Publishing Agro-Environmental Data to a Semantically Unified Environmental Information Space: a Case Study

Sasa Nešić¹, Andrea Emilio Rizzoli¹, Ioannis Athanasiadis², Marcello Donatelli³

Abstract

Publishing environmental data to Linked Open Data (LOD) requires environmental data providers to adopt a set of universally recognized linked data principles. These principles enable syntactic integration of the linked data regardless of the data publisher and origin. However, these principles do not provide any mechanism for semantic integration and interoperability of the linked data. In order to turn the linked environmental data into a semantically unified environmental information space, the linked environmental data must be accompanied by conceptualized semantics defined in domain conceptualizations known as domain ontologies. Taking into account the linked data principles and domain ontologies we proposed a set of requirements and a data-publishing workflow for environmental data providers. In this paper we present a case study supporting our approach by delivering: 1) an agro-environmental domain ontology and 2) a supporting software system that realizes the proposed data-publishing workflow. We demonstrate the publishing workflow on publishing some example, agro-environmental resources provided by JRC MARS, which is one of the leading European institutions for monitoring agricultural resources founded by the European Commission.

1. Introduction

As a result of pervasive sensor networks, advances in automated data acquisition, and the widespread use of computer based models and decision support systems, we have witnessed a proliferation of environmental resources on the Web in recent years. Typical examples of environmental resources available on the Web include observational sensor data, satellite images, processed data, catalogues, maps and reports as well as environmental models, software and decision support systems. The majority of these resources are however, generally used by closed communities, which are aware of them and are able to access them. Common practice has proven that environmental data are usually stored in non-reusable raw formats, situated in sparse locations and managed by different authorities, which ultimately raise obstacles in making environmental information accessible (Villa 2011). Moreover, apart from the problem of resource accessibility, there is also a big lack of semantic interoperability of environmental resources (Athanasiadis 2011). Related information assets, held by environmental resources published and managed by different authorities, are usually unaware of each other. As a consequence, the available environmental resources on the Web look more as sets of disconnected information islands than an integrated information space.

¹ IDSIA – Dalle Molle Institute for Artificial Intelligence, Manno-Lugano, Switzerland email: sasa@idsia.ch; andrea@idsia.ch, Internet: http://www.idsia.ch
² Democritus University of Thrace, Xanthi, Greece email: ioannis@athanasiadis.info, Internet: http://www.duth.gr/
³ JRC-IPSC Agric4cast Action, Ispra, Italy email: marcello.donatelli@jrc.ec.europa.eu, Internet: http://mars.jrc.ec.europa.eu/
Over recent years, the adoption of the Linked Data principles for publishing and collecting structured data on the Web (Bizer 2009; Berners-Lee 2006) has opened the possibility of creating a global, unified information space, connecting data from different sources and domains. Therefore, the adoption of the Linked Data principles for publishing environmental resources could also lead towards unification of the currently disconnected environmental resources. Moreover, employing rich semantics in the form of domain ontologies (Fensel 2001) to annotate resources and applying the Semantic Web resource description format (RDF)\(^4\) to represent resource descriptions could improve semantic integration of related environmental resources.

Having environmental resources published and annotated according to the Linked Data and the Semantic Web principles would facilitate a vision of the semantically unified environmental information space. By analyzing these principles and taking into account specifics of the environmental domain, we proposed a set of publishing requirements (Nešić 2011) that environmental resource providers should respect in order to get their resources fully contribute to this vision. In this paper we present a case study, which we conducted in order to verify the proposed publishing approach and the underlying publishing requirements on the example of publishing concrete agro-environmental resources (i.e., JRC MARS resources). The case study had two main objectives.

