

Ontological Representation of Cultivated Plants: Linking Botanical and Agricultural Usages*

Baptiste Darnala^{1,2}, Florence Amardeilh², Catherine Roussey³, Konstantin Todorov¹ and Clément Jonquet^{1,4}

¹LIRMM, University of Montpellier, CNRS, Montpellier, France

²Elzeard, Cité du Numérique, Bègles, France

³Université Clermont Auvergne, INRAE, UR TSCF, F-63000 Clermont-Ferrand, France.

⁴MISTEA, University of Montpellier, INRAE, Institut Agro, Montpellier, France

Abstract

Cultivated plants may be described from various viewpoints: botanical, agronomic, agricultural and more. These viewpoints often result into different specific formal representations (i.e., ontologies). Linking concepts describing these different viewpoints is difficult and demands domain expertise. Still it is necessary as it supports the agricultural planning processes. In our case, there exists no standard knowledge pattern to represent alignments between thesauri describing cultivated plants in agriculture and organism taxonomies or classifications; in addition, basic ontology mapping properties (e.g., from SKOS) are not sufficient. We have conceived the Crop Planning and Production Process Ontology (C3PO) to describe agricultural knowledge for diversified crop production. In this paper, we describe the ontological representation of the *Plant module* of C3PO, which addresses the aforementioned linking needs. It integrates crop usage information about cultivated plants from the *French Crop Usage* thesaurus and botanical –classification and nomenclature– information from the TaxRef taxonomy. This Plant module is valued in two systems both developed by Elzeard—a french SME: (i) a web application to support farmers in crop planting activity; and (ii) a web portal, *La Serre des Savoirs* (under development), which will share general agricultural information about crops.. The C3PO ontology and its knowledge graph are publicly available at <https://gitlab.com/serre-des-savoirs/c3po>.

Keywords

knowledge graphs, ontologies, ontology model, knowledge integration, agriculture, botanic taxonomy

1. Introduction

A plant is a complex system studied and observed by different experts (biologists, agronomists, botanists, farmers) who each use specific characteristics to describe it. Each of these viewpoints are captured into several ontologies or knowledge graphs (KG) such as: (i) TAXREF-LD [7], which represents –as linked data– the national repository on fauna and flora of metropolitan

INTERNATIONAL WORKSHOP ON KNOWLEDGE GRAPH GENERATION FROM TEXT (TEXT2KG 2022) and MODULAR KNOWLEDGE (2022), ESWC 2022, Hersonissos, Greece, May 29, 2022

✉ baptiste.darnala@elzeard.co (B. Darnala); florence.amardeilh@elzeard.co (F. Amardeilh); catherine.roussey@inrae.fr (C. Roussey); konstantin.todorov@lirmm.fr (K. Todorov); jonquet@lirmm.fr (C. Jonquet)

ORCID 0000-0001-7390-4850 (B. Darnala); 0000-0002-6306-4437 (F. Amardeilh); 0000-0002-3076-5499 (C. Roussey); 0000-0002-9116-6692 (K. Todorov); 0000-0002-2404-1582 (C. Jonquet)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

France and overseas territories; (ii) the French Crop Usage (FCU) [11] which represents a thesaurus of cultivated plant organised by agriculture usages in France (human food, industry, cattle feed).¹ Neither TAXREF-LD, nor FCU include all the plant characteristics into a unique knowledge representation. No semantic resource actually provide a fully integrated and unified view, making alignment/linking between such resources mandatory. Indeed, all this information is important for agriculture when farmers struggle to plan and optimise crop production.

