Key topic

Interoper-ability

Metadata & Ontologies

Scientific domain

Earth and environmental sciences

Leading organisation

INRAE

FAIR Use Cases 10/09/2024

FAIR-IMPACT
Expanding FAIR solutions across EOSC

eosc

Enabling interoperability between AgroPortal and PHIS information system data repository for enhanced phenomics data annotation and exchange

The goal of this use case is to illustrate the benefit of using AgroPortal ontologies to describe, annotate and structure phenomics data within the PHIS platform, an open source information system for Plant Phenomics. We like: (i) to ease the reuse of semantic artefact objects (classes, concepts, properties, etc.) within PHIS to describe data and (ii) enable the push back of knowledge objects created by domain scientists within PHIS to application or domain ontologies hosted in AgroPortal.

Context and materials

In recent years, **plant phenomics has generated a large number of datasets from both field and controlled experiments**, comprising hundreds of genotypes at various scales of organization. These datasets represent resources for the identification of new mechanisms and patterns [https://doi.org/10.1016/j.cub.2017.05.055]. Given the heterogeneity of the data and the difficulty of accessing information distributed across multiple sources, assembling the set is challenging. For these challenges, an ontology-based information system, called PHIS (Phenotyping Hybrid Information System), inspired by FAIR principles, has been proposed. PHIS is a solution to integrate, organize and manage multi-source and multi-scale phenomic data obtained in the field and in greenhouses [https://doi.org/10.1111/nph.15385]. It is based on the generic OpenSILEX technology, developed and operated by INRAE-MISTEA. It is an open source ontology-based information system designed for life science data. The software suite implements original management methods for semantic exploitation, FAIR data production and adopts an architecture suitable for heterogeneity and increasing data volumes. **PHIS is a specific instance of the OpenSILEX technology for distributed plant phenotyping across various plant categories (field, greenhouse)** developed in part as part of the H2020 projects EPPN2020, EMPHASIS ESFRI and the national infrastructure PHENOME.

One of the main obstacles to interoperability and data reuse is the accurate identification and definition of measured variables (see the work of the I-ADOPT RDA working group). The commonly measured variable “plant height” may have different definitions depending on the crop, may be measured by different methods, and may be expressed in different units. To address this challenge, the Entity-Feature-Method-Unit model (https://cropontology.org/) was adopted to facilitate standardization of the measured variables:

- **Entity**: refers to the object being targeted
- **Feature:** indicates the type of measurement, which includes physical quantities and observed qualities
- **Method:** describes the approach used to estimate the variable
- **Units**: Describes the units used to quantify the variable

Within PHIS, each of these building blocks of a measured variable is mapped as much as possible to reference ontologies such as Plant Ontology and Crop Ontology or SOSA Ontology for sensors. For example, air temperature is modeled according to this scheme as:
Air_Temperature_ShelterInstantMeasurement_DegreeCelsius
Here the Entity is Air, the Feature is Temperature, the Method used is an instantaneous measurement using a shelter and the units are °C.
For this variable in PHIS, the different components are mapped (when possible) to existing reference ontologies (here retrieved from AgroPortal) and unique internal URIs generated by PHIS are also associated with the variable and each of these components. In this specific case, no reference term was found for the method 'ShelterInstantMeasurement', so the system generates a new term and associates it with an internal URI. Ultimately, this term will be a candidate for extending an ontology. While the previous example demonstrates the flexibility and freedom of PHIS to create new terms when users are unable to find them for any reason (such as lack of IT skills, time constraints, or familiarity with existing repositories and resources), this approach does not promote reuse of existing terms and limits interoperability with other resources.

- **Variable URI**
- **Entity** (Air)
- **Characteristic** (temperature)
- **Method** (ShelterInstantMeasurement)
- **Unit** (degree Celsius)

This structured approach allows the creation of new variables by combining these building blocks, for example, by changing the method or unit. **To obtain the ontological terms relevant to the variable, PHIS users are encouraged to use AgroPortal.** It is a vocabulary and ontology repository created as a reference catalog to host, share and serve semantic artifacts for agri-food communities. AgroPortal allows users to search and browse terms in an intuitive interface. The semantic artifact catalog can be automatically called by tools via its API.