The first objective of the case study was to design and develop an agro-environmental domain ontology that would be used to semantically annotate the JRC MARS resources and thus enable their semantic integration into LOD. We designed the ontology independently from existing semantic annotation frameworks and knowledge bases. However, it can be easily aligned to any existing semantic framework that is based on the W3C-endorsed OWL specification. As an example, we provided an alignment/mapping of the developed ontology to the TaToo semantic framework (ParienteLobo 2011). The TaToo framework is an ontology-based annotation framework allowing for formal description of environmental resources. It is based on a minimal environmental resource model (MERM), which represents the largest common denominator between a set of heterogeneous description formalisms related to environmental resources.

The second objective of the case study was to design and develop the supporting data-publishing system that would utilize the developed ontology and realize proposed publishing workflow. The system was supposed to provide all necessary services and tools that would comply with the publishing requirements that we proposed (Nešić 2011). Moreover, the tools should have been designed to hide all specifics of the publishing process from the publisher.

With respect to the first objective of the case study, we first give an overview of the TaToo semantic framework (Section 2), then describe in details the JRC MARS ontology (Section 3), and conclude with the mapping of the JRC MARS ontology to the TaToo semantic framework (Section 4). With respect to the second objective of the case study (Section 5), we first present the architecture of the developed system and outline the main features of the system’s services and then we demonstrate the resource publishing on the example of some resources provided by JRC MARS. In addition, we also show how the system enables users of the agro-environmental resources to evaluate published resources as well as their annotations.

2. TaToo Approach

TaToo stands for “Tagging Tool based on a Semantic Discovery Framework”. The main goal of TaToo approach is to narrow the discovery gap that prevents a full and easy access to environmental resources on the Web. To achieve this goal, TaToo introduces a comprehensive framework for ontology-based, semantic tagging (annotation) of environmental resources (ParienteLobo 2011). The framework comprises a

\(^4\) RDF: http://www.w3.org/RDF/
novel, environmental resource model called MERM (Minimal Environmental Resource Model), a number of integrated environmental domain ontologies, and a service-oriented, supporting system providing intended tagging functionalities. The tagging process involves both environmental experts and general users of the environmental resources. In this way TaToo aims at enriching environmental resources by a sufficient amount of valuable semantic annotations. Moreover, in order to ensure validity of actual resources as well as validity of added resource annotations, TaToo introduces services and data representation structures for capturing and representing community-based evaluations of both resources and resource annotations. In the rest of this section we outline those parts of the MERM model that are key to the resource annotation and the resource and annotation evaluations.

### 2.1 Annotating environmental resources

MERM specifies the structure of environmental resources and their annotations managed by the TaToo system, and represents a reference data model for both services and user interfaces. It is formally described by the MERM ontology. Figure 1 shows the core concepts and properties of the ontology.

![Core concepts of the MERM ontology](image)

The `merm:Resource` concept represents an environmental resource managed by the TaToo system. Each resource is characterized by the resource access information, which is modeled by the `merm:ResourceAccessInfo` concept and its properties. This information can be very heterogeneous, ranging from a simple URL to a complex WSDL depending on the nature of the resource. Next, each resource is described by a set of basic information such as the resource author, owner, and creation date, which are modeled by the `merm:ResourceInfo` concept and its properties. Finally, each resource can be accompanied by a set of resource annotations, modeled by the `merm:Annotation` class. A resource annotation provides a semantic description of the resource including the resource’s type and the domain-specific knowledge held within the resource. The domain-specific knowledge is represented by a set of concepts from domain ontologies, which are linked to the resource annotation via `merm:subject` property. In order to ensure uniform integration/interpretation of the annotation concepts defined by different domain ontologies, TaToo introduces bridge ontology. This bridge provides the `bridge:Topic` concept that aligns domain-specific concepts to the MERM annotation subjects.
2.2 Evaluating resource and resource annotations