Elzeard (<https://elzeard.co>)—a French SME which develops an application dedicated to farmers (market gardeners and vegetable growers)—has conceived a modular ontology called *Crop Planning and Production Process Ontology* (C3PO) [3]. C3PO represents plot management and crop itineraries from an agricultural perspective. It is the backbone of a web application currently under development called Elzeard application to assist farmers in their planning and production activities and *La Serre des Savoirs*² a web portal under development to publish the culture and the crop itinerary publicly as a wiki portal. The domain of activities being extremely complex, the application requires clear and unambiguous knowledge to support farmers in their planning choices. C3PO's sub-part dedicated to cultivated plant representation, the *Plant module* describes all the knowledge specific to plant, Elzeard has aggregated and linked to agricultural production concepts. The module hierarchizes cultivated plants from a farmer perspective and collects information to describe objects at each different hierarchy level, e.g. plant/crop/family.

Our need is to get and aggregate data—structured as knowledge graphs—from different domains: farmers, agronomic and botanical knowledge. In this paper, we focus on the links between C3PO and TAXREF-LD for the integration of French botanic knowledge, and C3PO and FCU for the integration of the French agricultural usage knowledge. The difficulty is thus to align classes/concepts and individuals from different knowledge graphs, where each graph describes a different viewpoint. For example, farmers use the term 'solanaceous fruits' which somehow was borrowed from the scientific name 'solanaceae'. The related scientific taxon groups the plants tomatoes, potatoes, eggplants and peppers. However, from the farmer point of view, solanaceous never include potatoes because potato farming practices are very different from the three other ones. Our goal is, therefore, to borrow in *La Serre des Savoirs*'s knowledge graph, from scientific taxonomies or referential agricultural usages. Plant groups and families still coping with the concrete differences observed within the fields. We choose to use double typing to construct this alternative hierarchies, to link individuals to external knowledge graphs and to add the specific properties to each instance. As explained later, the elements of the Plant module's hierarchy are both instances of owl:Class and skos:Concept thus can be described with SKOS properties [8] and inherit properties from OWL descriptions. Plus, it permits to link with other knowledge graphs either relying on SKOS or OWL.

In this paper, we present C3PO's Plant module, how it links data from different semantic resources and how it builds a knowledge graph useful for a specific task, such as agricultural production. Section 2 covers related work on plant description models. Section 3 explains the ontological representation of the Plant module of C3PO and its links with TAXREF-LD and FCU. Section 4 describes the data model and instantiation.

¹<https://doi.org/10.15454/QHFTMX>

²'Greenhouse of knowledge', in French.

2. Related Semantic Resources for Plants

C3PO is built with Semantic Web technologies, i.e. it provides the integration of ontologies to model the plants or extract information from open knowledge graphs with facility. Plant module represents vegetable crops and links them to already existing knowledge graphs and ontologies. When building C3PO's Plant module, we searched the term 'plant' in the AgroPortal ontology repository [6] which returned results in multiple ontologies containing plant related knowledge:

- In the first category, we found resources describing plants within a taxonomy of organisms / biological entities: TAXREF-LD, previously cited and the NCBI Taxonomy which describes the standard nomenclature and classification of international organism.³
- In the second category, we found resources describing plants in farming usage: FCU previously cited and the GECO ontology [12], the backbone of an agro-ecological knowledge base to describe new agricultural practices. Plants within GECO being based on a previous version of FCU.
- In the third category, we found resources presenting an experimental or productive viewpoint: FoodOn [4], an exhaustive 'farm-to-fork' ontology about food related knowledge which contains several crop descriptions and some specific cultivars; The Agronomy Ontology [1] an ontology for "representing agronomic practices, techniques, variables and related entities" which contains a representation of 'crop'; FOODIE [9] an ontology which represents a monitoring process of one crop on one area at one moment. The plant is represented by crop species classes. Neither FoodOn, Agronomy Ontology and FOODIE propose a hierarchy of crops.
- In the last category, we found resources describing composition of plants. Plant Ontology (PO) [5] describes plant's characteristics (anatomy, morphology, growth), it is composed of a structured collection of terms that describe structure and developmental stages of a plant. Plant Trait Ontology (TO) [2] is a vocabulary that describes phenotypic traits in plant. These ontologies are used to describe crops by the view of an existing unique plant in the Crop Ontology, a project that describes each crop by a specific ontology.

