
Publishing agro-environmental resources as linked data

Saša Nešić* and Andrea E. Rizzoli

IDSIA – DalleMolle Institute for Artificial Intelligence,
University of Lugano, Lugano, Manno 6928, Switzerland
E-mail: sasa@idsia.ch
E-mail: andrea@idsia.ch
*Corresponding author

Ioannis N. Athanasiadis

Democritus University of Thrace,
Xanthi 67100, Greece
E-mail: ioannis@athanasiadis.info

Abstract: Publishing environmental resources to the linked environmental data requires resource publishers to conform to a set of universally recognised linked data principles. These principles, however, do not consider domain-specific semantics that are crucial for building a semantically unified environmental information space over the linked environmental data. In this paper we present an approach to publishing agro-environmental resources to the linked environmental data that extends the universal linked data principles by taking into account domain semantics of the agro-environmental domain. The approach relies on an underlying environmental resource model and introduces a set of publishing principles that should enable semantic integration of the linked agro-environmental resources. According to the proposed principles this integration happens over time as a result of conjoint acting of the resource providers and resource users. In order to prove feasibility of the approach we developed a supporting software system called AGROPub. In this paper we demonstrate three main use-cases of the system:

- resource annotation and publishing;
- resource linking;
- resource evaluation, on the example of some real-world agro-environmental resource.

The processes realised by these three AGROPub use-cases, all together, should enable the full integration of the linked agro-environmental resources into the semantically unified environmental information space.

Keywords: linked data; semantic annotation and linking; environmental resource modelling.

Reference to this paper should be made as follows: Nešić, S., Rizzoli, A.E. and Athanasiadis, I.N. (2012) 'Publishing agro-environmental resources as linked data', *Int. J. Metadata, Semantics and Ontologies*, Vol. 7, No. 1, pp.25–36.

Biographical notes: Saša Nešić is a Postdoctoral Researcher at IDSIA research institute for artificial intelligence, Lugano, Switzerland. He received his PhD degree in Computer Science from the Faculty of Informatics, University of Lugano, Switzerland in 2010. His PhD thesis focused on a design of novel semantic document architecture for desktop data integration and management. His research interests include semantic Web technologies, knowledge engineering, information retrieval, e-learning and environmental modelling. Saša has also contributed to the design of several semantic-based information software systems developed within EU founded projects.

Andrea E. Rizzoli is a Senior Researcher at IDSIA and Professor at SUPSI (University of Applied Sciences of Southern Switzerland). He holds a MSc in Control Engineering (1989) and a PhD in Control Engineering and Informatics (1993). His research interests are in modelling and simulation of dynamic systems, in the development of environmental decision support systems for natural resources management and in the application of artificial intelligence and operations research techniques to environmental problems. He has authored more than 100 scientific publications. He also serves as Editor for the International Journal 'Environmental Modelling and Software'.

Ioannis N. Athanasiadis is a Lecturer with the Electrical and Computer Engineering Department of Democritus University of Thrace, in Xanthi, Greece. His areas of expertise include software engineering, intelligent information systems, environmental informatics, software agents, decision support systems, knowledge engineering, metadata and ontologies and machine learning.

1 Introduction

The availability of high quality and up-to-date information on environmental conditions of agriculture, often referred as agro-environmental information, has always been of major interest given its relevance to the ability to secure food supply. Typical sources of agro-environmental information include raw data provided by sensor networks and satellite imaging, processed data stored in databases and catalogues, information processing units such as mathematical models, software, web services and decision support systems and scholar papers, maps and reports such as those published in the JRC MARS bulletins (MARS, 2011). Despite the fact that there is a lot of agro-environmental information currently available on the Web, only closed user communities that are aware of them, typically benefit from them. Discovering and accessing agro-environmental information effectively and efficiently still requires a deep understanding of both domain complexity and computer systems that provide them (Villa et al., 2009). Moreover, it is still a great challenge to support communication and community organisation, while providing efficient data, software and visualisation infrastructure (Porter et al., 2011); agro-environmental information that is published and managed by different authorities are still greatly disconnected and unaware of each other. As a consequence of all this the existing agro-environmental information on the Web looks more as sets of disconnected information islands than a semantically unified information space (Nesic et al., 2011a).

To narrow the discovery gap that prevents a full and easy access to relevant agro-environmental information on the Web it is necessary to develop methods, models and tools to organise domain-specific (i.e., *environmental* and *agricultural*) semantics of this information. A number of research approaches (Carr et al., 2001; Rocha et al., 2004; Anyanwn et al., 2005; Popov et al., 2003) developed over the last decade have shown that employing domain ontologies (Fensel and Harmelen, 2001; Goumopoulos et al., 2009; Bartol, 2009) is a promising solution to model and organise domain semantics (knowledge). Moreover, in order to successfully integrate agro-environmental information on the Web into thesemantically unified environmental information space, it is necessary to represent data (resources) holding this information according to universally recognised principles. The linked data principles (Bizer et al., 2009) have been recognised as one of the most promising solutions. On the basis of these principles, the linked environmental data initiative has been started recently (Ruther et al., 2010).

In the quest of the long-term vision of thesemantically unified environmental information space, we present an

agro-environmental resource publishing system, named AGROPub, which facilitates uniform integration of various agro-environmental resources to the linked environmental data. The system relies on a set of publishing principles and an underlying environmental resource model that all together ensure this uniform resource integration. The integration process involves both resource providers, who publish the resources and resource users, who utilise the resources but also contribute to the resources annotation, linking and evaluation. Besides publishing resources to the linked data, the resource providers are also responsible to set initial sets of semantic annotations and semantic links between the published resources and some related resources from the linked data. However, the real resource integration to the linked environmental data happens over time as a result of users' interaction with the resources, including the resource annotation, linking and evaluation. In addition, the system is designed as an extensible, service-oriented architecture that intends to provide all necessary services to support both agro-environmental resource providers and resource users to contribute to this vision.

The rest of the paper is organised as follows. In Section 2 we present and discuss the set of publishing principles that the AGROPub resource publishing process is based on. Section 3 introduces main concepts of the underlying environmental resource model. In Section 4 we outline the architecture design of the AGROPub system and provide some facts about the prototype implementation. Then, in Sections 5, 6 and 7 we give detailed, step-wise descriptions of the resource annotation and publishing, the resource linking and the resource evaluation processes respectively. Section 8 reports on related research efforts focusing on environmental resource modelling and the linked environmental data initiative. We conclude the paper with summary remarks and plans for future work.