One of the main issues regarding the current Web is how to assess the quality of information resources. There have been proposed numerous guides and approaches regarding this subject (Smith 1997, Vaughan 2006, Phillips 2004). A resource quality can be evaluated with respect to a number of evaluation criteria, each of which is determined by a number of different indicators. In order to provide a comprehensive evaluation of the managed environmental resources, the TaToo system enables the evaluation of not only the actual resources but also their annotations. In other words, TaToo considers both environmental resources and resource annotations as information entities that can be evaluated and distinguished between two types of evaluations: 1) resource evaluation and 2) annotation evaluation. Figure 2 shows a part of the MERM ontology providing concepts and properties for modelling evaluations of resource and resource annotations. Each evaluation, either it is a resource evaluation or annotation evaluation, is characterised by the evaluation criteria, evaluation metric and evaluation value. Moreover, each evaluation is also characterised by the evaluator whose profile information are modelled within the TaToo user profile. The resource evaluations are linked to the instances of the merm:resource concept directly, while the annotation evaluations are linked to the instances of the merm:Annotation concept.

Figure 2: Evaluation concepts of the MERM ontology
3. JRC MARS Domain Ontology

The JRC MARS is a domain ontology that we developed aiming at conceptualizing various aspects of the agricultural resources. We have named the ontology after MARS (Monitoring Agricultural Resources) unit of the European Commission’s Joint Research Centre (JRC), whose resource annotation the ontology is supposed to support. The ontology’s namespace has also been chosen according to the Internet domain of the resource provider:

xmlns:jrc = “http://mars.ies.jrc.eu/software/JRC_MARS_physical_domain/”

The ontology’s development has been gone through several iterations so far. The current ontology version contains two main groups of concepts. The first group is composed of concepts dedicated to modeling resource types of agricultural resources monitored and managed by the JRC MARS unit. The second group provides concepts describing agricultural topics that are related to the managed resources. We now present core concepts of both groups respectively.

3.1 Modeling agricultural resource types

By analyzing agricultural resources managed by JRC MARS we have distinguished between five resource types: 1) Agricultural Software, 2) Agricultural Model, 3) Remote Sensing Data, 4) Weather Scenarios Data, and 5) Agricultural Maps.

1. **Agricultural Software** refers to software enabling remote sensing and meteorological observations, agro-meteorological modeling, and statistical analysis. It is specified in the ontology by the jrc:AgriculturalSoftware concept. Moreover, we distinguished between three types of agricultural software: application, library and service. They are specified in the ontology by jrc:SoftwareApplication, jrc:SoftwareLibrary and jrc:WebService concepts, which are sub-concept of the jrc:AgriculturalSoftware.

2. **Agricultural Model** refers to mathematical, process-based models that are used to simulate crop development and growth. It is specified in the ontology by the jrc:AgriculturalModel concept. Figure 3 shows the sub-conceptualization of the jrc:AgriculturalModel concept.

3. **Remote Sensing Data** refers to discrete units of raw data provided by sensor networks and satellite imaging. It is specified in the ontology by the jrc:RemoteSensingData concept.

4. **Weather Scenario Data** refers to data collections describing weather (e.g., temperature, precipitation, dryness, humidity and wind) in a certain area for a given time period. It is specified in the ontology by the jrc:WeatherScenarioData.

5. **Agricultural Map** refers to maps showing the spatial and temporal variability of productivity, yield, and canopy density of filled crops as well as observed temperatures and precipitations. It is specified in the ontology by the jrc:AgriculturalMap concept.
3.2 Modeling agricultural topics

Similar to any other domain, the agricultural domain is characterized by a number of domain specific topics (subjects). In reality, it is almost impossible to identify and conceptualize all topics in a given domain. There are no strict borders between “neighboring” domains as well as each domain evolves and over time. What is realistic is to focus on several aspects of the domain of interest (e.g., related to the available resource (data) set) and then try to identify relevant concepts. In case of the JRC MARS we focused on two aspects of the agricultural domain: agricultural processes and agrophysical systems. All agricultural topics that we introduced in the ontology are defined as members of the jrc:AgriculturalTopic concept’s sub-tree. The jrc:AgriculturalTopic is an upper-level, domain concept which is defined as a sub-concept of the MERM’s bridge:Topic concept.

