An INSPIRE-based vocabulary for the publication of Agricultural Linked Data

Raúl Palma[1], Tomas Reznik[2], Miguel Esbrí[3], Karel Charvat[4], and Cezary Mazurek[1]

[1]Poznan Supercomputing and Networking Center, Poznań, Poland
rpalma@man.poznan.pl mazurek@man.poznan.pl
[2]Masaryk University, Brno, Czech Republic
tomas.reznik@sci.muni.cz
[3]Atos, Madrid, Spain
miguel.esbri@atos.net
[4]Wirelessinfo, Litovel, Czech Republic
charvat@ccss.cz

Abstract. FOODIE project aims at building an open and interoperable agricultural specialized platform on the cloud for the management, discovery and large-scale integration of data relevant for farming production. In particular, the integration focuses on existing open datasets as well as their publication in Linked data format in order to maximize their reusability and enable the exploitation of the extra knowledge derived from the generated links. Based on such data, for instance, FOODIE platform aims at providing high-value applications and services supporting the planning and decision-making processes of different stakeholders related to the agricultural domain. The keystone for data integration is FOODIE data model, which has been defined by reusing and extending current standards and best practices, including data specifications from the INSPIRE directive which are in turn based on the ISO/OGC standards for geographical information. However, as these data specifications are available as XML documents, the first step to publish Linked Data required transforming or lifting FOODIE data model into semantic format. In this paper, we describe this process, which was conducted semi-automatically by reusing existing tools, and adhering to the mapping rules for transforming geographic information UML models to OWL ontologies defined by the ISO 19150-2 standard. We describe the challenges associated to this transformation, and finally, we describe the generated ontology, providing an INSPIRE-based vocabulary for the publication of Agricultural Linked Data.

1 Introduction

The agriculture sector has been of strategic importance for both European citizens (consumers) and European economy (regional and global) since the conception of the EU; it was one of the first sectors of the economy to receive the attention of EU policymakers[2]. And despite the fact that its contribution to the overall EU economy has slightly decreased during the previous years, agriculture (with forestry and fishing) represented about 1,7% of the EU-28 Gross Added Value and accounted for 4.9% of the total number of persons employed in 2013[3]. As a result, the EU has developed policies and innovation programs that tackle the challenges associated to improve the efficiency of agricultural activities with a limited environmental footprint (see [1]).

Along these lines, we claim that in order to make economically and environmentally sound decisions, the different stakeholders groups involved in the agricultural activities need integrated access to multiple and heterogeneous sources of information collected by multiple applications and devices. In this context, FOODIE project[1] aims at building an open and interoperable cloud-based platform addressing among others the integration of data relevant to farming production, particularly from open datasets, as well as their publication in Linked data format.

In order to build such platform, we defined the modeling approach for the categories of information the platform will have to deal with, including their thematic, spatial and temporal characteristics as well as their meta-information. Such approach relies on reusing and extending standards and best practices to specify FOODIE data model. In particular we reused data specifications from the INSPIRE directive[2], which in turn are based on ISO/OGC standards for geospatial services and formats[3], thus applying the ISO/OGC-approach of modeling physical things, so-called "features". The specifications are defined as UML models and are available in different XML-based formats (e.g., GML, XMI) and as Enterprise Architect (EA)[4] projects. Accordingly, FOODIE data model was specified in UML by extending and specializing INSPIRE data model for Agricultural and Aquaculture Facilities (AF) [6].

However, according to the methodological guidelines for Linked Data publication [5], we need to specify the model for the representation of the data elements and their relationships in RDF format. This usually involves the specification of a lightweight ontology (or vocabulary), reusing standard vocabularies wherever possible. In our case, this required transforming FOODIE UML data model into an RDF vocabulary. But in addition to reusing standard vocabularies, our requirement was to comply with standard rules for mapping ISO UML models to OWL ontologies. In this paper we describe this transformation process, and the resulting ontology.

2 Transformation

We evaluated different approaches in the literature (e.g., [8], [9]) for the transformation process, and decided to follow a semi-automatic one using ShapeChange tool. ShapeChange can process application schemas for geographic information from a UML model (e.g., XMI) and derive implementation representations, such as XML schemas, feature catalogs, and RDF/OWL. In our case, we were interested in the RDF/OWL processor that is based on the ISO 19150-2 standard [7] defining rules for mapping ISO geographic information UML models to OWL ontologies.

