

# Alignement et enrichissement d'ontologies utilisées par deux systèmes d'informations gérant des données expérimentales sur les procédés de transformation :

## OpenSILEX-EnviBIS et PO2 Bagatel

Emilie Fernandez<sup>1</sup>, Patrice Buche<sup>4</sup>, Julien Cufi<sup>4</sup>, Stephane Dervaux<sup>3</sup>, Liliana Ibanescu<sup>3</sup>, Eric Latrille<sup>1</sup>, Pascal Neveu<sup>2</sup>,  
Virginie Rossard<sup>1</sup>, Anne Tireau<sup>2</sup>, Magalie Weber<sup>5</sup>

<sup>1</sup> Laboratoire de Biotechnologie de l'Environnement, LBE, Avenue des Étangs, Narbonne

<sup>2</sup> Mathématiques, Informatique et STatistique pour l'Environnement et l'Agronomie, MISTEA, 2 place Pierre Viala, Montpellier

<sup>3</sup> Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA Paris-Saclay, 91120 Palaiseau

<sup>4</sup> Ingénierie des Agropolymères et technologie émergente, IATE, Univ. Montpellier, INRAE, SupAgro, Montpellier

<sup>5</sup> Biopolymères Interactions Assemblages, BIA, INRAE, Nantes

### Introduction

Dans le domaine de la bioraffinerie environnementale, il existe actuellement à INRAE beaucoup de données expérimentales : une partie de ces données est structurée selon l'ontologie EBO (Environmental Biorefinery Ontology) qui se trouve au coeur du système d'information EnviBIS, qui s'appuie sur la suite logicielle libre OpenSILEX [1]. Dans le domaine de la transformation d'agro-ressources, une partie des données expérimentales est structurée selon l'ontologie noyau PO2 (Process and Observation Ontology) [2,3,4] qui a été spécialisée dans plusieurs domaines : PO2DairyGels pour la fabrication des gels laitiers [5,6,7], PO2MeatyLab pour la fabrication des knacks et PO2Bioraf pour le domaine de la bioraffinerie [4,8,9]. Les évolutions récentes d'EBO pour représenter des données expérimentales dans des itinéraires de transformation ont amené les auteurs à comparer leurs approches.

Dans un contexte mondial de science ouverte, ces projets ont pour but de développer et mettre en œuvre un modèle de structuration des données et un vocabulaire commun et apporter aux scientifiques un ensemble d'outils informatiques pour gérer leurs données tout le long du cycle de leur vie afin qu'elles puissent répondre aux critères FAIR.

Le travail de comparaison des deux ontologies évoquées ci-dessus, EBO et PO2, nous a conduit à identifier les concepts en commun et ceux qui devront enrichir chacun des modèles pour une exploitation uniforme des données, en utilisant les standards et technologies du Web sémantique. Nous présentons en premier l'ontologie EBO et son environnement, l'ontologie PO2 et son environnement, puis nous comparons les modèles et donnons des pistes pour leurs évolutions.

### L'ontologie EBO et le système OpenSILEX et EnviBIS

La gestion des données en ligne et le suivi des bioprocédés a amené l'unité INRAE LBE, Laboratoire de Biotechnologie de l'Environnement, hébergeant l'infrastructure scientifique collective Bio2E, à mettre en place, depuis les années 2000, un système d'information qui assure la réutilisation et le partage des données issues de la recherche. Cela a conduit à élaborer une ontologie des connaissances dans le domaine de la bioraffinerie environnementale, nommée EBO pour Environmental Biorefinery Ontology, récemment mise en œuvre par le système d'information

EnviBIS, qui s'appuie sur la suite logicielle libre OpenSILEX [1] portée par l'unité MISTEA. OpenSILEX est une suite logicielle qui permet de construire des systèmes d'information et d'implémenter les principes FAIR [10]. Il est dérivé pour des communautés scientifiques spécifiques (phénotypage haut-débit des plantes, viticulture et oenologie, dépollution, etc.) dont EnviBIS. OpenSILEX est pilotée par des ensembles d'ontologies. L'ontologie coeur d'OpenSILEX s'appuie sur des ontologies de hauts niveaux (e.g. Dublin Core [11], FOAF [12], SOSA [13], PROV-O [14]) et permet d'être étendu par des ontologies de domaine comme l'Environmental Biorefinery Ontology. EBO s'appuie sur des thésaurus spécialisés dans la description des procédés de transformation biologiques et physiques et de la matière organique. Dans l'état actuel, les termes des thésaurus publiés représentent 60% de l'ensemble des termes nécessaires à la description des matières (composants).