In these resources, the plant hierarchy is absent or not sufficient for our farming use-case. The requirement is a hierarchy seen by a farmer that integrates a maximum of information that can help him in his production. As plants are a mix of all the information describes above, we need to be able to characterise plants either from a botanical or agronomic point-of-view. To do this, we need a generic model describing what a farmer understands about plants, which links and integrates external resources to obtain the most complete depiction.

3. C3PO's Plant Module Description

C3PO [3] is an ontology composed of several ontology modules, each describing a specific part of a farm and its processes. The design of plots, the administrative organisation, the management of cultivation processes, the description of supplies (input, equipment and plant materials), a

³<https://www.ncbi.nlm.nih.gov/taxonomy>

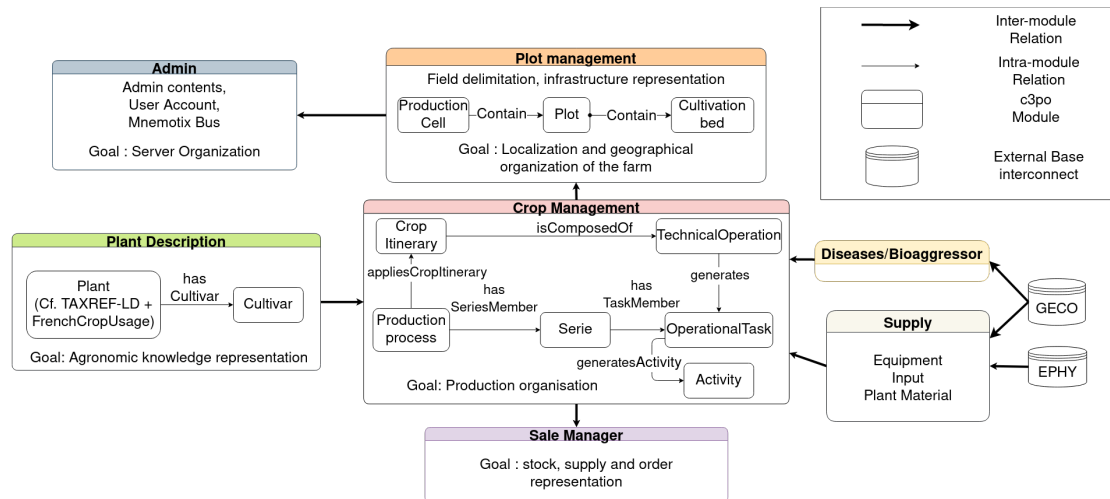


Figure 1: Overview of the Crop Planning and Production Process Ontology

sale manager module and one to describe the plants are the main modules. Figure 1 shows an overview of these ontology modules and their relations. The decomposition of C3PO in several ontology modules allows us to better manage the complexity of the farming domain.

The plant module describes the plant knowledge for C3PO and creates links to external resources to improve the plant representation. We proceed to present the specification of the module and the linking method that we have adopted.

3.1. Specification

In agriculture, according to farmers, a cultivated plant is generally part of a collection, i.e., a cultivated family. Both are defined as:

Cultivated Plant, a type of plant; the type gathers information about how the farmer will cultivate all the plants of this type. Examples of plant types are Carrot or Onion.

Cultivated Family, a set of plant types that are grouped based on some plant type characteristics. The characteristic could be botanical characteristic like the species (*Daucus carota*), or usage characteristic like Leaf vegetable.

3.2. Plant linking

We used the SAMOD [10] ontology development agile methodology. We have described the aforementioned linking knowledge pattern as a *modelelet* i.e. a “A modelelet is a stand-alone model describing a particular domain”. We have focused on linking TAXREF-LD and FCU through C3PO.