2.1 Pre-processing tasks

Source model FOODIE UML data model required some changes before processing it in ShapeChange. These changes led to the release of a new version (v4.3.2)[5] and include: (i) assignment of INSPIRE application schema stereotype to include the target

[1] http://foodie-project.eu/

[2] INSPIRE directive (http://inspire.ec.europa.eu/) aims at building a Pan-European spatial data infrastructure (SDI), requiring EU Member States to make available spatial data, from multiple thematic areas, according to established implementing rules using appropriate services [4].

[3] http://www.opengeospatial.org/standards/is

[4] http://www.sparxsystems.com/

[5] Available at https://git.man.poznan.pl/stash/projects/FOOD/repos/model/browse/

namespace; (ii) fixing inconsistent range usage for attribute *code*; (iii) naming target sides of aggregations and associations for the generation of named object properties. The model was then published as XMI from EA tool, but we had to remove manually the ASCII code for Carriage Return encoded as an XML character reference in the file.

ShapeChange configuration The primary mechanism for providing arguments to ShapeChange is the configuration file. The two main components of this file are the encoding rules and the mappings. The first drives (broadly) the conversion from an application schema in UML to another data structure. The second supports customized mappings from UML classes to target OWL elements, by enabling the specification of generic rules. Additionally, the ShapeChange processor relies on different base ontologies for the generation of the RDF model, and thus the configuration file includes namespaces definitions for these ontologies. In particular, the processor uses geo-spatial ontologies, including those based on ISO 19100 series standards GeoSPARQL OGC standard and INSPIRE specifications; and in line with the Linked Data publication guidelines, it reuses several standard vocabularies like rdf, skos, dublin core and PROV.

We used as starting point the sample configuration settings in http://shapechange.net/targets/ontology/uml-rdfowl-19150-2/ and customized it according to our needs[6]. In particular, we applied the following rules: (i) ontologies are created only for the selected schema; (ii) constraints on properties and classes are specified; (iii) feature types get a subClassOf declaration to the GeoSPARQL FeatureType class; (iv) feature types get a subClassOf declaration to the ISO 19150-2 FeatureType and ISO 19109 AnyFeature classes; (v) data types get a subClassOf declaration to the ISO 19150-2 Datatype class and code lists get a subClassOf declaration to ISO 19150-2 Codelist class; (vi) cardinality restrictions are specified; (vii) allValuesFrom restrictions are not specified; (viii) minCardinality is set to 0 for voidable properties; (ix) *dc:source* is included only on the ontology subject; (x) association names are not specified; (xi) the namespace abbreviation for the application schema is used for the ontology name and filename.

We included more than ten mapping entries in the configuration file for the classes and properties referenced in the model (see Section 3). We also fixed and added several namespaces in the configuration file, i.e., many of the namespaces for the geo-spatial ontologies were outdated or incorrect (e.g., INSPIRE, iso19150-2 and iso19109 ontologies), and we needed to include new namespaces for the mapping entries we created (e.g., iso19103, iso19108 and iso19115-citation ontologies).

Base ontologies The base INSPIRE ontology (the schema for basic types used by multiple themes)[7] was slightly modified to load it correctly, namely we: (i) added namespace of geosparql ontology (missing); (ii) fixed namespace of iso19150-2 based ontology, and removed the ontology import statement (because of few inconsistencies - see discussion below); (iii) fixed *VerticalPositionValue* datatype declaration; (iv) changed the ontology namespace to avoid multiple base prefixes.

The original iso19103 ontology[8] treated a few datatype as classes. For instance *Number* is defined as an equivalent class to the union of primitive numerical datatypes (*xsd:decimal*, *xsd:double*, *xsd:float* and *xsd:integer*), and as a result it was declared both

[6] https://git.man.poznan.pl/stash/projects/FOOD/repos/model/browse/shapechange-conf
[7] http://portele.de/ont/inspire/base.ttl
[8] http://def.seegrid.csiro.au/isotc211/iso19103/2005/basic

as class and datatype. We removed the class declarations, however they are still being treated as classes (as it was intended). This is possible in RDF, but in OWL terms this means that we have an OWL full ontology, as in all reasonable profiles (OWL 2 DL and below) datatypes and classes need to be disjoint.