3.2.1 Modeling agricultural processes

By an agricultural process we refer to a process that is involved or related to the crop development and growth. The jrc:AgriculturalProcess is a general, upper-level concept that we introduced to model the agricultural processes. As it was mentioned above, we consider the agricultural processes as a subset of the
agricultural topics thus the \textit{jrc:AgriculturalProcess} concept is defined as a sub-concept of the \textit{jrc:AgriculturalTopic} concept. Moreover, we have defined a list of concepts dedicated to more specific processes such as \textit{jrc:CropProcess}, \textit{jrc:SoilProcess}, \textit{jrc:DiseaseProcess}, and \textit{jrc:InsectProcess}, which model processes related to the crop development, soil management, disease treatments, and insect control.

Besides these concepts, we also added to the ontology a number of individual processes, that is, instances of the process concepts. Let us list few of them such as \textit{jrc:Drainage}, \textit{jrc:Denitrification}, \textit{jrc:Evaporation} and \textit{jrc:WaterRedistribution}, which are instances of the \textit{jrc:SoilProcess}, and \textit{jrc:Respiration}, \textit{jrc:BiomassFixation} and \textit{jrc:GrainFilling}, which are instances of the \textit{jrc:CropProcess}.

\subsection{Modeling agrophysical systems}

Agrophysical systems are biophysical systems that concern the biology of plants, animals, soil and an atmosphere involved in agricultural activities and biodiversity. We model general agrophysical systems by the \textit{jrc:AgrophysicalSystem} concept, which is defined as a sub-concept of the \textit{jrc:AgriculturalTopic} concept. The following is a list of sub-concepts of \textit{jrc:AgrophysicalSystem} that we introduced to the ontology to model more specific agrophysical systems: \textit{jrc:Crop}, \textit{jrc:Disease}, \textit{jrc:Insect}, \textit{jrc:Soil}, and \textit{jrc:Tree}, \textit{jrc:Weather};

\section{Mapping the JRC MARS ontology to the TaToo ontology framework}

The TaToo framework can be extended by an arbitrary number of domain ontologies. However, in order to plug-in new domain ontology to the TaToo framework, the ontology needs to be accompanied by an appropriate mapping/alignment to the MERM ontology (model). This mapping has to be created according to one of the two TaToo proved alignment strategies (ParienteLobo 2011). In short, the first mapping strategy is about using the \textit{rdfs:subClassOf} RDF construct to map concepts of the domain ontology to the corresponding MERM concepts, while the second strategy assumes the use of the \textit{owl:equivalentClass} and \textit{owl:sameAs} OWL constructs. In this section we demonstrate the second strategy on the example of the mapping the JRC MARS ontology.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{mapping_jrc_mars_onto_tattoo.png}
\caption{Mapping of the JRC MARS ontology to TaToo ontology framework}
\end{figure}
We localized the mapping of the JRC MARS domain to the TaToo ontology framework to the following five concepts: \texttt{jrc:AgriculturalModel}, \texttt{jrc:AgriculturalSoftware}, \texttt{jrc:RemoteSensingData}, \texttt{jrc:WeatherScenarioData}, \texttt{jrc:AgriculturalMap} and \texttt{jrc:AgriculturalTopic}. As Figure 4 shows, the first five concepts, which are upper-level concepts of the JRC MARS resource types, are defined as sub-concepts of the \texttt{merm:Annotation} concept, while \texttt{jrc:AgriculturalTopic}, which is an upper-level concept intended to model topics of the agricultural domain, is defined as a sub-concept of the \texttt{bridge:Topic} concept. By these sub-conceptualizations of the JRC MARS concepts from the MERM and bridge concepts, on one hand we enabled the TaToo system to seamlessly use the JRC MARS ontology and on the other hand we let the JRC MARS ontology evolves over time without any repercussions on the defined mapping.

