited to a specific range of operating conditions**

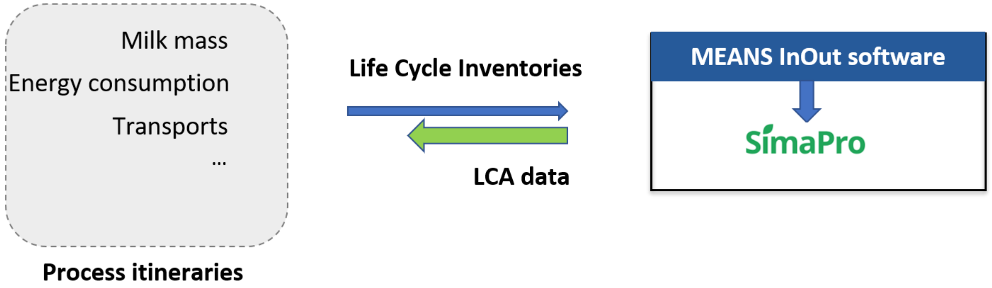
Objective: optimization of composite food packaging formulation



Münch et al. 2022

Problem to be solved : find what biomass and what loading rate to achieve the desired mechanical characteristics

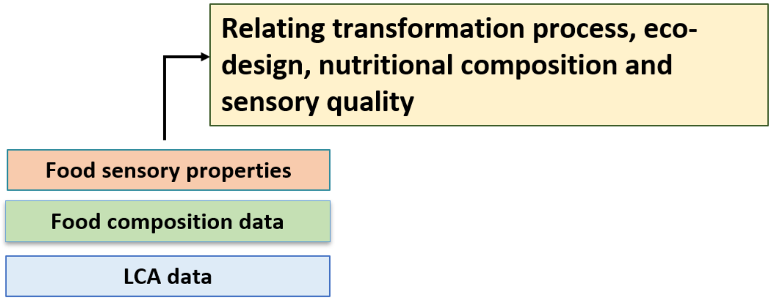
Objective: to study the environmental impacts of cheeses



Auberger et al. 2022

Problem to be solved : providing the process flowchart and inventory data needed for Life Cycle Impact Assessment

Objective: to design healthy, sustainable and acceptable food



Pénicaud et al. 2019

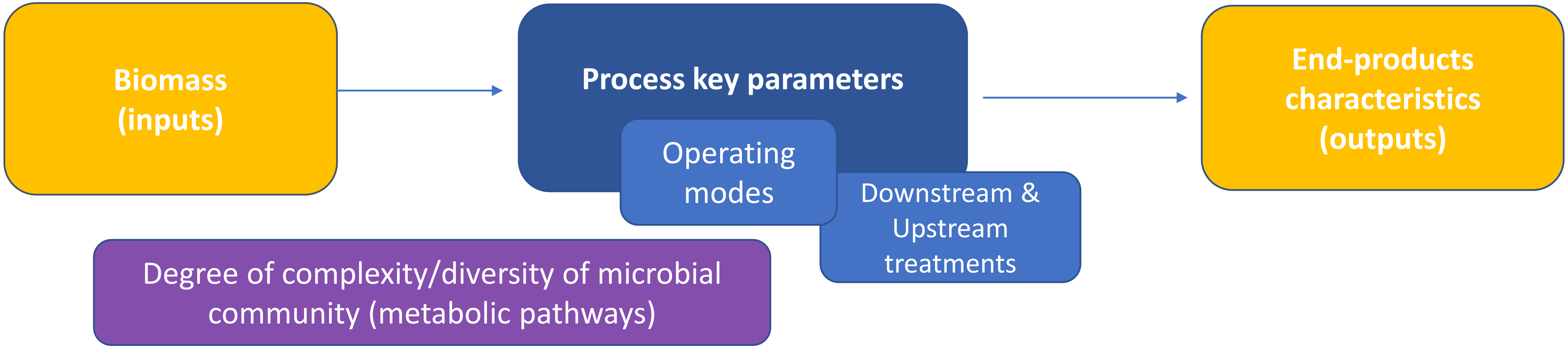
Problem to be solved: integration of heterogeneous data to provide a global quality score

What's next ?

Ongoing project FermentON

FermentON : extension of PO2/TransformON to bioprocesses

Scope: key concepts to describe **fermentation processes** (= using microbial agents) and, more broadly, **biological processes** (= using microbial agents and/or enzymes).



Examples of questions to be answered :

What combination of substrate/agent/target product and available equipment is needed to choose the cultivation method?

What products result from metabolic activity (intermediates and/or metabolites)? With which enzymes?