2 Publishing principles

In this section we present a set of publishing principles that environmental data providers should respect in order to get their data seamlessly integrated into a global information space, usually referred as the web of linked data. The seamless integration requires environmental data to be structured, semantically annotated, interlinked with related data and open to user access and evaluation. To achieve this, environmental data has to comply with the universally recognised Linked Data principles (Berners-Lee et al., 2006) and be represented using standardised content formats. Moreover, in order to be semantically integrated into the

unified environmental information space environmental data should be semantically annotated by domain semantics with a shared common understanding in the environmental domain. Domain ontologies have been recognised as one of the Semantic Web standards and a promising solution to providing a formal definition of shared domain semantics by means of modelling domain concepts and relationships between them (Fensel and Harmelen, 2001). By complying with the linked data principles and being annotated by the ontology-based domain semantics the environmental data can be considered as an integral part of the unified environmental information space, which in turn is a part of the Web of Data. In addition, in spite of the fact that the Web of Data has attracted most of the attention after the idea of engaging machines and intelligent software agents as the data users (Hendler, 2001), it is likely that the humans will remain the most important user of the Web Data as well. Therefore, the data on the Web of Data should be easily accessible by the Web users as well as equipped by properties for handling user opinions (i.e., evaluations) about it. We now present the set of publishing principles for the environmental data (Nesic et al., 2011a) that we came up with by an appropriate combination of linked data principles, ontology-based semantic annotations and end-user data evaluations.

2.1 Data identification

The vision of the linked data builds directly on the Web architecture (Jacobs and Walsh, 2004). Therefore, considering the linked environmental data as a part of the linked data, they also need to be identified by Unique Resource Identifiers (URIs). The linked data principles suggest using HTTP dereferencable URIs and avoiding other URI schemes such as URNs (Moats, 1997) and DOIs (Page et al., 1998). The HTTP URIs provide a simple way to create globally unique names in a decentralised fashion and serve not just as a name but also as a means of accessing the identified entity. Moreover, in order to achieve not only global uniqueness of data URIs but also revealing URIs within the environmental information space it is recommended that applied HTTP URI schemas include information about the environmental sub-domain to which the data belongs (e.g., agriculture or climate) and the information about the data provider.

2.2 Data representation

Having environmental data uniquely identified by HTTP dereferencable URIs, the next principle that the environmental data providers should comply with is a standardised data format. Linked data on the Web is represented by the Resource Description Framework (RDF) data model (RDF, 1999). There are several RDF serialisation formats, two of which are standardized by W3C: RDF/XML (Beckett, 2004) and RDFa (Adida et al., 2008). RDF/XML has been the most often used format so far in spite of the fact that it is difficult for humans to read and write it. RDFa is the RDF serialisation format that mixes RDF triples and HTML. It owes its popularity mostly to the fact that it enables both human-readable and

machine-processable data representations to coexist in the same Web document.

2.3 Semantic annotation

The RDF model is a generic, abstract data representation model that does not provide elements for expressing domain semantics. On the Semantic Web modelling domain semantics has been delegated to domain ontologies, which are formally described in languages such as the RDFS (McBride et al., 2004) and OWL (McGuinness and van Harmelen, 2004). The semantic annotation of environmental data requires linking concepts from environmental domain ontologies to the RDF descriptions of the data. Such domain-specific concepts are usually referred as the semantic annotations of the data. The semantic annotations can be added to the data manually or inferred automatically by analysing the data as well as the context in which the data is used. However, regardless of the way the semantic annotations are being generated (i.e., manually or automatically) they have to adhere to shared domain ontologies in order to ensure shared understanding of the annotated data. Alternatively, if the semantic annotations refer to newly introduced domain ontology, apart from the actual ontology the data provider has also to provide an appropriate set of ontology mappings to existing well-established domain ontologies.

2.4 Semantic linking

The last step in publishing environmental data to the unified environmental information space is setting an initial set of explicit semantic links between the newly published data and semantically related data available on the Web. This initial set of semantic links is supposed to act as an activator for the linking process that should last as long as the data is available on the Web. We refer to the links as explicit as there should be explicit RDF descriptions of them. Moreover, the links are semantic because they are established over typed relationships formally defined by ontological properties. The ontologies that define these linking properties can be the same domain ontologies as the ones used for the semantic annotation but also upper ontologies such as SKOS (SKOS, 2009). What should be important when setting the initial set of semantic links for the newly published data is to identify related data that is well interlinked on the Web. This ensures that the data will be discovered and further interlinked. Similar to semantic annotations, semantic links can be created manually by humans or inferred by intelligent software agents, which crawl the Web and infer semantic relationships between data.

2.5 Data evaluation

After the initial set of the semantic links has been assigned, the data publishing process is over. However, there is still one more requirement that environmental data providers should respect if they want their data to meaningfully contribute to the linked data vision. Besides the actual data, they should decide on models and representation formats to handle user evaluations of their data. Similar to semantic annotations and

links, data evaluations should also be represented as structured formally defined information. Following the same analogy as for semantic annotations and semantic links, ontologies could be a solution to formally describe data evaluations. In this case data evaluations could be serialised in one of the RDF serialisation formats and also considered as a part of the linked data corpus.

In the following Section 4 we present the underlying environmental resource model that enables the above discussed publishing principles. Then in Section 5 we describe a software system that provides services and tools necessary for realising the publishing approach based on the proposed publishing principles and the underlying environmental resource model.

3 The underlying environmental resource model

In order to publish linked environmental data according to the presented publishing principles (Section 3), we need an appropriate environmental resource model. The model should provide structures (i.e., entities and relationships) capable of handling information about data identification, semantic annotations, semantic links and evaluations according to the publishing principles described in Section 2. Moreover, the model should be serializable to some of the RDF serialisation formats. Among a number of existing environmental resource models, some of which we discussed in Section 2, we have selected the Minimal Environmental Resource Model (MERM) (Pariente et al., 2011) as a model that could suit best the proposed publishing principles. However, the original model did not completely match all requirements from the publishing principles, so we needed to extend it. In the rest of the section we first introduce and present the core concepts and properties of the model and then describe the extensions that we added.

