

# Publishing and Linking Semantically Annotated Agro-Environmental resources to LOD with AGROPub

Saša Nešić<sup>1</sup>, Andrea Emilio Rizzoli<sup>1</sup>, and Ioannis N. Athanasiadis<sup>2</sup>

<sup>1</sup>IDSIA, Manno, Switzerland  
{sasa, andrea}@idsia.ch

<sup>2</sup>Democritus University of Thrace, Xanthi, Greece  
ioannis@athanasiadis.info

**Abstract.** Publishing agro-environmental resources to a linked open data (LOD) cloud requires publishers to adopt a set of universally recognized linked data principles. These principles, along with semantic annotations based on shared domain ontologies can ensure the semantic integration of agro-environmental resources. In this paper we present a resource-publishing system, called AGROPub, that we developed to aid agro-environmental resource providers to annotate, publish and integrate their resources to LOD. The system comprises services and tools that enable resource providers to annotate their resources by relevant concepts from selected agro-environmental domain ontologies, to generate and publish RDF descriptions of the resources to LOD and to link the published resources to related resources from LOD. In addition to the services and tools dedicated to resource providers, AGROPub provides services and tools that enable consumers of the agro-environmental resources to search and annotate published resources by adding their own annotations as well as to evaluate them based on given criteria.

**Keywords:** agro-environmental resource publishing, annotation, linked-open data

## 1 Introduction

The availability of high quality and up to date information on environmental conditions and agriculture has always been of major interest for many institutions, especially given the relevance of such information to our ability to secure food supply. Typical sources of information range from raw data provided by sensor networks and satellite imaging, to processed data stored in databases and catalogues, to information processing units such as mathematical models, software, web services, decision support systems, and to scholar papers, maps and reports, as those published in the JRC MARS bulletin<sup>1</sup>. In spite of being

---

<sup>1</sup> The European Joint Research Center periodically publishes the Bulletin of the Monitoring Agriculture ResourceS Unit (MARS) on agrometeorological analyses for Europe. For more see: <http://mars.jrc.it/mars/Bulletins-Publications>

available on the Web, the great majority of the agro-environmental resources is typically accessed only by closed communities which are aware of them. Accessing agro-environmental resources effectively and efficiently still requires a deep understanding of both domain complexity and computer systems that provide them[14]. Moreover, related resources that are published and managed by different authorities are still greatly disconnected and unaware of each other. As a result, the existing agro-environmental resources on the Web look more as sets of disconnected information islands than an integrated information space.

To narrow the discovery gap that prevents a full and easy access to agro-environmental resources on the Web it is required to develop tools to organize the knowledge hidden in such resources. Recent research efforts [12, 1, 11] have shown that employing rich semantics in the form of domain ontologies is a promising solution. We rely on the TaToo approach<sup>2</sup> that proposes a comprehensive framework for ontology-based, semantic tagging (annotation) of environmental resources [11]. Among others, the TaToo framework provides a novel environmental resource model, called MERM (Minimal Environmental Resource Model) and a number of shared domain ontologies. Moreover, in order to successfully integrate agro-environmental resources on the Web into a globally unified agro-environmental information space, it is necessary to represent (i.e. identify and describe) the resources in accordance to some globally recognized principles. Linked open data (LOD) principles have been recognized as one of the most promising initiatives regarding this issue [4].

In the quest of this long-term vision, we present here AGROPub, a system that facilitates integration of agro-environmental resources into the semantically unified, global environmental information space [9] as a part of LOD. The system is designed as an extensible, service-oriented architecture that intends to provide a variety of services supporting both agro-environmental resource providers and consumers to contribute to this vision. The resource providers are responsible for the resource publishing to LOD as well as setting up some initial links between their resources and other related resources from LOD. However, the real resource integration into LOD happens over time as a result of the user interaction with the resources (e.g., searching and reusing, annotating, interlinking, and evaluating). In this paper our focus is on the AGROPub services and tools that enable agro-environmental resource providers to publish their resources to LOD, add initial semantic annotations, and setup initial links to existing, related resources from LOD.