Overall, we found some issues with the ontologies based on the ISO 19100 series standards. They are in provisional state, although they were created between 2012 and 2013, and in many cases the versions changed drastically. For instance, ISO 19100 series standards define UML profiles that include a list of stereotypes and basic types to be used in application schemas. Accordingly, the ISO 19150-2 based ontology defined classes for these stereotypes, including ⟨⟨datatype⟩⟩, ⟨⟨codelist⟩⟩, ⟨⟨featureType⟩⟩, and the base class ⟨⟨anyFeature⟩⟩. However, the latest version of this ontology does not declare all these classes, as it did in the previous version (used in ShapeChange). Additionally, the ontologies miss several elements from the standard and in most cases the ontologies are only available as OWL full ontologies (e.g., treating datatypes as classes). We tried unsuccessfully to reach the developers to discuss these issues.

2.2 Post-processing tasks

We had to make some manual fixes in the ontology after executing the transformation, including updating incorrect namespaces added automatically by the processor rules (hard-coded), adding missing prefixes and removing unnecessary imports of ontologies to avoid ending up with a heavy ontology. Additionally, as ShapeChange only processes the selected schema (i.e., FOODIE data model), we had to add manually the ontology elements (corresponding to the UML elements) of the base INSPIRE schemas, particularly those from the Agriculture and Aquaculture Facilities theme. Finally, we removed an axiom generated to constraint the cardinality of the property *rdfs:label* in a class expression (*rdfs:label* was the mapped property for the UML element "name") because *rdfs:label* is a predefined annotation property so it can only be used in annotations.

3 Ontology

FOODIE ontology is available at https://git.man.poznan.pl/stash/projects/FOOD/repos/model/browse/foodie.ttl?raw, and its classes along with their parent stereotype class are depicted in Figures 1, 2 and 3. In the reminder we describe the main ontology elements (ontology classes are in italics): For the purposes of FOODIE, we found the lack of

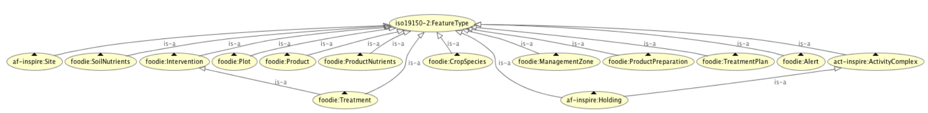

Fig. 1. FOODIE ontology subclasses of ISO 19150-2 *FeatureType* class (and ISO 19109 *AnyFeature* and geosparql *Feature* classes)

a feature on a more detailed level than *Site* that is already part of the INSPIRE AF data model. The main motivation was to represent a continuous area of agricultural land with one type of crop species, cultivated by one user in one farming mode (conventional vs. transitional vs. organic farming). Such concept is called *Plot* and represents the main element in the model, specially because it is the level to which the majority of agro data

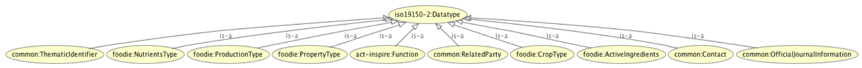

Fig. 2. FOODIE ontology subclasses of ISO 19150-2 Datatype class

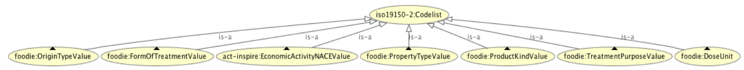

Fig. 3. FOODIE ontology subclasses of ISO 19150-2 Codelist class

is related. One lower level than *Plot* is the *ManagementZone*, which enables a more precise description of the land characteristics in fine-grained areas. The *Plot* has associated two kinds of data: (i) metadata information, including properties: code (id), validity (when the plot started and ceased to exist), geometry (spatial extent), description and originType (manual, system); (ii) agro-related information, including:

- *ProductionType*, representing production-related data, comprising properties: productionDate (when the information was added/changed in the knowledge base (KB)); variety (assemblage of cultivated individuals that are distinguished by characteristics significant for agriculture, e.g., morphological, physiological, cytological, chemical); productionAmount (physical quantity of produced variety).
- *CropSpecies*, representing the planted crop species, comprising properties: date (when it started/ended to be planted on the Plot); cropArea (spatial extent on the Plot); cropSpecies (designation under which it is commonly known).
- *Alert*, representing alerts generated by the models integrated in the platform, comprising properties: code; type (according to user-defined classification, e.g. phytosanitary); description; checkedByUser (indication of user awareness); alertDate (creation); alertGeometry (spatial extent for which it is applicable).
- *Intervention*, representing the basic feature type for any application with explicitly defined geometry, comprising properties: type (free text (e.g., tillage, pruning)[9]); description; notes (user-defined); status (free text); creationDate (in the KB); interventionStart/End (when started/ended in the real word); interventionGeometry (spatial extent); supervisor (entity with authority to guarantee its execution); operator (person who executed it); evidenceParty (entity who added it in the KB); price.