3.1 MERM Core

MERM is formally described by the MERM ontology. Figure 1(a) illustrates the core concepts and properties of the ontology. The central concept in the ontology is the *merm:Resource* concept representing an environmental resource. Firstly, each environmental resource is described by a set of basic information such as the resource author, owner and creation

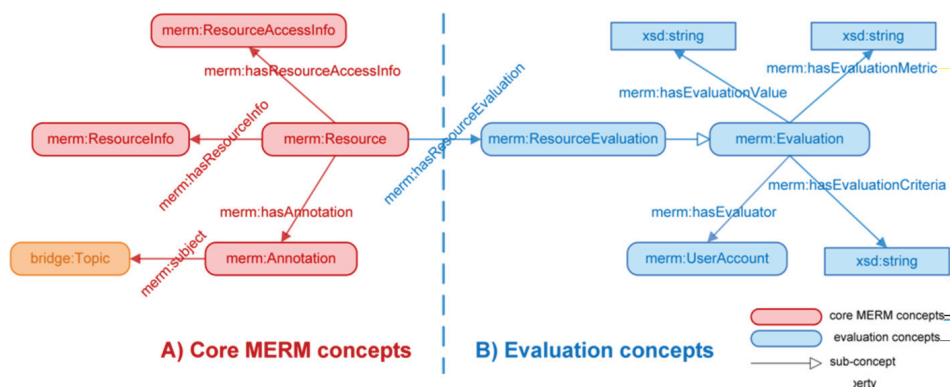
date, which are modelled by the *merm:ResourceInfo* concept and its properties. Secondly, each environmental resource that is represented by this concept is characterised by information of how to access the resource, which is modelled by the *merm:ResourceAccessInfo* concept and its properties. The resource access information can be very heterogeneous, ranging from a simple URL to a complex WSDL depending on the nature of the resource. Finally, each environmental resource can have a number of resource annotations, modelled by the *merm:Annotation* concept. The MERM resource annotation provides a description of the resource including a type of the resource (e.g., web service, document and software library) and (environmental) domain semantics characterising the resource. The domain semantics, as defined in MERM, are represented by concepts from environmental domain ontologies thus being completely in line with the semantic annotations defined in the our publishing principles. In the MERM domain semantics are linked to the resource annotation via *merm:subject* property. In order to ensure uniform interpretation of the domain semantics defined in different domain ontologies, MERM utilizes the *bridge:Topic* concept defined in the bridge ontology (Pariente et al., 2011). The role of the *bridge:Topic* is to mirror domain concepts to valid subject for the MERM resource annotation. In other words, in order to be added to the resource annotation domain concepts have to be align to the *bridge:Topic* concept (e.g., by using *owl:equivalentClass* or *owl:sameAs* OWL constructs) or sub-conceptualised from it (e.g., by using the *rdfs:subClassOf* RDF construct).

3.2 MERM extensions

The core MERM concepts and properties are capable of handling the information about the resource identification and the resource semantic annotations. As the publishing principles ask also for handling the semantic links and data evaluations we needed to extend MERM and add support for them too.

Identifying appropriate ontological properties that model semantic relationships between environmental data is a key issue in creating the semantic links. These properties actually models semantics of the links to be created. We refer to them in the extended MERM model as the linking properties. As the Semantic Web best practice (Berners-Lee et al., 2001), it is recommended

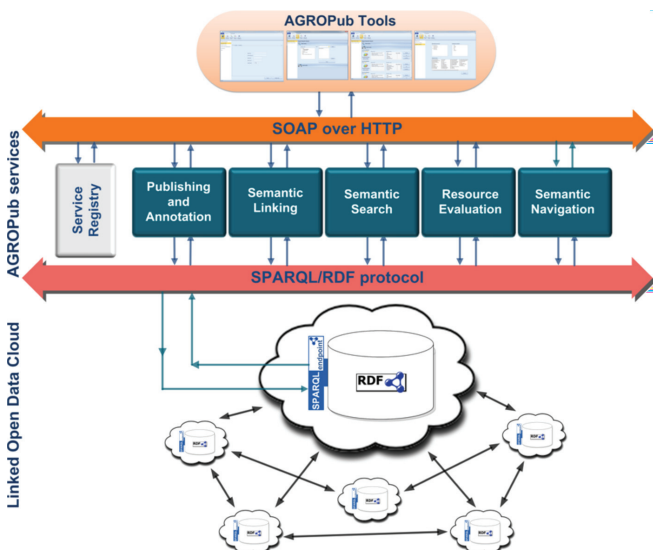
Figure 1 Extended MERM model: (A) core concepts and (B) evaluation concepts (see online version for colours)



that linking properties are taken from existing, well-established vocabularies (ontologies) rather than being newly introduced properties. However, if environmental data providers cannot find suitable properties among the existing ones then they should be able to introduce new properties by providing their formal ontological definitions. This actually means that the underlying environmental resource model should support a non-exhaustive list of the linking properties. Currently, we added an initial predefined list of the linking properties defined in the SKOS vocabulary: the *skos:related* property that can be used for general associative links between environmental resources, the *skos:broader* and *skos:narrower* properties that can be used for setting hierarchical links between environmental resources, the *skos:closeMatch* property that can be used to link sufficiently similar environmental resources and the *skos:exactMatch* property to link mutually equivalent environmental resources.

The quality of published linked environmental data can be assessed with respect to a number of evaluation criteria, such as *usefulness*, *accuracy* and *completeness*, each of which is determined by a number of different indicators. Figure 2(b) shows our extension of the MERM model providing concepts and properties for modelling user evaluations of the linked environmental resources. As we can see from the figure, the resource evaluation is modelled by the *merm:Re-sourceEvaluation* concept and its properties: *merm:evaluator* holding information about the user providing the evaluation, *merm:evaluationDate* holding information about the evaluation date, *merm:evaluationCriterion* holding information about the evaluation criterion, *merm:evaluationMetric* holding the information about the evaluation metric that is applied to the chosen evaluation criterion and *merm:evaluationValue* holding the evaluation value that must be valid against the evaluation metric. The resource evaluations are linked to the instances of the *merm:resource* via *merm:hasEvaluation* property.

Figure 2 Illustration of the AGROPub architecture (see online version for colours)



Both MERM core and MERM extensions comply with the RDF model so that their instances can be serialised to one of the RDF serialisation formats. This means that the extended MERM model is also in accordance with the data representation principle of the proposed publishing principles.

4 AGROPub: agro-environmental resource publisher

In order to support agro-environmental data providers to publish their data to the linked environmental data according to the proposed publishing principles and by applying the underlying environmental resource model, we have developed a supporting software system: the agro-environmental resource publishing system or, in short, AGROPub. Even though the system is dedicated primarily to agro-environmental data providers, it is designed according to the publishing principles that apply to publishing of any kind of environmental data. The adjustment of the system in order to support publishing environmental data other than the agro-environmental data is just a matter of using appropriate domain ontologies and potential small changes of the AGROPub user interface. Moreover, apart from supporting the publishing process AGROPub also provides functionalities that enable environmental data users to search and navigate across the linked environmental data. However, since our primary focus in this study is on the publishing we do not further elaborate the AGROPub search and navigation functionalities. In the rest of this section we first present the overall architecture of the system and then provide some facts about the prototype implementation.