Key topic

Interoper-ability

Metadata & Ontologies

Scientific domain

Earth and environmental sciences

Leading organisation

INRAE

Challenges and objective

While the plant phenomics community has embraced ontologies to standardize the description of experimental variables, **many users with a background in biology lack computer science skills and are unfamiliar with the use of ontologies or semantic artefacts, leading to difficulties in retrieving information** (as illustrated in the example above). Additionally, most available resources are not centralized, which further complicates the process of gathering information from multiple sources and mapping concepts. Currently, for PHIS users, **fetching ontology terms from AgroPortal or directly from multiple sources and ad-hoc vocabulary systems is a manual process**. This process requires going to another web application or tool, performing a search and manually copy/pasting the information found (if any) related to the selected ontology term. **This information is then used to fill in the necessary field for the mappings by specifying the mapping (more specific, more general)**. This manual process considerably prevents and slows down the reuse of standard ontology terms when describing objects within PHIS. The goal of our use case in FAIR-IMPACT T4.5 is to build a connector between PHIS and AgroPortal to ease the re-use of ontology terms when building variables and other scientific objects within PHIS.

Prototype connector between PHIS and AgroPortal

We (INRAE-MISTEA and INRAE-LEPSE) are working on a prototype (currently developed within the generic OpenSILEX technology and later moved to the PHIS instance) so PHIS users can easily describe, through Web interfaces, their measures, observations as scientific variables –using the model presented above– as well as their experimental vocabulary. **The connector allows users to search and grab from AgroPortal either a term URI and its related information (name, synonyms, definition) or create a new term within PHISand describe it with information and mappings coming from AgroPortal.** All the descriptions are stored in PHIS in RDF, the pivotal language of semantic knowledge graphs. Mapping between variable description and other domains ontologies are made by users with SKOS mapping relations (e.g., skos:exactMatch, skos:broadMatch).

The prototype connector under development aims to address this challenge by providing a semi-automatic ontology term fetching tool embedded within PHIS. The connector is executed within an OpenSILEX instance (here PHIS) and relies on AgroPortal API to get the information. **The connector will offer a number of features to make the fetching process easier and more efficient**, including:

- **An ergonomic search interface** that allows users to easily find the terms they are looking for.
- **Integration with pre-selected semantic artefacts** made available by AgroPortal, such as the AGROVOC thesaurus , the Crop Ontology or multiple references ontologies for plant sciences.
- **A mapping functionality that allows users to specify the links between terms from different ontologies** especially in the case a new term is created in PHIS and AgroPortal is only used to grab information and mappings to other close terms (Fig. 4). This connector will facilitate the search for new terms to PHIS users.

In the future, following the same philosophy and technical behavior (API calls) in addition to consuming the content from AgroPortal, the connector will open up the possibility of proposing content (e.g., terms and mappings) to AgroPortal ontologies and semantic artefacts. This will valorize the new intrinsic contributions made in creating scientific objects (variables, terms, properties, etc.) by PHIS data experts which would engage with external ontology experts and users who validate these terms.

The connector will provide a number of benefits to users of PHIS and will ultimately contribute the AgroPortal's content, including:

- **Reduced time and effort required** for reusing or mapping the terms used in PHIS.
- **Improved interoperability** between different dataset and ontologies.

The connector is currently under development as a demonstrator within FAIR-IMPACT T4.5.
Eventually, **the connector would be made completely generic to work with any instance of OntoPortal (**on the semantic artefact catalogue side) and on any instance of PHIS (on the data repository side).

Contributors

Anne Tireau
(INRAE)

Arnaud Charleroy
(INRAE)

Llorenc Cabrera-Bosquet
(INRAE)

Clement Jonquet
(INRAE)