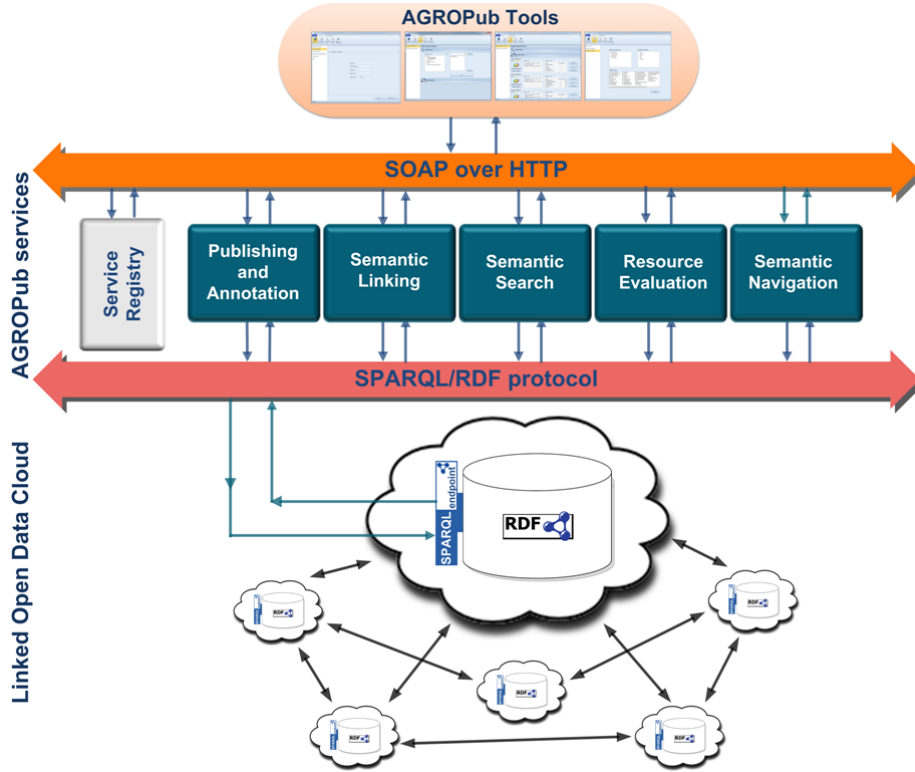
The rest of the paper is organized as follows. In Section 2 we introduce the architecture of the AGROPub system and outline main features of the AGROPub services and tools. Then, in Section 3 and Section 4 we give detailed, step-wise descriptions of the AGROPub resource publishing and linking use-cases respectively. We conclude the paper with a brief overview of related work, and plans for future work.

---

<sup>2</sup> <http://www.tatoo-project.eu>

## 2 AGROPub: Agro-Environmental Resource Publisher

Our main goal of developing the AGROPub system was to support agro-environmental resource providers in publishing and linking (integrating) their resources to the LOD cloud. Apart from that, we also wanted to provide services that would facilitate both resource providers and resource consumers in searching, navigating and evaluating linked agro-environmental resources. We designed the AGROPub system as a three-tier, client-server architecture (see Figure 1). It is composed of a data tier, a service tier, and a presentation tier.



**Fig. 1.** Illustration of the AGROPub architecture

The **data tier** consists of RDF repositories storing 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 the LOD cloud.

The **service tier** offers the AGROPub system functionalities. The number of provided services can be extended by adding an arbitrary number of new services, which have to comply with the AGROPub design principles. Moreover, to be

operational within the system each service has to be registered (through a service registry). The service registry provides registration and look up functionality for the AGROPub services as well as provides methods to enable and disable services. In a current design of the AGROPub system we provided a set of five services:

- (a) *Publishing and Annotation*,
- (b) *Semantic Linking*,
- (c) *Semantic Search*,
- (d) *Resource Evaluation*, and
- (e) *Semantic Navigation*.