We designed the AGROPub system as a three-tier, service oriented architecture (Fig. 2). The tiers are a *data tier*, a *service tier* and a *presentation tier*.

The *data tier* contains an RDF repository storing RDF instances (descriptions) of the published agro-environmental resources. The repository exposes an HTTP de-referenceable SPARQL endpoint, so that the RDF instances of the published resources can be referred and linked to other linked data on the Web.

The *service tier* provides the functionalities of the AGROPub system, which are implemented by a number of services. The core functionalities of the system are implemented by the following five services:

- 1 Semantic Annotation and Publishing;
- 2 Semantic Linking;
- 3 Resource Evaluation;
- 4 Semantic Search;
- 5 Semantic Navigation.

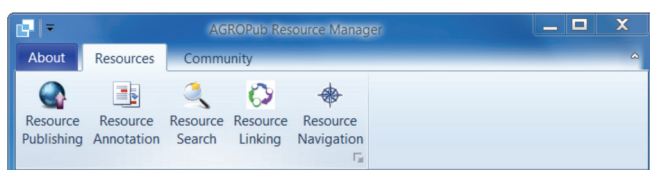
All the services are designed as the SOAP-based Web services, all of which expose a standardised Web interface. Besides the core services, the service tier can be extended by new services, which have to comply with the AGROPub design principles and have to be registered to the service

registry. The service registry provides registration and look-up functionalities for the AGROPub services as well as methods to enable and disable services in the system.

The *presentation tier* is a top tier in the AGROPub architecture that provides the user interfaces for the AGROPub services. It is technology- and platform-independent according to the service-oriented nature of AGROPub. It may contain different types of end-user applications, including Web-based, desktop and mobile phone applications.

In order to demonstrate the validity of the proposed architecture and to enable evaluations that would validate the proposed publishing principles, we have implemented the AGROPub prototype. The prototype has gone through several iterations in its development, following the incremental refining of the architecture's design. The current version of the prototype is a fully functional piece of software developed and available as an open source project (AGROPub, 2011). The prototype provides the implementation of all three tiers from the AGROPub architecture. The AGROPub repository is built on the top of the Sesame OpenRDF repository (Sesame, 2011). The AGROPub services are implemented as .NET WCF services (WCF, 2011). We also developed several end-user tools and encapsulated them in a desktop application suite called AGROPub Manager. Figure 3 shows a screenshot of the AGROPub Manager's main menu ribbon tab. As we can see from the figure, there are five tools that can be triggered directly from the main application menu: Resource Publishing, Resource Annotation, Resource Linking, Resource Search and Resource Navigation. The Resource Publishing tool enables resource providers to publish their resources to the linked environmental data by specifying information about the resource identification and access and providing an initial set of the semantic annotations to the resource. The Resource Annotation tool enables both resource providers and resource consumers to add new semantic annotations to environmental resources already published to the linked environmental data. Moreover, the Resource Linking tool enables AGROPub users to set the semantic links between related environmental resources from the linked data. Finally, the Search and Navigation tools enable users to search and navigate the linked environmental resources based on their semantic annotations and semantic links respectively. In addition, besides these five tools that can be invoked directly from the application's main menu, AGROPub Manager provides a tool that enables browsing of the resource content and semantics (i.e., semantic annotations) and a tool for the resource evaluation. Currently, these two tools can be triggered from the list of the search and navigation results.

Figure 3 The main menu of the AGROPub manager (see online version for colours)



In the following Sections 6–8 we demonstrate the AGROPub functionalities (i.e., services and tools) that are involved in the resource annotation and publishing, the resource linking and the resource evaluation processes respectively.

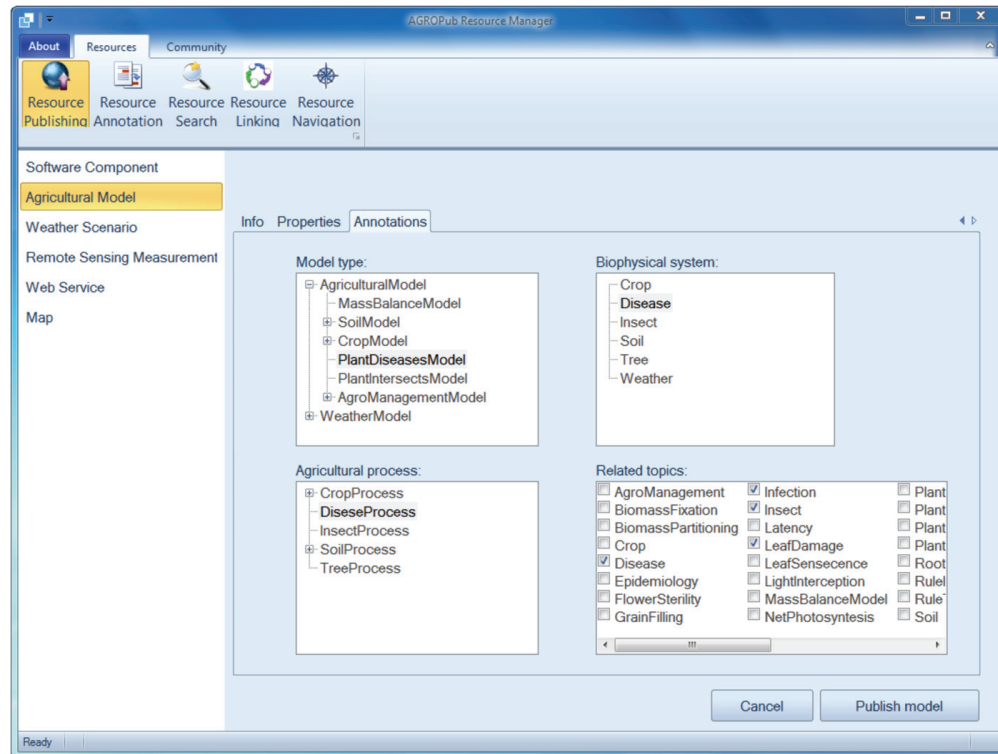
5 Resource annotation and publishing by AGROPub

In this section we demonstrate the resource-publishing process realised by the AGROPub system. Since the initial semantic annotation of the environmental resource is performed as a part of the publishing process, we consider the semantic annotation in this section too. However, we want to point out that once the resource has been published to the linked environmental data everybody who discovers the resource can annotate it with an arbitrary number of new semantic annotations.