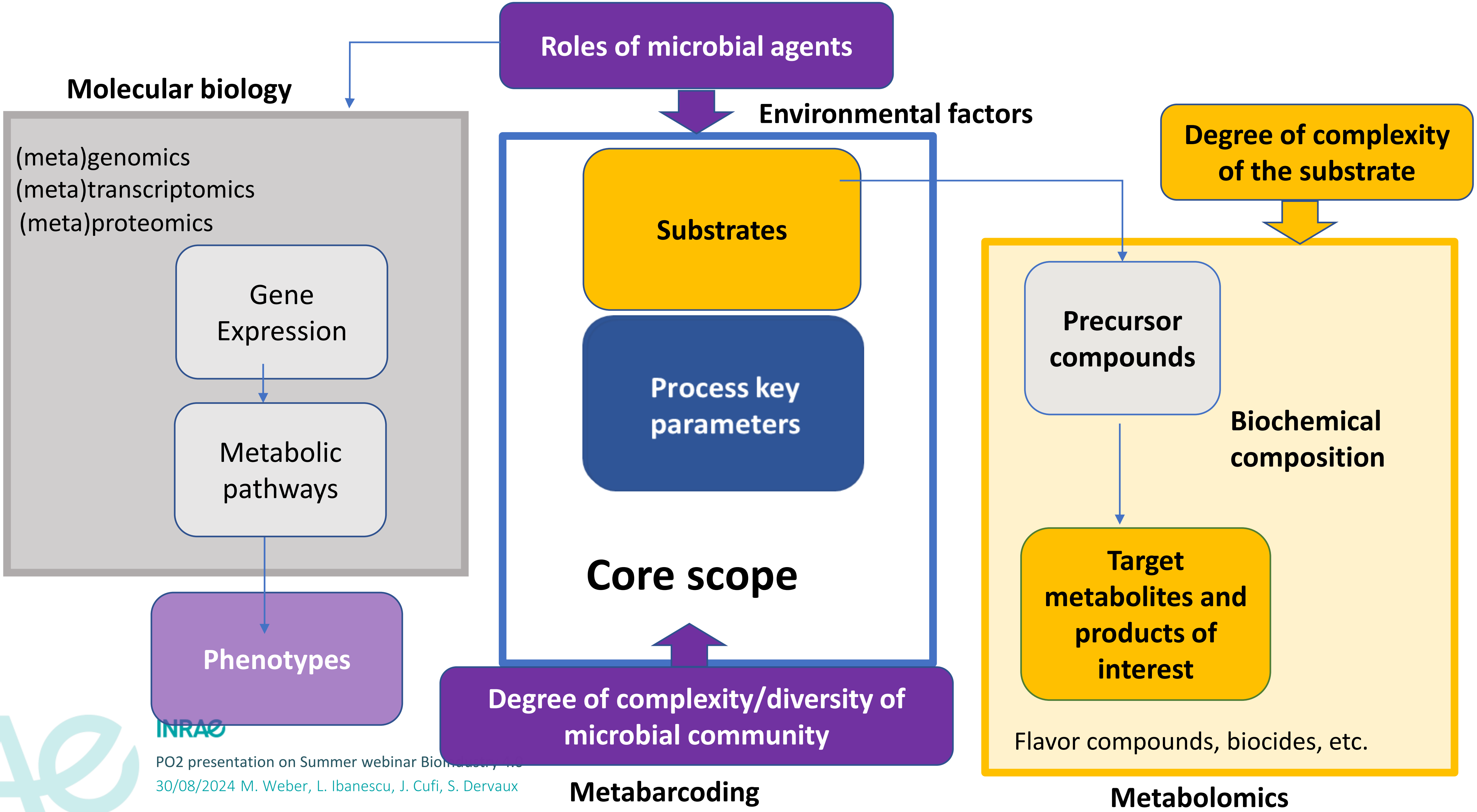
Use cases in FermentON

Food fermentation	Wine making	Composite making	Environmental biorefinery & biotech
(i) cheese, (ii) mixed dairy and plant-based yogurts, and (iii) fermented vegetables	A continuous line from vine to wine & a multiscale study of alcoholic fermentation	Production of PHA/PHBV for composite packaging materials	(i) Electrofermentation (ii) Design of soft sensors