Our requirement is a Plant module which represents a plant hierarchy described by farmers and merges information from different resources. We therefore created our instances as pivot

objects reifying the links between the resources to merge, these objects being themselves characterized by properties and classes in their own hierarchy. To do this, we choose to represent C3PO plant instances both as `skos:Concept` and `owl:Class`. SKOS allows the description of the hierarchy with broader / narrower relations plus linking to external objects with mapping properties `skos:*Match`. OWL allows to describe knowledge about things and relations between them. The double typing makes it possible to recover the competences of each one. It also makes the class generic enough on the level of the modelling to be able to connect information coming from external resources which are made in SKOS as in OWL. The hierarchy of `skos:broader/narrower` enables to retrieve any element and provides some information retrieval service.

This model is composed of several hierarchized classes specified in Subsection 3.1: `c3po:BotanicalFamily`, `c3po:UsageFamily` and `c3po:CultivatedPlant`; their instances will be double typed as `skos:Concept` and will be linked to others resources. `CultivatedPlant` instances are linked with `BotanicalFamily` and `UsageFamily` instances by a SKOS broader/narrower relation as `Plant` are part of a family, like `Onion` is part of `Alliaceae` family. SKOS is used to make the plant organisation as `owl:subClassOf` is not made to declare a family and its members.

FCU is formalized in SKOS with objects instances of `fcu:Crop`, a specialization of `skos:Concept`. Therefore, we can link C3PO plant instances and FCU instances with SKOS properties.⁴ The link between C3PO and TAXREF-LD is made with a C3PO property called `c3po:hasScientificName`. C3PO classes are a pivot between the external resources. Figure 2 shows the model.

4. Data model and instantiation

C3PO's knowledge graph contains 118 instances of `c3poplant:CultivatedPlant`, current crops of interest for *La Serre des Savoirs*'s clients. We were then able to link them manually to FCU and TAXREF-LD and involve three agronomy experts to validate them. 74 cultivated plants have both a link to TAXREF-LD and FCU, 6 only to TAXREF-LD, and 36 only to FCU. Figure 3 presents an example of the links for C3PO's instance `Onion_i`.

The individual `c3pokb:Onion-i` borrows its scientific name '`Allium cepa`' from `txrf:name/81339`. Those two individuals are linked by the property `c3poplant:hasScientificName`. The individual `c3pokb:Onion-i` borrows its preferred common name "`Oignon`"@fr from `txrf:taxon/81339/10.0`. Those two individuals are linked by the property chain composed of `c3poplant:hasScientificName` and `txrfp:hasReferenceName`. The individual `c3pokb:Onion-i` borrows its french preferred common name "`Onion`"@en from `fcu:Oignons`. Those two individuals are linked by the property `fcu:hasRelatedCrop` and `skos:exactMatch`. `fcu:hasRelatedCrop` is declared to link with a `fcu:Crop`, declare a `skos:*Match` add more semantic to the link.

At the end, this choice of modelling extends the C3PO's knowledge graph and adds more labels. Moreover, it improves the description by grouping cultivated plants under multiple representations. For example, the onion of C3PO is under family of `Alliaceae` and the link with TAXREF-LD extends the family description because the taxon is under `Amaryllidaceae` family. These two descriptions come from different botanical representations. As an agricultural advice

⁴Later, we will link C3PO instances with other resources such as the NCBI Taxonomy.