The services are designed as Windows Communication Foundation (WCF) Web services, all of them exposing a standardized Web-service interface.

The **presentation tier** is a top tier of the AGROPub architecture, which provides the user interface for the AGROPub services. The presentation tier is technology- and platform-independent according to the service-oriented nature of AGROPub. It may contain different kind of applications, including Web-based applications, desktop applications, and mobile phone applications. In the current version of the AGROPub prototype, as a part of the presentation tier we developed several tools and encapsulated them in a desktop application namely AGROPub Manager.

In the rest of this section, we report on the main features of the AGROPub services and the AGROPub Manager.

## 2.1 AGROPub Services

The **Publishing and Annotation** service provides with the functionality to generate and publish RDF descriptions of agro-environmental resources. The RDF resource descriptions are generated in accordance to the TaToo specifications of the Minimal Environmental Resource Model (MERM) [11]. The MERM model describes environmental resources by combining a set of basic resource information, such as a resource Web page, resource author and resource creation date, and a set of ontology-based annotation tags that represent domain-specific semantics of the resources. The annotation tags, actually, link concepts from a domain ontology to resources (in the form of URIs). As an example domain ontology, used to test our system, we used the JRC ontology [10]. This ontology provides concepts and properties for modeling different types of agro-environmental resources (e.g., agricultural models and agricultural software) as well as concepts and properties describing agricultural systems, processes and products. After generating the RDF descriptions the service stores them into the AGROPub RDF repository and finishes the publishing process. Besides the resource providers, the publishing and annotation service is supposed to be used by the resource consumers to add their own annotations to the linked agro-environmental resources.

The **Semantic Linking** service provides functionalities for associating related agro-environmental resources regardless of whether their RDF descriptions

are stored into the same or different RDF repositories. The service is capable of generating two types of links: generic and property-based links [9]. The generic links are generated automatically whenever a new resource is published. When the resource publishing is finished, the publishing and annotation service triggers the semantic linking service which then crawls available RDF stores, identifies other resources related to the published one, and adds the appropriate generic links. The service determines the semantic relation between resources based on their ontology-based annotations. The property-based links are generated from information (i.e. resources to be linked and a linking property) provided manually by the AGROPub users.

The **Semantic Search** service implements a simple, concept-matching search [12] of linked agro-environmental resources. Similar to the resource annotation process, a user specifies a query by selecting concepts from the domain ontologies used by the AGROPub system. The system takes the user query and invokes the service, which in turn creates a corresponding SPARQL query, and executes it against available RDF repositories. The search result contains resources that are annotated by the query concepts. In the next service version, we plan to enhance the search algorithm by taking into account structure (i.e. ontology-graph traversal) [3] and formal semantics of the underlying domain ontologies (i.e. RDFS/OWL reasoning) [1].

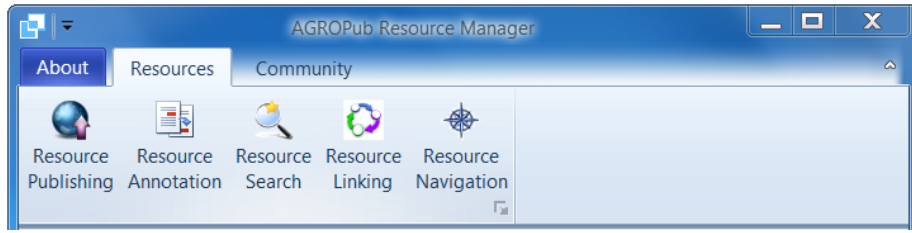
The **Semantic Navigation** service provides functionalities that enable users to navigate across linked agro-environmental resources by following links among them. The navigation process requires the existence of a navigation user interface that is a part of the AGROPub presentation tier, through which the user can interact with the navigation service. The navigation starts by the user selecting a resource from the search results and then clicking on one of the resource's links. This user action invokes the semantic navigation service, which takes as input the resource's URI and the link's data, forms a navigational SPARQL query, and executes the query against the available RDF repositories.