The resource publishing starts by the user clicking on the Resource Publishing button from the main menu of AGROPub Manager. The user is then presented a new window with two panes (Fig. 4). The left pane contains a navigation bar with a list of supported agro-environmental resource types (e.g., agricultural models, agricultural software and agricultural maps). That list currently contains the resource types defined in the MARS JRC ontology (Nesic et al., 2011b), which is an agro-environmental domain ontology that we use for semantic annotation of resources considered in this study. The AGROPub system and the underlying environmental resource model do not restrict the publishing process to any specific domain ontology. However, in the current prototype at the user interface level there is no support for adding an arbitrary domain ontology that could afterwards be used for the resource annotation. Adding this support will be one of our priorities in the next version of the prototype. Let us now go back to the publishing process. After selecting a resource type, the right pane of the application window (Fig. 4) displays the resource-publishing control composed of three tabs, each of which corresponds to one step of the publishing process. We now briefly explain each of the three steps on an example of publishing an agro-environmental resource of the agricultural model type.

5.1 Specifying general resource info

The first step in publishing the resource is to specify a set of general information about the resource including the resource's URI, author, owner, creation date and the URL of the Web page describing the resource. This set of information corresponds to the information held by the properties of the *merm:resourceInfo* and *merm:resourceAccessInfo* concepts defined in the underlying model (Section 4). The only mandatory information to be provided in this step is the resource URI. The resource URI must be globally unique, which is ensured by the system validating the entered resource URI against an HTTP based URI schema proposed in (Nesic et al., 2011a). In case of resources whose content is accessible through HTTP, the resource URL and URI are identical.

Figure 4 Resource annotation and publishing by the AGROPub manager (see online version for colours)

5.2 Specifying type-specific resource properties

In this step the user specifies values for the properties that are specific to the selected resource type. These properties are defined in the same domain ontology that provides the specified resource type. In case of the considered resource example, that is, the agricultural model resource type among other properties the user should specify the input and output parameters of the agricultural model, the agricultural process that the model describes and the software components that implement the model.

5.3 Specifying semantic annotations

This is the last step in the publishing process in which the user specifies the ontology-based semantic annotations for the resource to be published. The user selects concepts from the domain ontology supported by the AGROPub system. These concepts will be then added to the resource as values of the *merm:subject* property defined in the resource model. In case of the resource example that we consider here and the JRC ontology, which is supported by AGROPub, the user can select concepts from four sets of concepts. The right pane of the resource annotation window (Fig. 4) contains four boxes, each of which containing a set of concepts from the JRC ontology describing specific aspects of the selected resource type. In case of the agricultural model resource type those are: model types (e.g., soil model, crop model and plant disease model) related agro-physical systems (e.g., crop, soil and weather), related agricultural process (e.g., biomass fixation, drainage and grain-filling) and related agricultural topics (e.g., leaf disease, insect, latency and flower sterility).

Having passed through these three steps, the user finishes with the specification of the resource description and can initiate the resource publishing. By clicking on the 'Publish model' button, the AGROPub Manager collects the data specified in the three steps and calls the AGROPub publishing and annotation service. The service then generates the RDF description (i.e., a set of RDF triples) of the resource, according to the underlying environmental resource model and stores it into the AGROPub RDF repository. When the service finishes with the publishing, it notifies the user by sending back a confirmation message. If the user attempts to publish a resource that has already been published, the service sends an appropriate warning message. Considering that each resource published to the linked data has to be unique, it is not possible to publish the same resource (i.e., the resource identified by the same URI) twice. Figure 5 exhibits a snippet of the RDF resource description generated by the AGROPub system for the considered example resource. Let us now briefly describe this RDF snippet. *Line 2* of the snippet contains the opening tag of the resource's RDF description that holds the information about the resource's URI. *Lines 3–12* contain a set of basic resource info, including the resource's creator, owner, description Web page, creation date, publisher and publication date. Then, *Lines 13–38* contain the resource's semantic annotations. First, *Line 14* shows that the published resource is of the *jrc:PlantDiseasesModel*. *Line 15* and *Line 16* hold information about the annotation's creator and creation date respectively. *Line 17* and *Line 18* hold information about the related biophysical systems and agricultural processes respectively and finally *Lines 19, 20 and 21* hold related domain specific (i.e., agricultural)

Figure 5 A snippet of the RDF resource description generated by AGROPub (see online version for colours)

```

2 <merm:Resource rdf:about="http://www.ape.it/mars01/res#33">
3   <merm:supports>
4     <rdf:Description rdf:about="http://mars.ies.jrc.eu/software/res#info_63444700_0">
5       <dc:creator>Marcello Donatelli</dc:creator>
6       <dc:owner>Andrea Rizzoli</dc:owner>
7       <dc:description>http://www.apesimulator.it/</dc:description>
8       <dc:date>6/7/2011 12:00:00 AM</dc:date>
9       <merm:Publisher>Sasa Nesic</merm:Publisher>
10      <merm:PublicationDate>11/18/2011 3:37:07 PM</merm:PublicationDate>
11    </rdf:Description>
12  </merm:supports>
13  <merm:hasAnnotation>
14    <jrc:PlantDiseasesModel rdf:about="http://mars.ies.jrc.eu/res#annotation_63444700_1">
15      <dc:creator>Sasa Nesic</dc:creator>
16      <dc:date>6/28/2011 3:11:22 PM</dc:date>
17      <jrc:BiophysicalSystem rdf:resource="http://mars.ies.jrc.eu/JRC_MARS_physical_domain.owl#Disease"/>
18      <jrc:AgriculturalProcess rdf:resource="http://mars.ies.jrc.eu/JRC_MARS_physical_domain.owl#DiseaseProcess"/>
19      <merm:subject rdf:resource="http://mars.ies.jrc.eu/JRC_MARS_physical_domain.owl#Infection"/>
20      <merm:subject rdf:resource="http://mars.ies.jrc.eu/JRC_MARS_physical_domain.owl#Insect"/>
21      <merm:subject rdf:resource="http://mars.ies.jrc.eu/JRC_MARS_physical_domain.owl#LeafDamage"/>
22    </jrc:PlantDiseasesModel>
23  </merm:hasAnnotation>
24  <skos:related rdf:resource="http://www.ape.it/mars01/res#91"/>
25  <merm:hasResourceEvaluation>...</merm:hasResourceEvaluation>
33 </merm:Resource>

```

topics defined in the JRC MARS ontology. Lines 24 to 32 contain RDF descriptions of the resource semantic link and the resource evaluation that we describe in details in the next two sections. Finally, Line 33 contains the closing tag of the resource's RDF description.

As we pointed out at the very beginning of this section, the semantic annotations can be added to a resource not only during the resource publishing but also after that. The AGROPub system, actually, enables everyone who wants to add semantic annotation to the linked environmental resources to take advantage of the annotation functionalities provided by the publishing and annotation service. The annotation process do not defer from the above-described third step of the publishing process. The annotation user interface is provided as a part of the AGROPub Annotation tool that is a part of the AGROPub Manager.