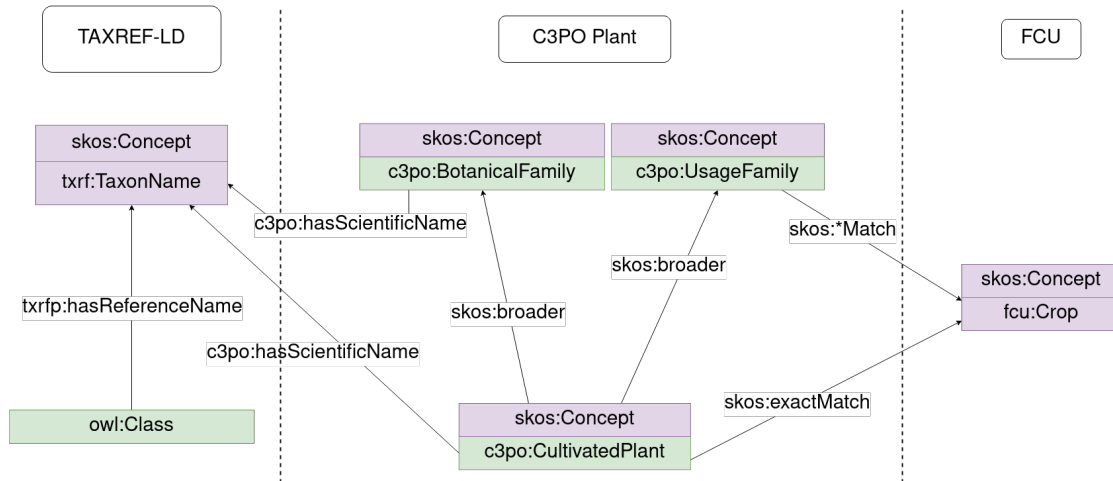


Figure 2: Knowledge pattern (modele) interlinking C3PO, TAXREF-LD and FCU.

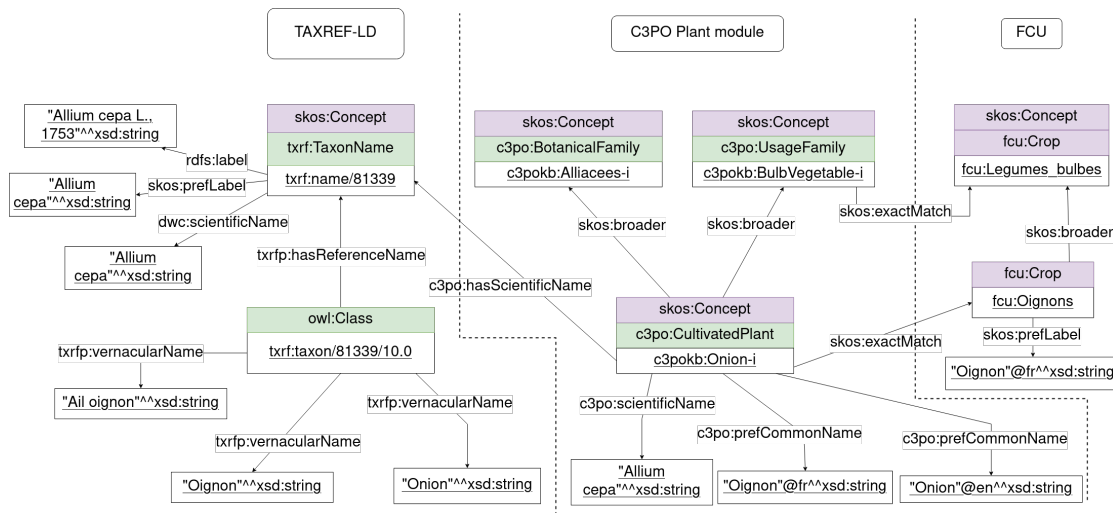


Figure 3: Example of instantiation with onion as cultivated plant.

can be based on different representations, having larger knowledge will help farmers in their research. For example, we consider the following use-cases:

- Inputs management. These are products bought or made by farmers and used for agricultural production, like seeds or fertilizers. For chemical products, there are official regulations, which may depend on the plant but also on the botanical and usage families.
- Manage the crop rotations. Rotation rules are plant or family dependent. It is important to know the membership of Cultivated plant in Cultivated family to infer crop succession. A rotation system will combine this membership and the historical information stored in the crop and plot management modules.