Fermented food & Beverage

Bioproducts, matter & energy

Scope to be covered

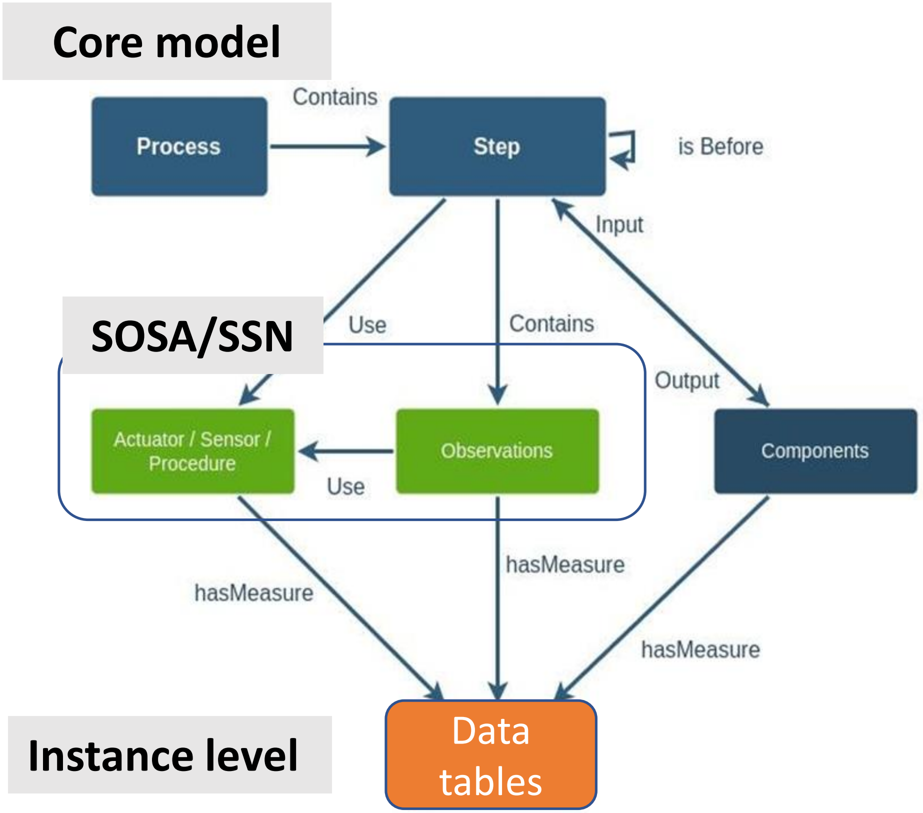


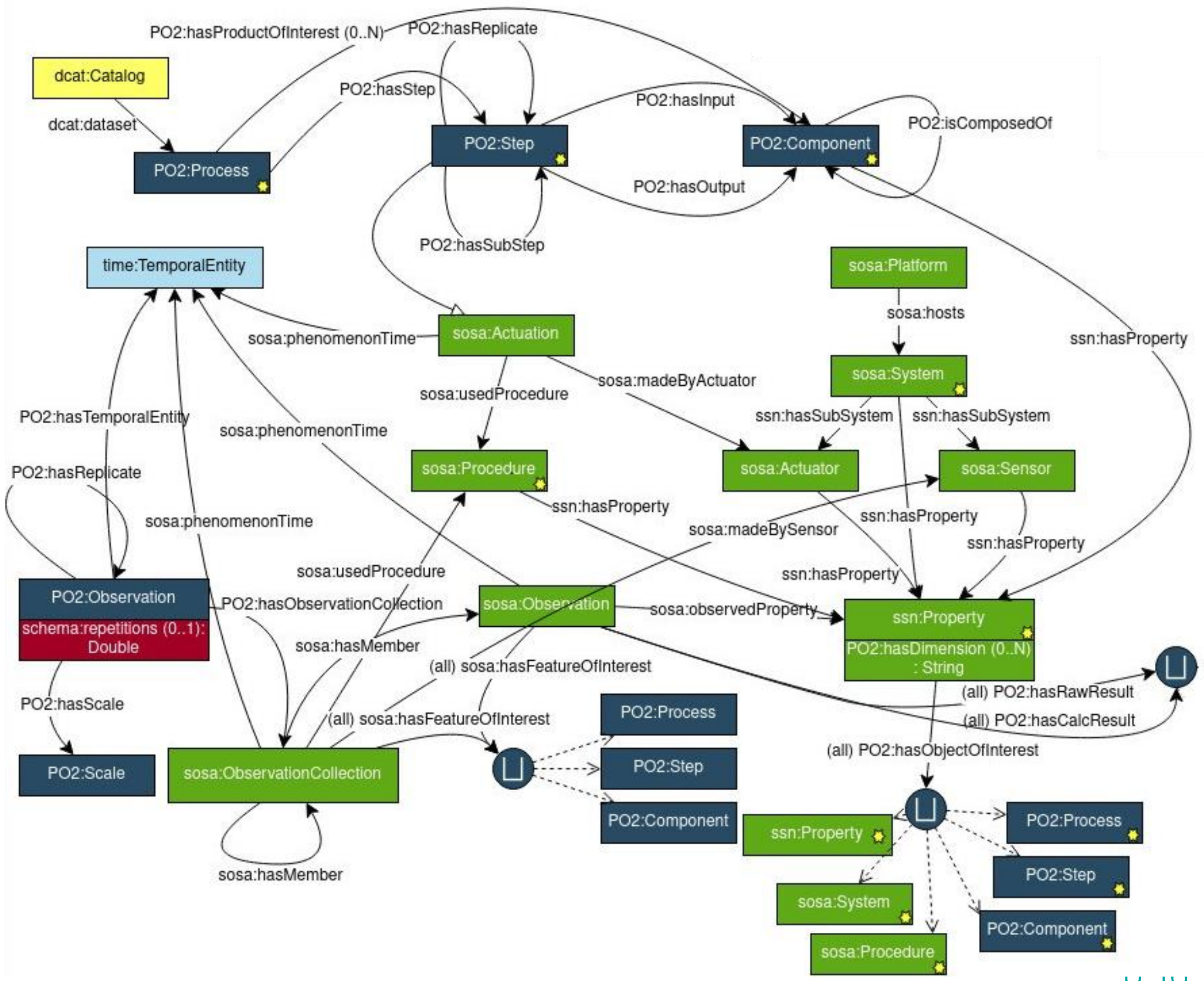
How to produce FAIR by-design and AI-ready data in practice ?

Deep dive into the ecosystem !

How to produce FAIR by-design and AI-ready data in practice ?

Simplified ontology representation VS reality …





How to produce FAIR by-design and AI-ready data in practice ?