6 Resource linking by AGROPub

After having a resource published to the linked environmental data, according to the proposed publishing principles, it is recommended that the resource provider also provides an initial set of semantic links connecting the newly published resource to some related resources from the linked environmental data. Along with the initial semantic annotations, which were specified during the resource publishing, the initial set of semantic links will further enhance the resource discovery and foster integration of the resource within the linked environmental data. The semantic links that we consider in this section are the links defined by the linking properties from the underlying environmental resource model. Moreover, these links are supposed to be added manually by using the AGROPub system. We now explain the linking process considering again the same example resource as the one considered for the resource publishing.

The resource linking user interface is triggered from the detailed resource view (Fig. 6). Initially, the user interface

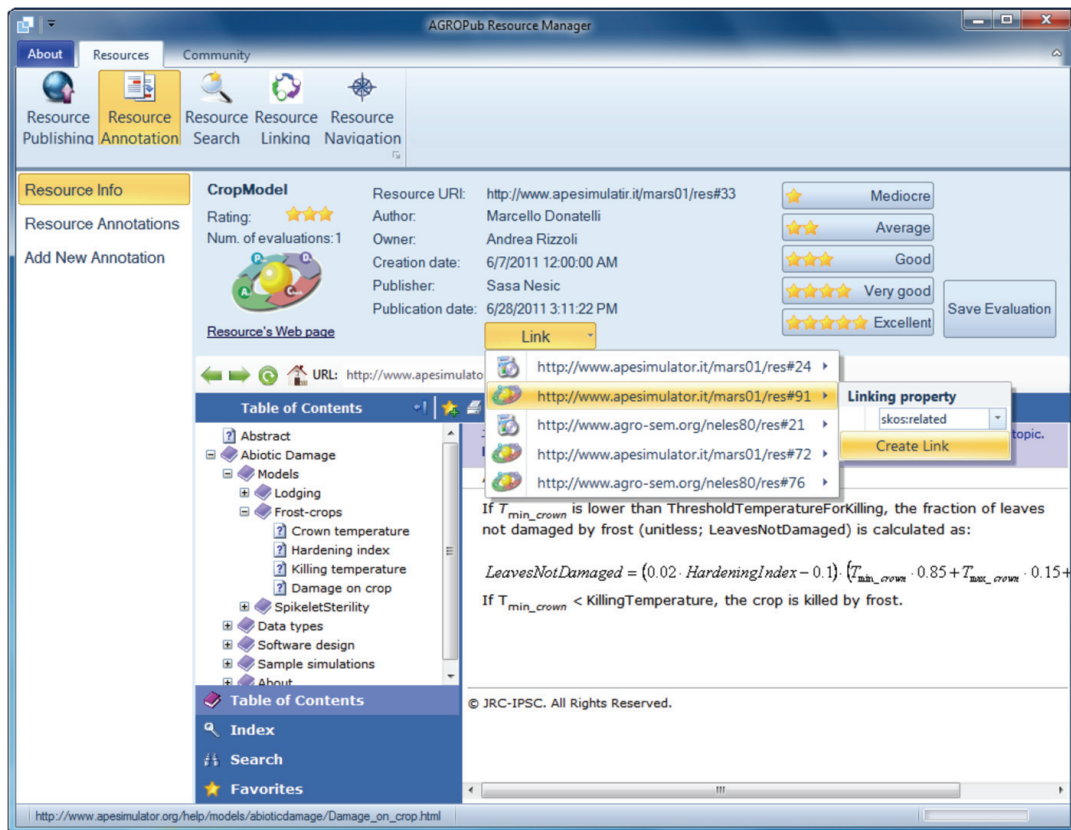
element (i.e., a pop-up list) that should provide a list of related resources for linking is empty. This is because the system does not know in advance what other resources the user thinks of as related to the resource to be linked. Therefore, the linking process requires some prior exploration (searching and navigation) of the linked resources and caching those resources that could be considered for the linking later on. We now go step-by-step through the linking scenario starting from the resource detail view. This first step could be omitted if the user already knows the resource she or he wants to link.

6.1 Browsing resource details

The AGROPub Manager offers two ways to request and browse details of the resource published to the linked environmental data. The first way is to type in the resource URI and the second one is to select the resource URI from the list of obtained semantic search and navigation results. In case of setting the initial set of links by the resource provider as a post-publishing task, the first way is a proper one. As we can see from the Figure 6, the resource-linking user interface is realised as a cascading popup with two cascades. The first cascade contains a combo-box filled with a list of available linking properties. The second cascade contains a list of cached resources, which is initially empty.

6.2 Caching related resources

After observing the resource details the user needs to find resources that she or he might consider for linking afterwards. Therefore, the next step in the linking scenario is to explore available linked environmental resources and to cache potential candidates for linking. To perform this step the user utilises the search and navigation utilities of the AGROPub system. When the user identifies potential related resources, the AGROPub Manager enables him or her to cache the resource URIs into the application's internal memory. The user can keep on searching and navigating among available

Figure 6 Resource linking by the AGROPub manager (see online version for colours)

resources as long as she or he caches a satisfactory number of related resources.

6.3 Selecting a linking property

Having cached related resources the user moves back to the resource detailed view and opens the linking popup. Next, he needs to select the linking property over which the semantic link will be established. Currently, the AGROPub system comes with the predefined list of linking properties, which come from the underlining environmental resource model (Section 4.2). For the next version of the AGROPub prototype we plan to enable users to use custom linking properties added to the system in a similar way as the users will be able to add some custom domain ontologies for the semantic annotation.

6.4 Selecting a resource

The last step in the linking process is to select a related resource from the list of cached resources. After the user selects the resource and clicks on the 'Create Link' button the AGROPub Manager collect the information from the linking user interface and calls the AGROPub semantic linking service.

The semantic linking service generates and stores the semantic link (i.e., the RDF triple) in the AGROPub RDF repository. In case the link between the two resources had already existed in the repository, it will not be generated and stored again. Let us now go back to the RDF snippet of the

considered resource example (Fig. 5). *Line 24* exhibits the RDF triple of a semantic link added to the resource. As we can see from the figure, the semantic link links the considered resource and the resource identified by the `http://www.apesimulato.it/mars01/res#91` URI via the `skos:related` linking property.

7 Resource evaluation by AGROPub

The AGROPub system provides the Resource Evaluation service that generates the resource evaluations based on the evaluation information, which corresponds to the evaluation schema defined within the underlying resource model (Section 4.2). To specify this information the user accesses the resource evaluation interface that is available as a part of the resource detail view (Fig. 6, upper right corner) of the AGROPub Manager. After receiving the evaluation information from the Manager, the Resource Evaluation service generates and stores the evaluations (i.e., RDF instances of the resource evaluation schema) to the AGROPub RDF repository.