The intervention has direct and indirect associations to the following entities:

- *Treatment* comprising properties: quantity (applied physical quantity); tractorId (vehicle for machine applying it); machineId (machine applying it); motionSpeed (recommended speed for its application); pressure (recommended pressure for its application); flowAdjustment (indication if flow adjustment was needed for its application); applicationWidth (width in which a machine is capable to apply it); areaDose (maximum application rate); formOfTreatment (id of its application, e.g., manual, aerial, from a code list); treatmentPurpose (rationale why it was used, e.g., weed, pest, from a code list); treatmentDescription.

[9] In the cases of free text properties, it was not feasible to provide common code lists (e.g., values vary from country to country or from farm to farm)

- *TreatmentPlan* comprising properties: treatmentPlanCode; description; type; campaign (period to which it was designed); treatmentPlanCreation (in the KB); notes.
- *ProductPreparation* comprising properties: productQuantity (physical quantity of the applied product); solventQuantity (physical quantity of solvent applied); safetyPeriod (when a dissolved product may be used).
- *Product*, comprising properties: productCode; productName; productType (free text); productSubType (detailed classification - as free text); productKind (origin, e.g., organic, mineral - from a code list); description; manufacturer; safetyInstructions; storageHandling (for safe storage); registrationCode (id according to the national or other relevant registration scheme); registerUrl (link to the national (or other) registry); nutrients (id of nutrients, i.e., chemical elements and compounds necessary for plant growth, represented by *NutrientsType* class comprising properties for the amount of nitrogen, phosphorus pentoxide, potassium oxide and other chemical elements.
- *ActiveIngredients* with properties: code, ingredientName, and ingredientAmount.

4 Conclusion

The publication of Agricultural Linked Data is unfortunately still not a common practice. One of the key issues for this is the lack of vocabularies for modeling this data in RDF format. In FOODIE project, we have developed an ontology for this task which complies and adheres to existing standards for the representation of geo-spatial data relevant for agriculture. In particular, we extended and specialized INSPIRE UML data model for Agricultural and Aquaculture Facilities and transformed this model into a lightweight ontology. We conducted this process semi-automatically reusing ShapeChange tool, which enables the transformation of UML models in XMI into RDF. The transformation required several pre and post processing tasks, in order to build the final ontology. We described in detail this process, the challenges associated, and finally, we presented the resulting ontology. At the moment we are working on the application of this ontology in FOODIE platform.

Acknowledgments The research reported in this paper has been supported by the EU FOODIE project (http://foodie-project.eu/, CIP-ICT-PSP-2013-7, Pilot B no.621074).

References

1. European Comission. Agricultural and rural development, research and innovation, challenges. http://ec.europa.eu/agriculture/research-innovation/challenges/index_en.htm, 2015.
2. Eurostat. Enlargement countries agriculture, forestry and fishing statistics. http://goo.gl/Sm121D, 2014.
3. Eurostat. Enlargement countries agriculture, forestry and fishing: tables and figures. http://goo.gl/m3mkFA, 2014.
4. A. Fichtinger. INSPIRE Roadmap and Implementation. In *INSPIRE-GMES Information Brochure*, pages 2–10. Technische Universität München, 7th edition, Oct 2011.
5. B. Hyland, G. Atemezing, and B. Villazon-Terrazas. Best practices for publishing linked data. http://www.w3.org/TR/ld-bp/, 2014.
6. INSPIRE Thematic WG Agricultural and Aquaculture Facilities. D2.8.III.9 data specification on agricultural and aquaculture facilities. http://goo.gl/eWi6rq, Dec 2013.
7. ISO. Standard 19150-2:2015: Geographic information - Ontology - Part 2. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57466, July 2015.
8. S. Tschirner, A. Scherp, and S. Staab. Semantic access to inspire. In *Terra Cognita'11 Workshop Foundations, Technologies and Applications of the Geospatial Web*. Citeseer, 2011.
9. L. Van den Brink, P. Janssen, and W. Quak. From geo-data to linked data: Automated transformation from gml to rdf. *Linked Open Data-Pilot Linked Open Data Nederland*, 2013.