The **Resource Evaluation** service enables AGROPub users to express their opinion on linked agro-environmental resources that are accessible through the AGROPub system. A user is enabled to evaluate not only the actual resources but also the resource annotations added by others. Both resource and annotation evaluations are characterized by an evaluation criterion and a corresponding evaluation value. The service generates and stores the evaluations as RDF instances of the resource evaluation schema, which is defined as a part of the MERM model [11].

## 2.2 AGROPub Manager

The AGROPub Manager is a Windows-based, desktop application that comprises a set of tools enabling users to interact with the AGROPub services. Figure 2 shows a screenshot of the application main menu ribbon tab. The tools are accessible through five menu buttons namely: Resource Publishing, Resource Annotation, Resource Search and Resource Navigation. As their names suggest the first two tools enable users to publish their resources as well as to annotate

already published resources. The next tool, Resource Search, enables users to specify their queries and search for desired resources. The tool also provides a detailed view of resource information and annotations. Among others, the detailed resource view contains user-interface elements that enable the resource evaluation and linking. The last tool, Resource Navigation, enables users to navigate across linked resources by clicking on links connecting them. Users can specify the initial resource by entering the resource's URI or by selecting one of the resources from the list of previously obtained search results.



**Fig. 2.** The main menu tab of the AGROPub Manager

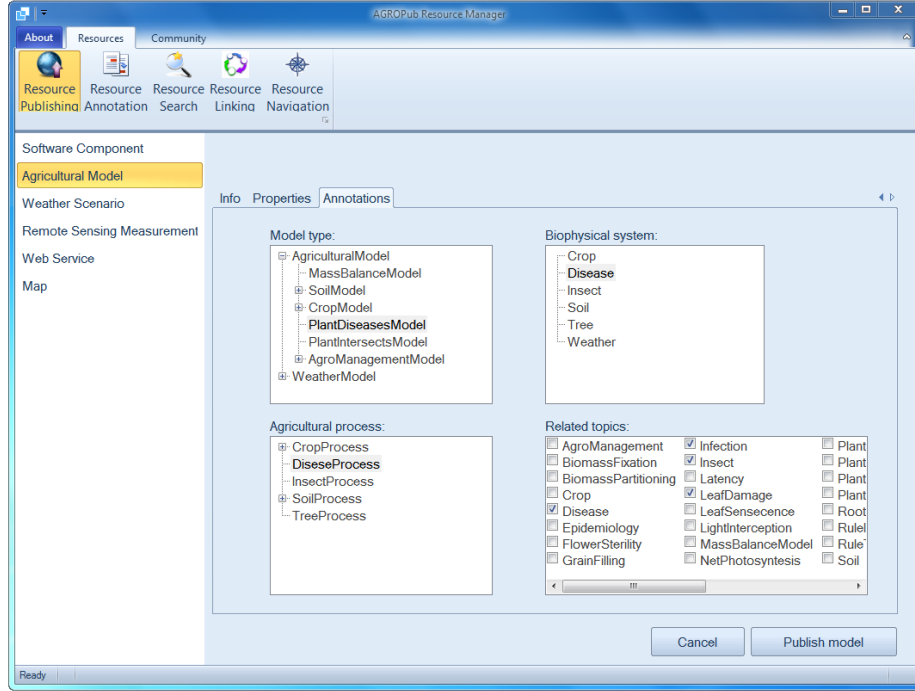
In the following two sections we provide a closer look at the user actions composing the AGROPub resource publishing and linking use-cases. For each user action we refer to the employed tools, user-interface elements, and services of the AGROPub system.

### 3 Publishing Agro-environmental Resources

In this section, we present the resource publishing use-case of the AGROPub system. The resource publishing starts by the user selecting the Resource Publishing button from the main menu ribbon tab of the AGROPub Manager. The user is then faced to the new window with two panes (see Figure 3). The left pane contains a navigation bar with a list of supported agro-environmental resource types (e.g., agricultural models, software components and remote sensing measurements). After selecting a resource type, the right pane displays a resource-publishing control tab, containing forms for specifying a resource description.