Figure 7 shows the resource evaluation RDF snippet of the same example resource that has been considered so far. The evaluation's description is nested between the evaluation opening (*Line 26*) and closing (*Line 31*) tags. Similar to the opening tags of the whole resource description and the resource's annotation description, the opening tag of the evaluation description holds the information about the evaluation's URI that uniquely identifies the evaluation. *Line 27* contains information about the evaluator. Then, *Line*

Figure 7 A snippet of the RDF resource evaluation generated by AGROPub (see online version for colours)

```

25 <merm:hasResourceEvaluation>
26   <rdf:Description rdf:about="http://mars.ies.jrc.eu/model/res#resource_evaluation_63444700_2">
27     <merm:hasEvaluator>Sasa Nesic</merm:hasEvaluator>
28     <merm:hasEvaluationCriteria>usefulness</merm:hasEvaluationCriteria>
29     <merm:hasEvaluationMetric>5-point Likert scale</merm:hasEvaluationMetric>
30     <merm:hasEvaluationValue>Very good</merm:hasEvaluationValue>
31   </rdf:Description>
32 </merm:hasResourceEvaluation>

```

28 and Line 29 contain information about the evaluation criterion and evaluation metric respectively. In this concrete example, the specified evaluation criterion is *usefulness* and the corresponding evaluation metric is a *5-point Likert scale*. The last line in the evaluation description (Line 30) contains the evaluation value, which is *Very good* in this example.

8 Related work

While there are several generic-purpose tools for annotating web resources such as Annotea Project (2011) and Delicious (2011) there has been a transition towards domain-specific tools that are able to incorporate functionality required by certain communities. For example, SemTag and Seeker have demonstrated how to perform automated semantic tagging of large corpora (Dill et al., 2003). Connotea (2011) and Mendeley (2011) focus on the scientific publications lifecycle. Confolio (2011) have been used for annotating and organising educational resources (Ebner et al., 2009). Protonotarios et al. (2011) have introduced a methodology for content integration from various agricultural repositories to a single network, by suggesting a common metadata vocabulary and a uniform method for the different steps of the process.

Related to the agro-environmental domain, there are several services that have been recently deployed, as those for open sensor architectures (Uslander et al., 2010), reusable geospatial services (Granell et al., 2011) and modelling water resources (Goodall et al., 2011). Lately there has been an institutional shift towards Linked Open Data. UN Food and Agriculture Organisation has published its multilingual agricultural vocabulary (FAO, 2011) (AGROVOC) into Linked Open Data (Caracciolo et al., 2011). Also, the European Environmental Agency has started shifting their data assimilation tasks using RDF (Roug, 2011). Similarly governmental data in the US and the UK are published as Linked Open Data (see data.gov and data.gov.uk projects). US EPA collections are available in RDF through data.gov portal.

Despite this trend, community involvement either in the form of feedback from the experts or as comments from the public is not yet incorporated in the process. There are not any services or tools available for annotating and evaluating such linked open data services. This need is addressed by AGROPub system presented in this paper, which aims to narrow the search and discovery gap of agro-environmental information.

9 Conclusions

In this paper we presented an approach for publishing agro-environmental resources to the linked environmental data, which is a result of our efforts to facilitate integration of agro-environmental information into the semantically unified environmental information space. The approach is based on the underlying environmental resource model and conforms to the set of publishing principles. We designed and developed a prototype software system called AGROPub, which realizes the proposed approach. The system provides services and tools enabling both resource providers and resource users to participate in the integration of agro-environmental resources to the linked environmental data. In the proposed approach this integration is supposed to happen over time as a result of conjoint acting of resource providers and resource users.

By having developed the AGROPub system and by publishing some example resources, we showed the feasibility of the approach. However, we are aware that potential success of the approach depends on real benefits it brings to agro-environmental resource users. In other words, the approach will be accepted and settled in the community as long as it aids the agro-environmental resource users to commit their tasks more efficiently and effectively. Therefore, one of our main objectives in the future will be to conduct a comprehensive usability evaluation of the proposed approach and the supporting software system.

References

- Adida, B., Herman, I. and Sporny, M. (2008) *RDFa Primer*, Available at: <http://www.w3.org/TR/xhtml-rdfa-primer/> (Accessed 24 November 2011).
- Agropub (2011) Available at: <http://sourceforge.net/p/agropub/> (Accessed 24 November 2011).
- Annotea Project (2011) Available at: <http://www.w3.org/2001/Annotea/> (Accessed 24 November 2011).
- Anyanwn, K., Maduko, A. and Sheth, A.P. (2005) 'Semrank: ranking complex relationship search results on the semantic web', *Proceedings of the 14th International WWW Conference*, Chiba, Japan, pp.117–127.
- Athanasiadis, I.N., Rizzoli, A.E., Donatelli, M. and Carlini, L. (2011) 'Enriching environmental software model interfaces through ontology-based tools', *International Journal of Applied Systemic Studies*, Vol. 4, Nos. 1–2, pp.94–105.