Producing FAIR data requires Ontology driven Information System

These Information System address main needs like

- Data Entry
- Data Querying

… and relies on a Software Architecture who ensure

- Security and Access Management
- Interoperability

Data entry

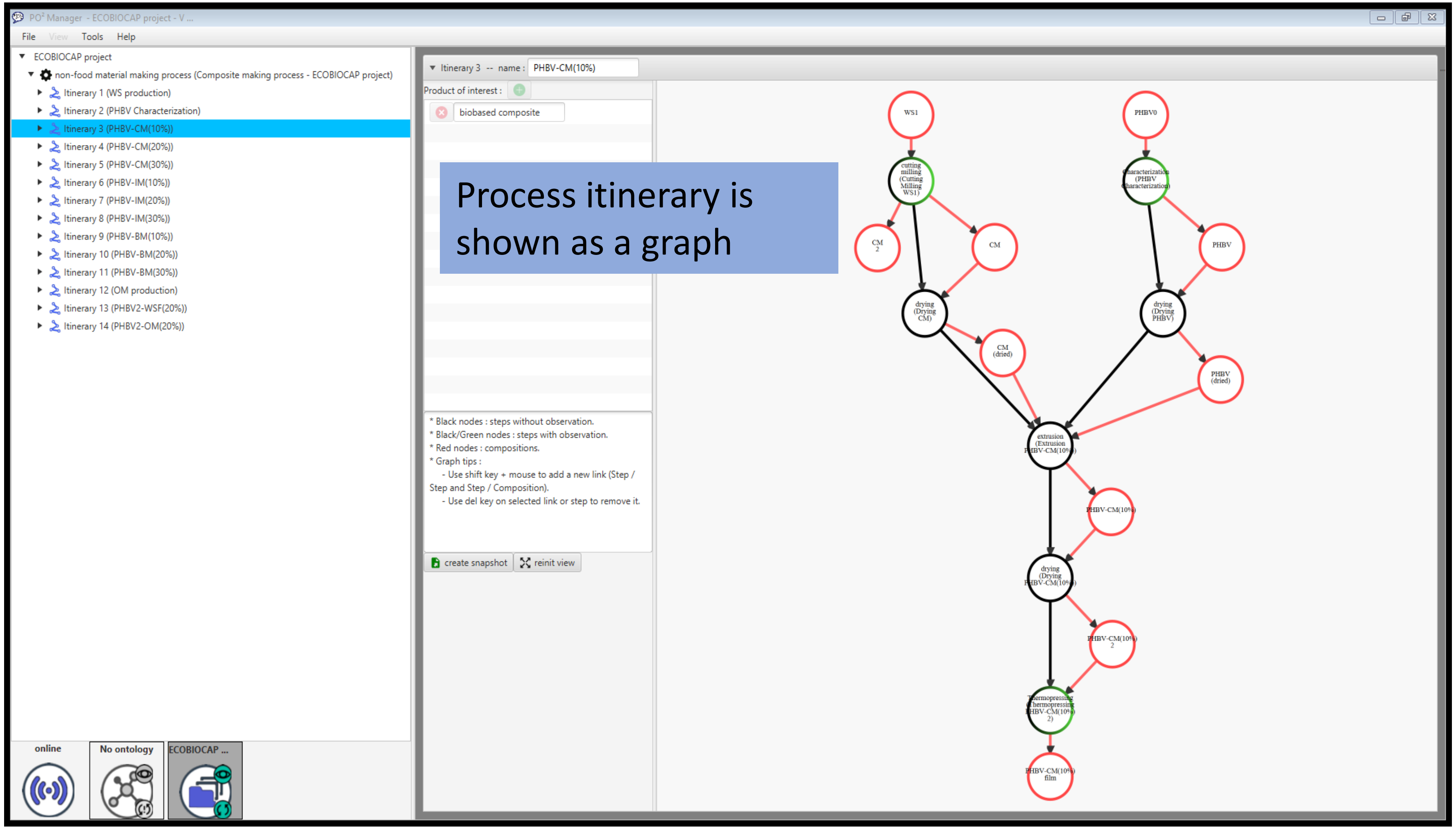
Desktop Application PO² Manager

- Main features
  - Data entry with visualization directly through the application
  - Import/Export of Excel template files
  - Data quality checks
    - Terminological consistency checks
    - Data checks on allowed value range (including dimensions for SI units)
  - Semantization process *i.e* remote publication to save and share data
- Benefits
  - Ontologies are stored on a centralized database

    → **One shared vision applied to ≠ context**
  - Data understanding through visualisation of the studied process
  - Standardized vocabulary used but no « Semantics prerequisites » required
  - Linked with international W3C standards
    - (e.g., Semantic Sensor Network Ontology, Time Ontology, ...)