5. **AGROPub: Agricultural Resource Publishing system**

In order to facilitate a vision of a semantically unified, global environmental information space (Nešić 2011), which should happen as a part of the LOD cloud, we have developed a system called AGROPub. AGROPub is designed as an extensible, service-oriented architecture that provides a variety of services supporting both agro-environmental resource providers and resource consumers. In particular, it provides services that enable resource providers to publish RDF descriptions of their resources (generated according to the MERM model) to LOD and to set up some initial links between their resource and related resources available in LOD. Since the real resource integration to LOD is more likely to happen over time as a result of the user interaction with the resources (e.g., searching and reusing, annotating, interlinking, and evaluating), the system also provides services for searching, annotating, interlinking and evaluating agro-environmental resources published to LOD.

Figure 5 illustrates the architecture of AGROPub. We designed it as a tree-tier, service oriented architecture composing of the data tier, service tier, and presentation tier.

1. The **data tier** consists of a number of RDF repositories storing resource descriptions (i.e. RDF triplets) of the published agro-environmental resources. Each of the RDF repositories from the data tier exposes an HTTP de-referenceable SPARQL endpoint, so that the published resources can be referred and linked to other resources from LOD.

2. The **service tier** comprises a set of five core services: a) Publishing and Annotation, b) Semantic Linking, c) Semantic Search, d) Resource Evaluation, and e) Semantic Navigation and a service registry that provides registration and look up functionality for the AGROPub services. The service tier can be extended by an arbitrary number of new services, whose design has to comply with the AGROPub design principles. To be operational within the system, a new service first needs to be registered to the service registry.

3. The **presentation tier** is a top tier in the architecture providing the user interface for the AGROPub services. It is technology and platform independent and can include Web-based applications, desktop applications, and mobile phone applications.

In order to prove the architecture's implementability and to enable experimental validation of the proposed services, we have developed the AGROPub prototype. The prototype is fully functional software, providing the implementation of all three tiers from the AGROPub architecture (Figure 5). While the implementation of the data tier relied mainly on the deployment of an existing RDF repository (Sesame), we implemented the service and user interface tiers completely from scratch. The services and the communication between them and the presentation tier are implemented according to the WCF (Windows Communication Framework).

---

\footnote{Sesame: http://www.openrdf.org/}

7/26/11, SasaNesic.doc
Foundation) principles. At the presentation tier, the prototype offers several tools that are integrated in a desktop application namely AGROPub Manager. More information, snapshots and demos of the tools can be found at our project Web page\(^6\).

Figure 5: Illustration of the AGROPub architecture

In the rest of this section we provide more details on the AGROPub services that are involved in the two AGROPub use-cases (i.e., resource publishing and resource evaluation), which were part of the case study that we present in this paper.

5.1 Publishing JRC MARS resources with AGROPub

Publishing JRS MARS resources with AGROPub includes generating RDF resource descriptions according to MERM, adding ontology-based annotations to the generated resource descriptions, storing the resource descriptions to the RDF repository, and setting up an initial set of links between the generated resource descriptions and related resources from LOD. Therefore, the publishing process requires the agro-environmental resource providers to interact with the Publishing and Annotation sand the Linking services. This interaction is enabled by the resource publishing, annotation, and linking tools, which are provided by the AGROPub Manager.

\(^6\) SDArch Web page: www.semanticdoc.org/AGROPub
Figure 6: An AGROPub-generated resource description

Figure 6 exhibits a snippet of an example RDF resource description, which is generated and published by the AGROPub system. Let us now briefly describe this example. Line 2 contains the opening tag of the RDF resource description of the published resource. The opening tag holds the information of the resource’s URI. Lines 3 to 12 contain a set of basic resource info, including the resource’s creator, owner, description Web page, creation date, publisher, and publication date. Then, Lines 13 to 38 contain the resource’s annotation information. Line 14 shows that the published resource is of the jrc:CropModel type. Line 15 and 16 hold information about the annotation’s creator and creation date respectively. Line 17 to 20 hold domain specific information, that is, a list of domain topics related to the annotated resource. In this example, domain topics are defined in the JRC MARS ontology. Line 21 to 37 contains RDF descriptions of two annotation evaluations that we describe in details in the next section. Finally, Line 38 contains the closing tag of the resource’s RDF description.