La recherche dans le domaine de la bioraffinerie environnementale au LBE s'appuie sur trois objets d'études : procédé, matière et écosystème microbien qui décrivent les thèmes de recherche tels que la digestion anaérobie, les pré-traitements, la bio-méthanisation, le digestat, le biohydrogène, les écosystèmes microbiens, les photogranules, la bio-électrochimie et les microalgues [19]. Ces objets d'études sont reliés par des liens sémantiques, par exemple une matière est une entrée d'une étape, et un procédé de transformation est composé d'étapes.

EBO est développée à l'aide des standards du web sémantique en utilisant RDF, RDFS et OWL et elle s'appuie sur des ontologies de références [11–14] intégrées à l'outil EnviBIS. Ce système permet l'acquisition de données hétérogènes issues de capteurs, d'observations ou d'appareils analytiques. Enfin, ces données sont rattachées à un modèle de variables comportant cinq éléments : une entité, une entité de contexte, une caractéristique, une méthode et une unité.

### **L'ontologie PO2 et son écosystème**

L'ontologie noyau PO2 (Process and Observation Ontology) [2,3,4], spécialisée dans plusieurs ontologies de domaines, permet de représenter et stocker de manière uniforme des données expérimentales de procédés issues des projets de recherche de plusieurs unités de INRAE [4-9].

L'ontologie noyau PO2 réutilise des concepts des ontologies de haut niveau (BFO [16], SOSA [13], IAO [17] et TimeOntology [18]) et propose des concepts génériques reliés par des relations sémantiques et dont un sous-ensemble sont spécialisables pour chaque ontologie de domaine. Les sept concepts qui peuvent être spécialisés sont : Component, Step, Attribute, Process, Material, Method, Scale.

PO2 BaGaTel [5] est une base de données au format RDF structurée selon le modèle de l'ontologie noyau PO2 et intégrant les données sur les procédés, comme par exemple les impacts environnementaux de certains procédés, les qualités sensorielle et nutritionnelle de différents aliments ou les caractéristiques physicochimiques de différentes biomasses étudiées dans des projets de recherche du département Transform de INRAE.

Deux logiciels [3,15] sont développés pour assister les experts à utiliser le modèle PO2:

1. PO2 Manager est un outil qui permet de construire et de visualiser une ontologie de domaine qui spécialise le noyau PO2. PO2 Manager aide aussi à créer et à visualiser un jeu de données structuré avec cette ontologie de domaine et de sauvegarder ce jeu de données dans la base de données PO2 BaGaTel.
2. SPO2Q (Simple PO2 Query) est un outil d'interrogation permettant de créer à l'aide de formulaires des requêtes SPARQL et de les lancer sur la base RDF PO2 BaGaTel. Ces formulaires assistent les experts à l'usage de SPARQL.

L'ontologie noyau PO2 est actuellement en cours d'être spécialisée pour obtenir l'ontologie de domaine PO2-TransformON qui rassemblera le vocabulaire nécessaire à la description de procédés

de transformation et la caractérisation de la biomasse ou des produits qui en sont issus (aliment, bioproduits, déchets). Ce vocabulaire contrôlé et unifié pour l'ensemble des domaines de recherche du département Transform a pour objectifs d'assurer une interopérabilité des données de la base PO2 BaGaTel pour réaliser des connexions entre les différents projets.

### **Comparaison des modèles et leurs évolutions**

Suite à la comparaison des modèles EBO et PO2 et l'analyse détaillée des besoins d'intégration des données de procédés de transformation, plusieurs évolutions des modèles ont été identifiées et sont en cours d'implémentation.

D'une part, l'ontologie EBO réutilise une partie des concepts de PO2, et notamment les concepts de Process, Step et Component afin d'offrir une représentation des enchaînements d'opérations au sein d'un procédé. Les discussions menées au sein du groupe de travail PO2/EBO ont également permis d'introduire dans EBO la notion de rôle au niveau des composants d'entrées et sorties d'étapes de procédé. Cela permettrait de spécifier les matières et les énergies dans les réactions biochimiques mises en œuvre dans les bioprocédés.

D'autre part, une nouvelle description des variables dans EBO est en cours d'étude, en s'inspirant de l'ontologie I-ADOPT [20] proposée par le groupe de travail d'une organisation communautaire mondiale sur le partage des données de recherche, Research Data Alliance, pour se rapprocher de l'ontologie SOSA [13] sur les capteurs et actionneurs.

Par ailleurs, les vocabulaires utilisés dans EBO et PO2 décrivent le domaine des procédés de transformation avec des périmètres qui ne sont pas exactement les mêmes mais qui ont en partie les mêmes objets d'études. Nous avons identifié une partie des termes/concepts communs et le travail en cours est de préciser quels sont ceux qui vont enrichir chacun des modèles.