Data entry

PO² Manager User Interface



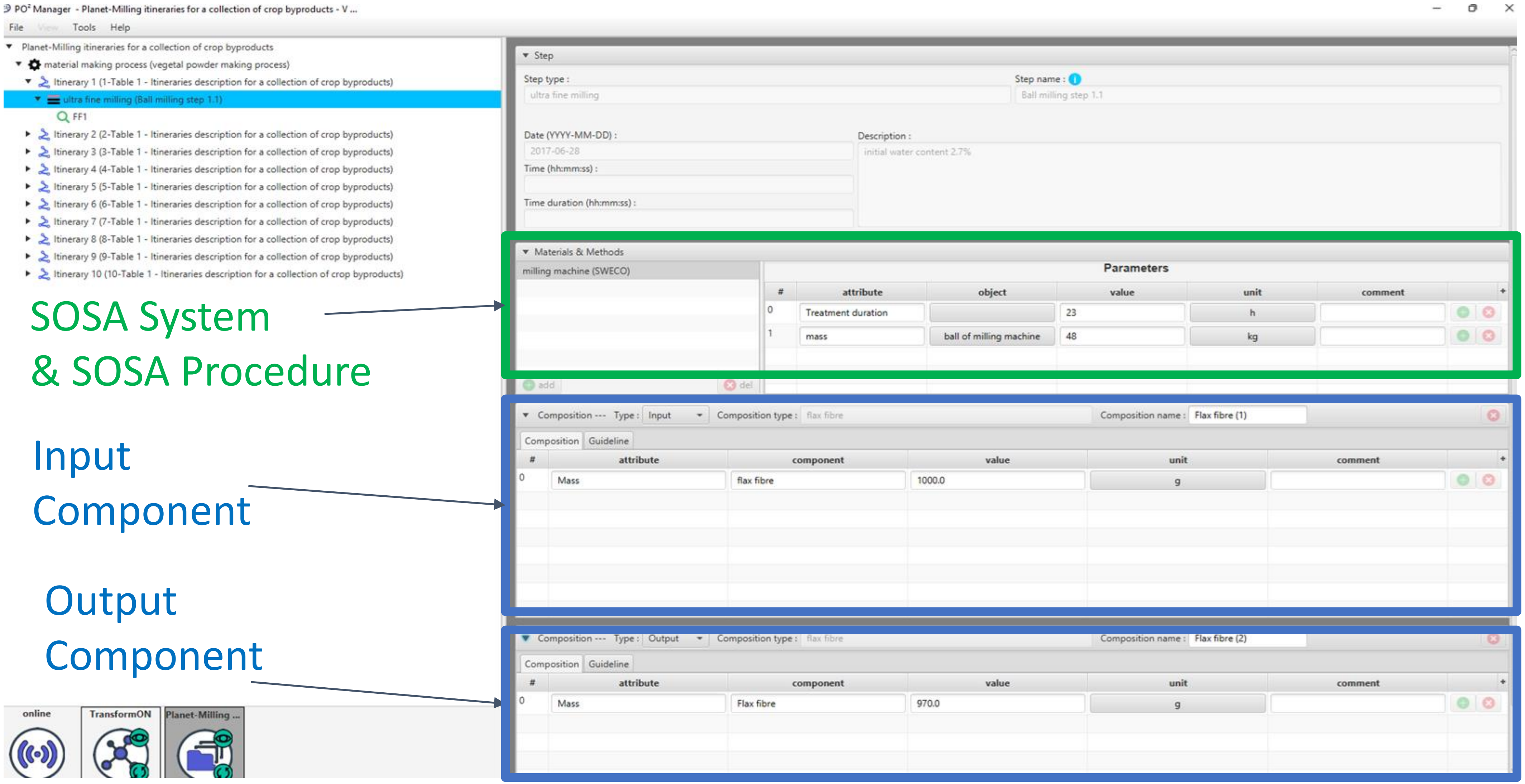
Red Circle : Component

Black Circle : Step

Black + Green: Observations

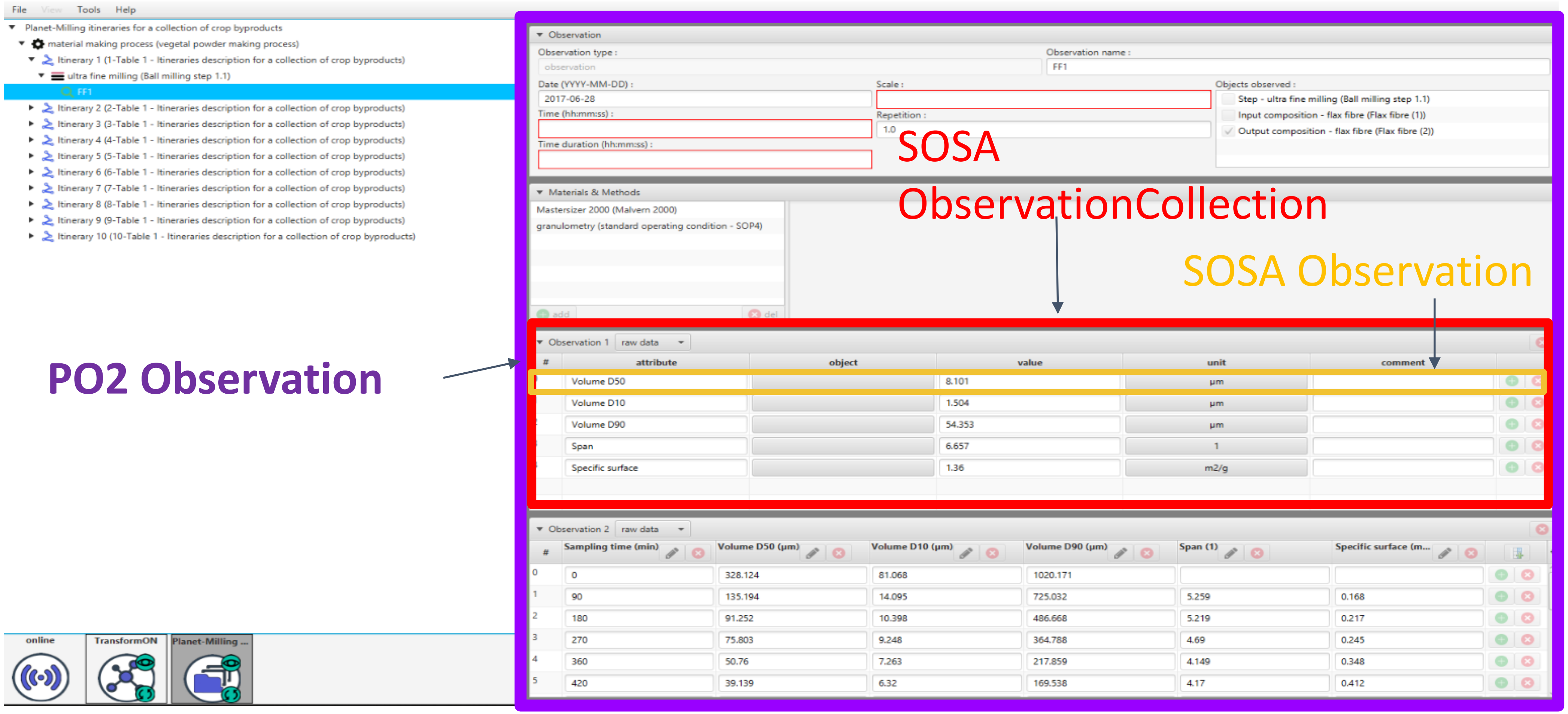
Data entry

PO² Manager User Interface : link with underlying ontologies



Data entry

PO² Manager User Interface : link with underlying ontologies



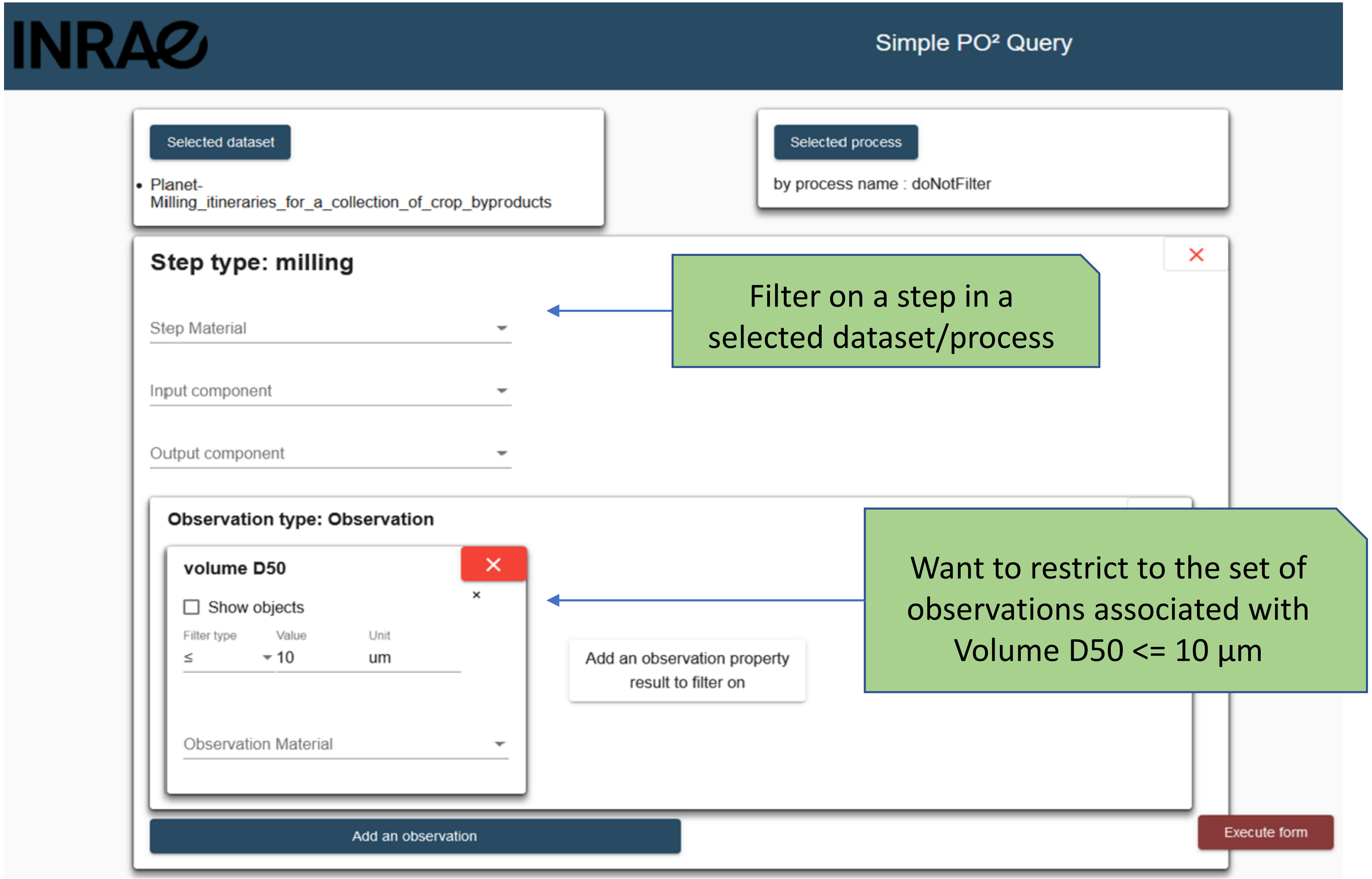
Data query

Web application SPO²Q

- Main Features
  - Search engine with two modes :
    - **Simplified** for standard data users via classic web forms
      - No « technical » knowledge required
    - **Advanced** for « power users »
      - Assisted complex query writing
      - Fine-tuning of the SPARQL query
  - Automatic conversion : querying data with ≠ units of measure
  - Save most frequent queries for re-execution
  - Query result export (CSV, JSON, …)
- Benefits
  - Controlled vocabularies used to query data
  - Query public and private project's data
  - Cross querying between opened data sources