In case the user wants to publish an agricultural model, s/he will follow the following steps:

- (a) **Specifying resource info:** The first step in publishing the resource is to specify a set of basic information about the resource including the resource's URI, the URL of the Web page describing the resource, the resource's author, the resource's owner, and the resource's creation date. The only mandatory information to be provided in this step is the resource URI, which is the only mandatory information in a resource description. The resource URI must be globally unique, which we ensure by validating the entered resource URI



**Fig. 3.** Resource annotation and publishing with the AGROPub Manager

against the URI schema that we proposed in [9]. Only in case of resources whose content can be obtained through HTTP, the resource URIs and URLs are identical.

- (b) **Specifying resource properties:** In this step the user specifies values for the properties that are specific to the selected resource type. In case of the agricultural-model, among others, the user specifies the inputs and outputs of the model, the model's algorithm and the software components that implement the model.
- (c) **Specifying resource annotations:** This is the last step in which the user specifies the ontology-based annotations for the resource to be published. Basically, the user selects concepts from a domain ontology [2] that is loaded in the AGROPub manager. The right pane of Figure 3 shows concepts from the JRC ontology that we used in this example to annotate an agricultural model resource. The concepts from the ontology are shown into four boxes each of which is dedicated to a specific aspect of the resource annotation: agricultural model types, related biophysical systems, related agricultural process, and related agricultural topics.

Having passed through these three steps, the user finishes with the specification of the resource description and can initiate resource publishing. By clicking on the “*Publish model*” button, the AGROPub Manager collects the user-specified

model description and calls the publishing and annotation service. When the service finishes the publishing process, the user is notified by a confirmation message. If the user attempts to publish a resource that has already been published, the AGROPub Manager displays an appropriate warning message. Since each resource published to the LOD has to be unique, it is not possible to publish the same resource (i.e. the resource with the same resource URI) more than once.

## 4 Linking Agro-environmental Resources

After having published a resource to the LOD, it is also recommended that the resource publisher provides an initial set of the resource's outgoing links, connecting it to some related resources. Along with the initial resource annotations, specified during resource publishing, the initial set of outgoing links will enhance the resource discoverability and foster further integration of the resource in the LOD. Note that the links we consider in this section are the property-based links which are supposed to be added manually by the AGROPub users. Linking resources by the generic links [9], which are generated automatically, is not in the scope of this paper. To demonstrate the linking use case, consider the same agricultural model used in the resource publishing use-case (Section 3). The AGROPub user is offered with the following options:

1. **Browsing resource details:** The AGROPub Manager offers two ways to access the resource to be linked. One is to manually specify the resource URI and request its details and the other one is to select the resource from the previously obtained search results. The resource detailed view offers a resource-linking user interface, which is realized as a cascading popup with two cascades (see Figure 4). The first cascade contains a combo-box filled with a list of predefined linking properties. The second cascade contains a list of cached resources. Initially, this list is empty.
2. **Caching related resources:** After observing the resource's details the user needs to find resources that s/he might consider for linking afterwards. Therefore, the next step in the linking use-case is to explore available linked resources and to identify potential candidates for linking. To perform this step the user utilizes the search and navigation utilities of the AGROPub Manager. When the user identifies a resource that is related to the considered resource, the AGROPub Manager enables him/her to cache the resource into its internal memory. The user can keep on searching and navigating among available resources as long as s/he caches a satisfactory number of resources.
3. **Selecting a linking property:** Having cached related resources the user moves back to the resource's detailed view and opens the linking popup. The next step is to select the linking property over which the link will be established. Currently, the user is constrained to select the property from the predefined list of properties such as `owl:sameAs`, `skos:related`, `skos:closeMatch` and `skos:exactMatch`. All these properties come with



formal specifications from standard vocabularies. For the next version of the AGROPub System we plan to consider the possibility of using custom, user-defined properties for linking resources.