After publishing the resource, the resource’s description contains only initial annotation (Lines 13 to 38) provided by the resource publisher at the publishing time. Once the resource is published, a number of resources annotations, provided by various users, can be added to the resource’s description. The same apply to the annotation evaluations.

5.2 Evaluating JRC MARS resources with AGROPub

Evaluating JRC MARS resources with AGROPub includes the evaluation of the actual resources and the evaluation of resource annotations. Both types of the evaluation are community-based evaluations, performed by the AGROPub users. AGROPub provides the Evaluation service that generates the resource evaluations and annotation evaluations based on the evaluation criteria and evaluation values specified by a user. The resource and annotation evaluations are generated as RDF instances of the MERM evaluation concepts (Section 2.2) and are linked to the resource RDF descriptions. To specify the evaluation criteria and values, the user uses the resource evaluation tool that is provided by the AGROPub Manager.
Figure 7 shows a snippet of the same RDF resource description discussed in the previous Section (Figure 6). The snippet contains a user evaluation of the resource’s annotation. The evaluation’s description is nested between the opening tag (Line 21) and closing tag (Line 28). Similar to the opening tags of the resource and annotation descriptions, the opening tag of the evaluation description holds the information about the evaluation’s URI that uniquely identifies the evaluation. Line 23 contains information about the evaluator. Then, Line 24 and Line 25 contain information about the evaluation criterion and evaluation metric respectively. In this concrete example, the specified evaluation criterion is Completeness and the corresponding evaluation metric is Likert scale. The last line in the evaluation description (Line 26) contains the evaluation value, which is Very good in this example.

The syntax of the resource evaluation does not defer from the annotation evaluation discussed in this example. It also contains information about the evaluator, evaluation criterion, metric and value. The description of the resource evaluation is nested within the resource description, at the same level as the resource basic info and resource annotation (Figure 6, Line 39).

6. Conclusions

This paper presents the outcome of the case study that we conducted to demonstrate the publishing of environmental resources to LOD and to verify the proposed requirements that the resource publishers should respect. The case study delivered the agro-environmental domain ontology, namely the JRC MARS ontology, and the supporting resource-publishing system called AGROPub. We described the main concepts of the ontology and explained how the ontology is mapped to the MERM model, which we use to describe the JRC MARS resources. Then, we discussed the architecture of the developed supporting system and illustrated the results of the publishing on the example of publishing the JRC MARS resources.

The case study proved the feasibility of the proposed resource-publishing approach and the underlying publishing requirements. Having in mind that improved resource discovery is one of the main objectives of the envisioned, semantically unified environmental information space, our main focus in future work will be on evaluating search and discovery of environmental resources published to LOD according to our approach. In addition, we also plan to evaluate the developed system (AGROPub) by conducting a usability evaluation of the system’s services and tools.

7. Acknowledgment

TaToo is a Collaborative Project (contract number 247893) co-funded by the Information Society and Media DG of the European Commission within the RTD activities of the Thematic Priority Information Society Technologies.
Bibliography


Nešić, S., Rizzoli, A.E., Athanasiadis, I. N. (2011): Towards a semantically unified environmental information space, in Environmental Software Systems - Frameworks of eEnvironment, 9th IFIP WG 5.11 International Symposium, ISESS 2011, Brno, Czech Republic, 2011; Springer, IFIP Advances in Information and Communication Technology Volume 359, pp.407-418;