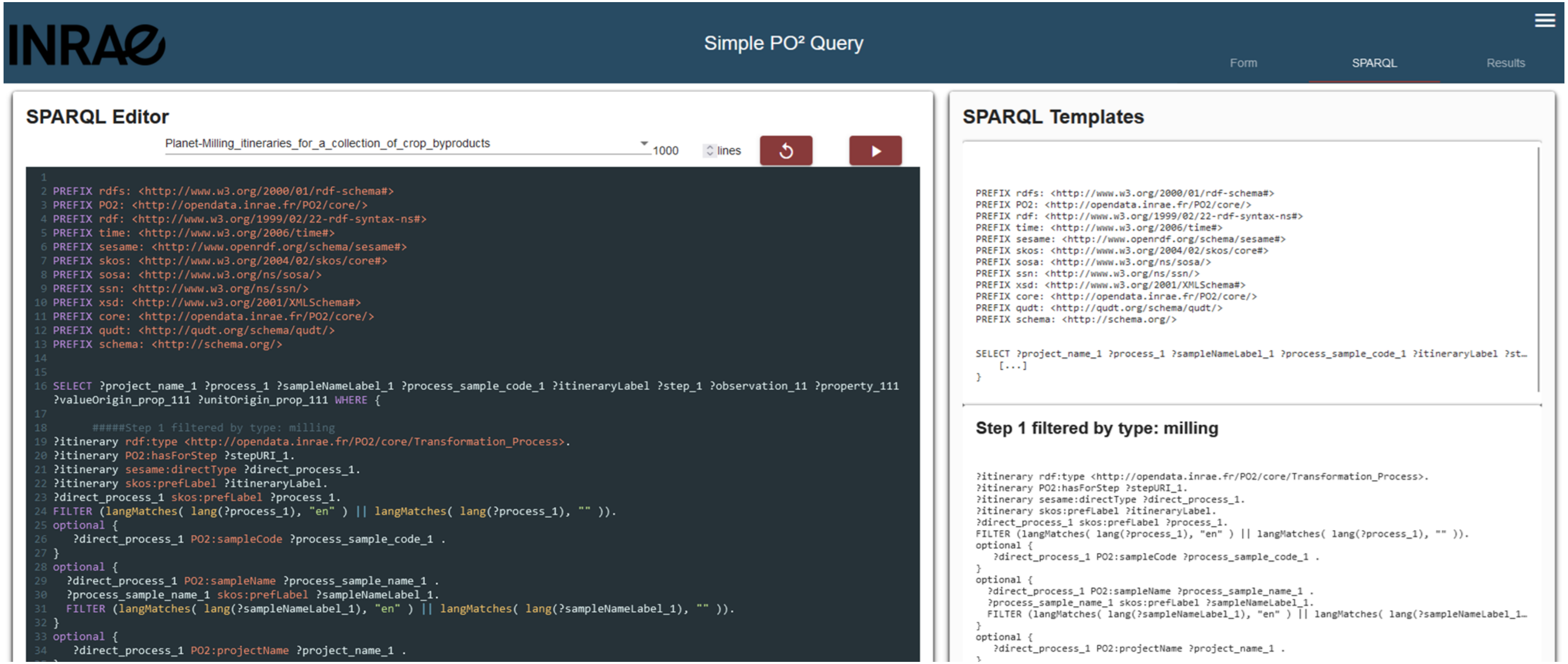
Data query

SPO²Q User interface : Simple mode



Data query

SPO²Q User interface : Advanced mode

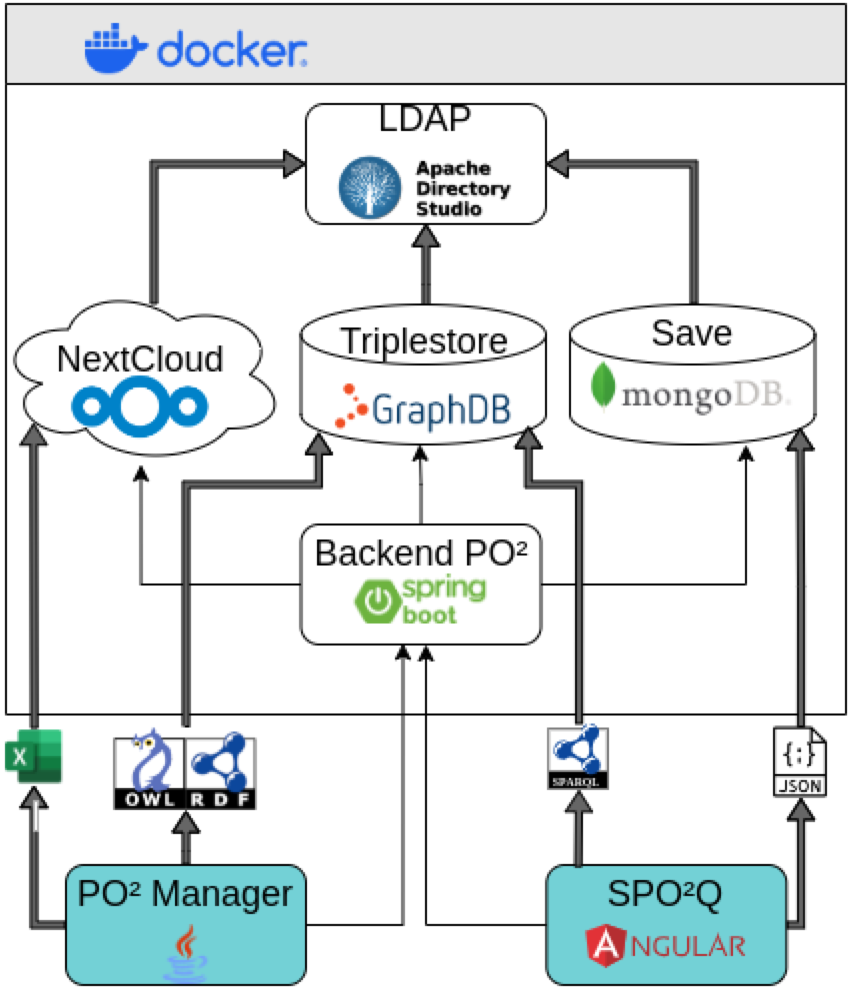


Security and Access Management

- **Data is FAIR but not necessarily immediately open**

→ Possibility of private project creation

- Account management internal LDAP shared with PO2 Manager and SPO²Q
    - One single account per user
    - Flexibility in account creation (not necessarily institutional)
- Management of read/write access by users
- Querying private and public data



What's next ?

- Collaborative ontology construction
  - → Share and improve domain ontology by integrating scientific point of views
  - → Workflow construction based on Git/Gitlab
- Develop and enforce AI applications
  - → Improve quality checks and user's feedback
- Large dataset management
  - → Be able to perform semantic queries on large dataset (Ex : Long times series)
    - Proof of Concept « Best of two worlds » in progress
      - BigData technology stack for performance
      - Semantic technology stack for knowledge representation

      → SANSA https://sansa-stack.net/
- Interface improvement (UX design)

In a nutshell

Summary

Data of interest

STRUCTURING data

STANDARDISING metadata

QUERYING &
LINKING data (W3C)

AI-READY Data

Re-USE

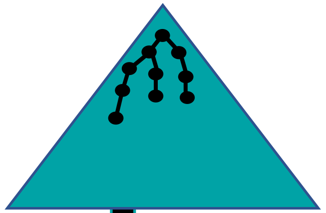
SHARE (PUBLISH)