### **Conclusion et perspectives**

En conclusion, l'alignement entre EBO et PO2 facilitera, à court terme, les échanges d'information entre les deux systèmes et permettra à plus long terme à la communauté scientifique du domaine, de publier, de partager et d'échanger des données à l'aide des outils du web sémantique. L'utilisation d'ontologies de référence permettra l'exploitation des données de bioraffinerie environnementale comme celle de l'unité INRAE LBE et plus généralement de données expérimentales de procédés de transformation comme celles du département INRAE Transform.

### **Bibliographie**

[1] OpenSILEX: <https://github.com/OpenSILEX/>

[2] J. Dibie, S. Dervaux, L. Ibanescu, J. Raad (2018) . "Process and Observation Ontology", [10.15454/1.4702114192525708E12](https://doi.org/10.15454/1.4702114192525708E12), Portail Data INRAE, V4

[3] "Process and Observation Ontology" <http://quantum.agroparistech.fr/PO2/>

[4] M. Munch, P. Buche, S. Dervaux, J. Dibie, L. Ibanescu, et al.. Combining ontology and probabilistic models for the design of bio-based product transformation processes. *Expert Systems with Applications*, Elsevier, 2022, pp.117406. [10.1016/j.eswa.2022.117406](https://doi.org/10.1016/j.eswa.2022.117406). [hal-03662183](https://hal.archives-ouvertes.fr/hal-03662183)

[5] C. Pénicaut, L. Ibanescu, T. Allard, F. Fonseca, S. Dervaux, B. Perret, H. Guillemin, S. Buchin, C. Salles, J. Dibie, E. Guichard (2019). Relating transformation process, eco-design, composition and sensory quality in cheeses using PO2 ontology. In *Int. Dairy Journal*. Volume 92, pp. pp. 1–10. [10.1016/j.idairyj.2019.01.003](https://doi.org/10.1016/j.idairyj.2019.01.003). [hal-02043246](https://hal.archives-ouvertes.fr/hal-02043246)

- [6] S. Aubin, P. Bisquert, P. Buche, J. Dibie, L. Ibanescu, Clément Jonquet, C. Roussey (2019). Recent progresses in data and knowledge integration for decision support in agri-food chains. Dans Actes de IC 2019 (Journée francophone d'Ingénierie des Connaissances), pp. 43-59. [<hal-02609466>](#)
- [7] L. Ibanescu, T. Allard, S. Dervaux, J. Dibie, E. Guichard, C. Pénicaud, J. Raad (2018). A Use Case of Data Integration in Food Production. Dans Proceedings of the EKAW 2018 Posters and Demonstrations Session co-located with 21st International Conference on Knowledge Engineering and Knowledge Management (EKAW 2018), volume 2262, pp. 17-20, CEUR-WS.org, 2018. [<ceur-ws.org/Vol-2262/ekaw-poster-11.pdf>](#)
- [8] C. Fabre, P. Buche, X. Rouau, C. Mayer-Laigle. Milling itineraries dataset for a collection of crop and wood by-products and granulometric properties of the resulting powders. Data in Brief, Elsevier, 2020, 33, pp.106430. [\(10.1016/j.dib.2020.106430\)](#). [\(hal-03004903\)](#)
- [9] M. Munch, P. Buche, S. Dervaux, A. Breysse, M-A Berthet, et al.. Biocomposites from poly(3-hydroxybutyrate-co-3-hydroxyvalerate) and lignocellulosic fillers: Processes stored in data warehouse structured by an ontology. *Data in Brief*, Elsevier, 2022, pp.108191. [\(10.1016/j.dib.2022.108191\)](#). [\(hal-03650668\)](#)
- [10] Le Centre pour la Communication Scientifique Directe: principes FAIR <https://www.ccsd.cnrs.fr/principes-fair/>
- [11] Dublin Core <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>
- [12] FOAF (Friend of a Friend Ontology) <http://xmlns.com/foaf/spec/>
- [13] SOSA (Semantic Sensor Network Ontology) <https://www.w3.org/TR/vocab-ssn/>
- [14] PROV-O (The PROV Ontology) <https://www.w3.org/TR/prov-o/>
- [15] S. Dervaux, J. Dibie and L. Ibanescu (2018). PO2VocabularyManager - A Collaborative Tool to Assist Users in Building a PO2 Domain Ontology Linked with Existing Resources. In SemFAEN 2018 : Semantic for Food, Agriculture, Environment and Nutrition.
- [16] BFO (Basic Formal Ontology) <https://basic-formal-ontology.org/>
- [17] IAO (Information Artifact Ontology) <https://www.ebi.ac.uk/ols/ontologies/iao>
- [18] Time Ontology <https://www.w3.org/TR/owl-time/>
- [19] Thèmes de recherche de LBE <https://www6.montpellier.inrae.fr/narbonne/Themes-de-recherche>
- [20] I-ADOPT (Interoperable Descriptions of Observable Property Terminologies) <https://www.rd-alliance.org/group/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg/wiki/i-adopt-0>