- Bartol T. (2009) 'Assessment of classification and indexing of agricultural journal based on metadata in AGRIS and CAB abstracts databases', *Int. Journal of Metadata, Semantics and Ontologies*, Vol. 4, pp.4–12.
- Beckett, D. (2004) *RDF/XML Syntax Specification*, Available at: <http://www.w3.org/TR/REC-rdf-syntax/> (Accessed 24 November 2011).
- Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N. and Weitzner, D.J. (2006) 'A Framework for Web Science', *Foundation and Trends in Web Science*, Vol. 1, No. 1, pp.1–130.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The semantic web', *Scientific America*, pp.29–37.
- Bizer, C., Heath, T. and Berners-Lee, T. (2009) 'Linked data – the story so far', *Int. Journal on Semantic Web and Information Systems*, Vol. 5, No. 3, pp.1–22.
- Caracciolo, C., Morshed, A., Stellato, A., Johannsen, G., Jaques, Y. and Keizer, J. (2011) 'Thesaurus maintenance, alignment and publication as linked data: the AGROOVOC use case', *Metadata and Semantic Research, CCIS*, Springer-Verlag, Vol. 240, pp.489–499.
- Carr, L., Bechhofer, S., Goble, C. and Hall, W. (2001) 'Conceptual linking: ontology-based open hypermedia', *Proceedings of the 10th International Conference on World Wide Web (WWW '01)*, pp.334–342.
- Confolio (2011) Available at: <http://www.confolio.org> (Accessed 24 November 2011).
- Connotea (2011) Available at: <http://www.connotea.org/> (Accessed 24 November 2011).
- Delicious (2011) Available at: <http://www.delicious.com/> (Accessed 24 November 2011).
- Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., Rajagopalan, S., Tomkins, A., Tomlin, J.A. and Zien, J.Y. (2003) 'Semtag and seeker: bootstrapping the semantic web via automated semantic annotation', *Proceedings of the 12th International Conference on World Wide Web (WWW '03)*, Budapest, Hungary, pp.178–186.
- Ebner, H., Manouselis, N., Palmer, M., Enoksson, F., Palavitsinis, N., Kastrantas, K. and Naeve, A. (2009) 'Learning object annotation for agricultural learning repositories', *Proceedings of 9th Int. Conference on Advanced Learning Technologies (ICALT 2009)*, Riga, Latvia, pp.438–442.
- FAO (2011) Available at: <http://aims.fao.org/standards/agrovoc/linked-open-data> (Accessed November 24, 2011).
- Fensel, D. and Harmelen, F. (2001) 'An ontology infrastructure for the semantic web', *IEEE Intelligent Systems*, Vol. 16, No. 2, pp.37–46.
- Goodall, J.L., Robinson, B.F. and Castronova, A.M. (2011) 'Modelling water resource systems using a service-oriented computing paradigm', *Environmental Modelling and Software*, Vol. 26, pp.573–582.
- Goumopoulos, C., Kameas, A.D. and Cassells, A. (2009) 'An ontology-driven system architecture for precision agriculture applications', *Int. Journal of Metadata, Semantics and Ontologies*, Vol. 4, Nos. 1–2, pp.72–84.
- Granell, C., Diaz, L. and Gould, M. (2010) 'Service-oriented applications for environmental models: reusable geospatial services', *Environmental Modelling and Software*, Vol. 25, No. 2, pp.182–198.
- Hendler, J. (2001) 'Agents and the semantic web', *IEEE Intelligent Systems*, Vol. 16, No. 2, pp.30–37.
- Jacobs, I. and Walsh, N. (2004) *Architecture of the World Wide Web*, Available at: <http://www.w3.org/TR/webarch/>, 2004.
- MARS (2011) *MARS Bulletins and Publications*, Available at: <http://mars.jrc.ec.europa.eu/mars/Bulletins-Publications> (Accessed 24 November 2011).
- McBride, B., Brickley, D. and Guha, R.V. (2004) *RDFS – RDF Schema*, Available at: <http://www.w3.org/TR/rdf-schema/> (Accessed 24 November 2011).
- McGuinness, D.L. and van Harmelen, F. (2004) *OWL Web Ontology Language*, Available at: <http://www.w3.org/TR/owl-features/> (Accessed 24 November 2011).
- Mendeley (2011) Available at: <http://www.mendeley.com/> (Accessed 24 November 2011).
- Moats, R. (1997) *Rfc 2141: Urn Syntax*, Available at: <http://tools.ietf.org/html/rfc2141> (Accessed 24 November 2011).
- Nesic, S., Rizzoli, A.E. and Athanasiadis, I.N. (2011a) 'Towards a semantically unified environmental information space', *Proceedings of the International Symposium on Environmental Software Systems – ISESS 2011*, Springer, Vol. 359, pp.407–418.
- Nesic, S., Rizzoli, A.E., Athanasiadis, I.N. and Donatelli, M. (2011b) 'Publishing agro-environmental data to a semantically unified environmental information space: a case study', *Proceedings of the 25th International Conference on Informatics for Environmental Protection – EnviroInfo 2011*, Vol. 2, pp.856–867.
- Page, L., Brin, S., Motwani, R. and Winograd, T. (1998) *The Pagerank Citation Ranking: Bringing Order to the Web*, Technical report, Stanford Digital Library Technologies Project, Stanford, California, USA.
- Pariente, T., Fuentes, J.M., Sanguino, M.A., Yurtsever, S., Avellino, G., Rizzoli, A.E. and Nesic, S. (2011) 'A model for semantic annotation of environmental resources: the tattoo semantic framework', *Proceedings of the International Symposium on Environmental Software Systems (ISESS 2011)*, Springer, Vol. 359, pp.419–427.
- Popov, B., Kiryakov, A., Manov, D., Ognyanoff, D. and Goranov, M. (2003) 'KIM – semantic annotation platform', *Proceedings of the 2nd International Semantic Web Conference*, Florida, USA, pp.834–849.
- Porter, C., Janssen, S. and Athanasiadis, I.N. (2011) 'Software challenges for the agricultural model intercomparison project', *3rd International Workshop on Software Research and Climate Change, Proceedings of ECOOP 2011*, Lancaster, UK.
- Protonotarios, V., Gavrilut, L., Athanasiadis, I.N., Hatzakis, I. and Sicilia, M.A. (2011) 'Introducing a content integration process for a federation of agricultural institutional repositories', *Metadata and Semantics Research, CCIS*, Springer-Verlag, Vol. 240, pp.467–477.
- RDF (1999) *RDF – Resource Description Framework*, Available at: <http://www.w3.org/RDF/> (Accessed 24 November 2011).
- Rocha, C., Schwabe, D. and Aragao, M.P. (2004) 'A hybrid approach for searching in the semantic web', *Proceedings of the 13th International WWW Conference*, New York, USA, pp.374–383.
- Roug, S. (2011) 'EEA and linked environment data – a status report', *In EnviroInfo 2011. Shaker Verlag*, Ispira, Italy, pp.380–389.

- Ruther, M., Bandholtz, T. and Logean, A. (2010) ‘Linked environmental data for the life sciences’, *Proceedings of the Workshop on Semantic Web Applications and Tools for Life Sciences*, Berlin, Germany, Vol. 698.
- Sesame (2011) *Sesame OpenRDF Project*, Available at: <http://www.openrdf.org/> (Accessed 24 November 2011).
- SKOS (2009) *SKOS – Simple Knowledge Organization System*, Available at: <http://www.w3.org/2009/08/skos-reference/skos.html> (Accessed 24 November 2011).
- Uslander, T., Jacques, P., Simonis, I. and Watson, K. (2010) ‘Designing environmental software applications based upon an open sensor service architecture’, *Environmental Modelling and Software*, Vol. 25, No. 9, pp.977–987.
- Villa, F., Athanasiadis, I.N. and Rizzoli, A.E. (2009) ‘Modelling with knowledge: a review of emerging semantic approaches to environmental modelling’, *Environmental Modelling and Software*, Vol. 24, No. 5, pp.577–587.
- WCF (2011) *Windows Communication Foundation*, Available at: <http://msdn.microsoft.com/en-us/library/dd456779.aspx> (Accessed 24 November 2011).