Open data publication option

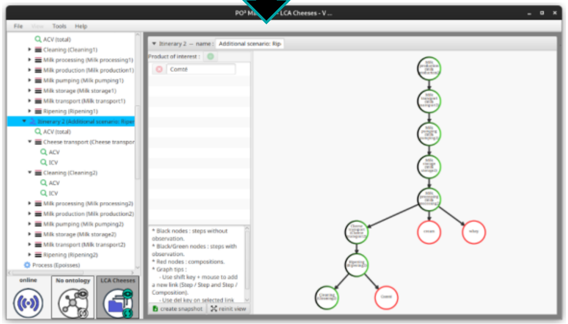
STATISTICAL MODELLING & MACHINE LEARNING

KNOWLEDGE DISCOVERY

PO2 core and domain ontologies



PO2 Manager



SPO2Q

Graph DB

FAIR by-design Data

PO2/TransformON : An ontology for modelling biomass transformation process and characterization of food, feed, bioproducts & biowaste

npj | science of food

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › npj science of food › articles › article

Article | Open Access | Published: 04 September 2023

PO2/TransformON, an ontology for data integration on food, feed, bioproducts and biowaste engineering

Magalie Weber, Patrice Buche, Liliana Ibanescu, Stéphane Dervaux, Hervé Guillemin, Julien Cufi, Michel Visalli, Elisabeth Guichard & Caroline Pénicaud

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE
Liberté Égalité Fraternité

Ouvrir la science !

2023 Open Science Research Data Award from the French Minister for Higher Education, Research and Innovation

INRAE
science for people, life & earth

TRANSFORM

MIA PARIS-SACLAY

CATI DIISCICO

Documentation and support:
https://quantum.mia-ps.inrae.fr/PO2/

Core team for IT development &knowledge engineering

Magalie Weber
IR INRAE
BIA Nantes

BIA
Biopolymères
Interactions
Assemblages

Julien Cufi
IR INRAE
Dpt Transform

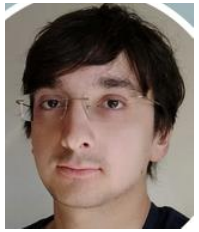
Liliana Ibanescu
MC AgroParisTech
MIA Paris-Saclay

Stéphane Dervaux
IE INRAE
MIA Paris-Saclay

Mélanie Münch
IR INRAE
STLO Rennes

STLO

UMR iAte
ingénierie des Agropolymères & technologies émergentes



Alrick Oudot
IE INRAE
IATE Montpellier

Patrice Buche
IR HDR NRAE
IATE Montpellier

Luc Menut
AI INRAE
IATE Montpellier

Hervé Guillemin
IR INRAE
PAM Dijon/Poligny

& PLASTIC Platform

PAM
Food and wine sciences & Technology

Softwares

- Dervaux et al. 2023: PO2 Manager, an annotation tool to edit biomass transformation and characterization itineraries using the Process and Observation (PO2) Ontology. ⟨swh:1:dir:b9983f579ca087d946a9c3e4818e4507de395937⟩. ⟨hal-04313202⟩
- Dervaux et al. 2024: PO2 Engine, an application programming interface associated with the Process and Observation (PO2) software ecosystem. ⟨swh:1:dir:076aa8ecc70160ac3c9c333570c8d495d7e2ff19⟩. ⟨hal-04387669⟩
- Dervaux et al. 2024: Simple PO2 Query (SPO2Q), a querying tool to retrieve biomass transformation and characterization itineraries using the Process and Observation (PO2) Ontology. ⟨swh:1:dir:360ce0db73073210c9e95ee9ee44cb50fc31d9bd⟩. ⟨hal-04501660⟩

Reused standards and ontologies

- SOSA/SSN https://www.w3.org/TR/vocab-ssn/
- BFO https://obofoundry.org/ontology/bfo.html
- IAO https://obofoundry.org/ontology/iao.html
- Time Ontology https://www.w3.org/TR/vocab-owl-time-rel/
- QUDT https://qudt.org/
- UCUM https://ucum.org/
- Schema.org https://schema.org/
- DCAT https://www.w3.org/TR/vocab-dcat/
- I-ADOPT https://i-adopt.github.io/ontology/

Bibliographic references

- 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 [Communication MTSR2016]. ⟨10.1007/978-3-319-49157-8_13⟩. ⟨hal-01418097⟩
- Pénicaud et al. 2019: Relating transformation process, eco-design, composition and sensory quality in cheeses using PO2 ontology. *International Dairy Journal*, 92, pp.1-10. ⟨10.1016/j.idairyj.2019.01.003⟩. ⟨hal-02043246⟩
- 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*, 2022, 13 (1), ⟨10.4018/IJAEIS.309136⟩. ⟨hal-03738973⟩
- Auberger et al. 2022 Adapting MEANS-InOut LCA software to food engineering, in relation to the PO² food ontology and PO²-BaGaTel food engineering database. *13th International Conference on Life Cycle Assessment of Food 2022 (LCA Foods 2022)*, Oct 2022, Lima, Peru. ⟨hal-03837121⟩
- Münch et al. 2022a: Combining ontology and probabilistic models for the design of bio-based product transformation processes. *Expert Systems with Applications*, Elsevier, 203, pp.117406. ⟨10.1016/j.eswa.2022.117406⟩. ⟨hal-03662183⟩
- 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*, 16 (1), pp.1-15. ⟨10.1504/IJMSO.2022.131129⟩. ⟨hal-04115029⟩
- Münch et al. 2022c: Biocomposites from poly(3-hydroxybutyrate-co-3-hydroxyvalerate) and lignocellulosic fillers: Processes stored in data warehouse structured by an ontology. *Data in Brief*, 42, pp.108191. ⟨10.1016/j.dib.2022.108191⟩. ⟨hal-03650668v2⟩
- Weber et al. 2023: PO2/TransformON, an ontology for data integration on food, feed, bioproducts and biowaste engineering. *npj Science of Food*, 2023, 7, pp.47. ⟨10.1038/s41538-023-00221-2⟩. ⟨hal-04197618⟩