5. Conclusions and Future Work

Linking and integrating data from different resources is a complex process that needs curation. This paper proposes modelling multiple viewpoints inside a single hierarchy using *double typing* with `skos:Concept` and `owl:Class`. This allows to model the hierarchy, and in the meantime align our instances with external resources and inherit class properties. C3PO Plant module's instances are generic enough to be linked to other knowledge graphs such as NCBI or EPPO Global Database.⁵

More generally, other scientific domains (legal, health) could display the problem of connecting data organised by different viewpoints and our methodology using a pivot class and double typing could be considered.

Acknowledgements

We acknowledge support from the National Office for Biodiversity with MesclunDurab grant and the Nouvelle-Aquitaine Region with "Social Innovation AMI" and "Digital Prototypes" grants. This work was also partially achieved with support of the *Data to Knowledge in Agronomy and Biodiversity* (D2KAB – www.d2kab.org) project that received funding from the French National Research Agency (ANR-18-CE23-0017) and the project "Partages de Connaissances" (PACON) of the transverse programme MetaBio funded by INRAE. We also thank Dr. Kevin Morel (INRAE), Matthieu Hirshy (ACTA) and Juliette Raphel (Elzeard) for curating the alignments as well as all contributors from MesclunDurab and D2KAB projects for their constructive feedback.

References

- [1] Aubert C., Buttigieg P.L., Laporte M.A., Devare M., Arnaud E., (2017) CGIAR Agronomy Ontology, <http://purl.obolibrary.org/obo/agro.owl>, licensed under CC BY 4.0.
- [2] Elizabeth Arnaud et al. "Towards a reference plant trait ontology for modeling knowledge of plant traits and phenotypes". In: *International Conference on Knowledge Engineering and Ontology Development*. Vol. 2. SciTePress. 2012, pp. 220–225.
- [3] Baptiste Darnala et al. "Crop Planning and Production Process Ontology (C3PO), a new model to assist diversified crop production". In: *Integrated Food Ontology Workshop (IFOW'21) at the 12th International Conference on Biomedical Ontologies (ICBO)*. 2021.
- [4] Damion M. Dooley et al. "FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration". In: *npj Science of Food* 2.1 (Dec. 2018), p. 23. issn: 2396-8370. doi: 10.1038/s41538-018-0032-6.
- [5] Pankaj Jaiswal et al. "Plant Ontology (PO): a controlled vocabulary of plant structures and growth stages". In: *Comparative and functional genomics* 6.7-8 (2005), pp. 388–397.

⁵<https://gd.eppo.int>

- [6] C. Jonquet et al. “AgroPortal: an ontology repository for agronomy”. In: *Computers and Electronics in Agriculture* 144 (2018), pp. 126–143.
- [7] F. Michel et al. “A Model to Represent Nomenclatural and Taxonomic Information as Linked Data. Application to the French Taxonomic Register, TAXREF”. In: *Proceedings of ISWC 2017 Workshop on Semantics for Biodiversity (S4Biodiv 2017), Oct 2017, Vienna, Austria* (2017), pp. 1–12.
- [8] Alistair Miles and Sean Bechhofer. “SKOS simple knowledge organization system reference”. In: (2009).
- [9] Raúl Palma et al. “An INSPIRE-based vocabulary for the publication of Agricultural Linked Data”. In: *International Experiences and Directions Workshop on OWL*. Springer, 2015, pp. 124–133.
- [10] Silvio Peroni. “A simplified agile methodology for ontology development”. In: *OWL: Experiences and Directions—Reasoner Evaluation*. Springer, 2016, pp. 55–69.
- [11] Catherine Roussey et al. “A methodology for the publication of agricultural alert bulletins as LOD”. In: *Computers and Electronics in Agriculture* 142 (2017), pp. 632–650. issn: 0168-1699. doi: <https://doi.org/10.1016/j.compag.2017.10.022>.
- [12] V. Soullignac et al. “GECO, the French Web-based application for knowledge management in agroecology”. In: *Computers and Electronics in Agriculture* 162 (2019), pp. 1050–1056.