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 pushes the “Create Link” button the AGROPub Manager calls the linking service, which generates and stores the link (i.e. the RDF triple) in the RDF repository. If the link between the two resources already exists in the repository, it will not be generated again.

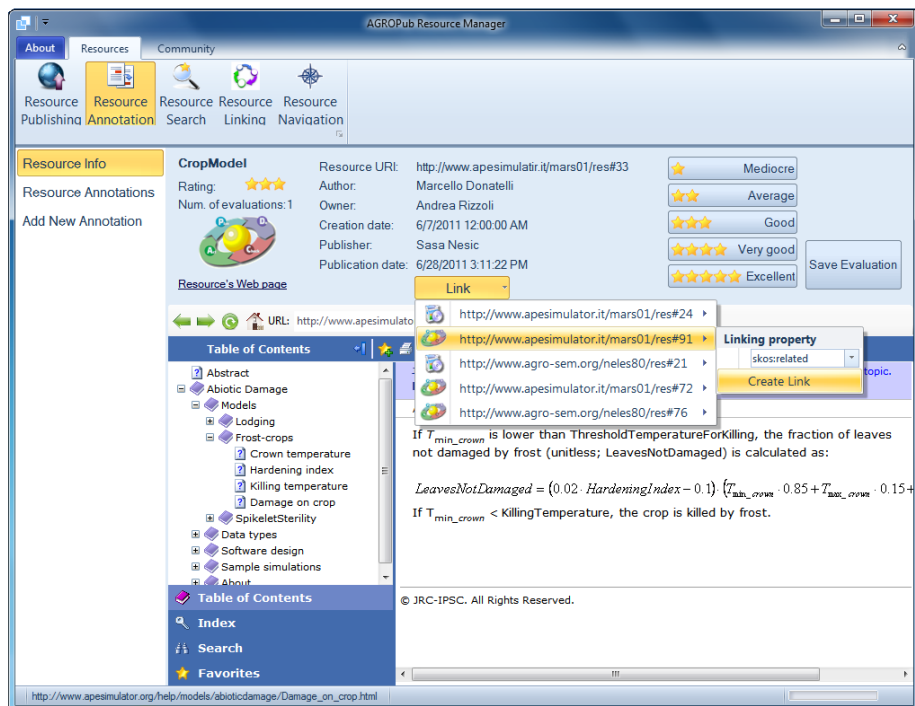


Fig. 4. Resource linking with the AGROPub Manager

## 5 Related Work and Discussion

While there are several generic-purpose tools for annotating web resources (as Annotea<sup>3</sup> and Delicious<sup>4</sup>), there has been a transition towards domain-specific tools that are able to incorporate functionality required by certain communities.

<sup>3</sup> <http://www.w3.org/2001/Annotea>

<sup>4</sup> <http://www.delicious.com>

For example, SemTag & Seeker have demonstrated how to perform automated semantic tagging of large corpora [5]. Connotea<sup>5</sup>, and Mendeley<sup>6</sup> focus on the scientific publications lifecycle. Confolio<sup>7</sup> have been used for annotating and organizing educational resources [6].

Related to the agro-environmental domain there are several services that have been recently deployed, as those for open sensor architectures [13], reusable geospatial services [8], and modelling water resources [7]. However, there are not any services or tools available for annotating and evaluating such 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.

In this paper, we presented the development of the AGROPub system, which is a result of our ongoing efforts to facilitate integration of agro-environmental resources into the semantically-unified environmental information space. The system enables agro-environmental resource providers to annotate their resources by ontology-based, semantic annotations and then publish and link resource descriptions to LOD. Besides the resource providers, the system provides services that enable resource consumers to search, navigate and evaluate linked agro-environmental resources. We designed the system to be extensible in terms of adding new services as well as independent from applied domain ontologies. Therefore, as a part of our future work we plan to investigate on potential new services and to test the system with different domain ontologies. In order to evaluate the system and the underlying annotation, linking, search, evaluation and navigation approaches, we plan to conduct two evaluation studies. The first study will focus on a usability evaluation of the system's services and tools. The main focus of the second study will be on a discoverability of the linked agro-environmental resources within LOD.

## Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement Number 247893.

## References

1. K. Anyanwu, A. Maduko, and A. P. Sheth. Semrank: ranking complex relationship search results on the semantic web. In *Proceedings of the 14th International WWW Conference*, pages 117 – 127, 2005.
2. I. N. Athanasiadis, A. E. Rizzoli, M. Donatelli, and L. Carlini. Enriching environmental software model interfaces through ontology-based tools. *International Journal of Applied Systemic Studies*, 4(1/2):94–105, 2011.

<sup>5</sup> <http://www.connotea.org>

<sup>6</sup> <http://www.mendeley.com>

<sup>7</sup> <http://www.confolio.org>

3. S. Auer, S. Dietzold, and T. Riechert. Ontowiki - a tool for social, semantic collaboration. In *In Proceedings of the 5th International Semantic Web Conference - ISWC*, pages 736 – 749, 2006.
4. C. Bizer, T. Heath, and T. Berners-Lee. Linked Data - The story so far. *Int. Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.
5. S. Dill, N. Eiron, D. Gibson, D. Gruhl, R. Guha, A. Jhingran, T. Kanungo, S. Rajagopalan, A. Tomkins, J. A. Tomlin, and J. Y. Zien. Sematic and seeker: bootstrapping the semantic web via automated semantic annotation. In *Proceedings of the 12th international conference on World Wide Web, WWW '03*, pages 178–186, New York, NY, USA, 2003. ACM.
6. H. Ebner, N. Manouselis, M. Palmer, F. Enoksson, N. Palavitsinis, K. Kastrantas, and A. Naeve. Learning object annotation for agricultural learning repositories. In *Proc. 9th Intl Conf on Advanced Learning Technologies (ICALT 2009)*, pages 438–442, 2009.
7. J. L. Goodall, B. F. Robinson, and A. M. Castronova. Modeling water resource systems using a service-oriented computing paradigm. *Environmental Modelling and Software*, 26:573–582, 2011.
8. C. Granell, L. Diaz, and M. Gould. Service-oriented applications for environmental models: Reusable geospatial services. *Environmental Modelling and Software*, 25(2):182–198, 2010.
9. S. Nešić, A. E. Rizzoli, and I. N. Athanasiadis. Towards a semantically unified environmental information space. In *Proceedings of the the International Symposium on Environmental Software Systems - ISESS 2011*, volume 359, pages 407 – 418. Springer, 2011.
10. S. Nešić, A. E. Rizzoli, I. N. Athanasiadis, and M. Donatelli. Publishing agro-environmental data to a semantically unified environmental information space: a case study. In *In Proceedings of the 25th International Conference on Informatics for Environmental Protection - EnviroInfo 2011*, page In Press, 2011.
11. T. Pariente, J. M. Fuentes, M. A. Sanguino, S. Yurtsever, G. Avellino, A. E. Rizzoli, and S. Nešić. A model for semantic annotation of environmental resources: The tatoo semantic framework. In *Proc. of the International Symposium on Environmental Software Systems (ISESS 2011)*, volume 359, pages 419 – 427. Springer, 2011.
12. C. Rocha, D. Schwabe, and M. P. de Aragão. A hybrid approach for searching in the semantic web. In *In Proceedings of the 13th international WWW conference*, pages 374 – 383, 2004.
13. T. Uslander, P. Jacques, I. Simonis, and K. Watson. Designing environmental software applications based upon an open sensor service architecture. *Environmental Modelling and Software*, 25(9):977–987, 2010.
14. F. Villa, I. N. Athanasiadis, and A. E. Rizzoli. Modelling with knowledge: a review of emerging semantic approaches to environmental modelling. *Environmental Modelling and Software*, 24(5):577–587, 2009.