EcoPortal: a proposition for a semantic repository dedicated to ecology and biodiversity

Nicola Fiore^{1 [0000-0002-9538-2966]}, Barbara Magagna^{2 [0000-0003-2195-3997]} and Doron Gold-farb^{2 [0000-0003-1183-6041]}

¹ LifeWatch Italy, University of Salento, Lecce, Italy ² Umweltbundesamt GmbH, Vienna, Austria nicola.fiore@unisalento.it, barbara.magagna@umweltbundesamt.at, doron.goldfarb@umweltbundesamt.at

Abstract. This paper presents the joint effort of LifeWatch Italy and LTER-Europe to design EcoPortal, a semantic repository focused on ecology and biodiversity as well as on ecosystem observation mainly in the European context. It is our aim to offer a space to collect domain ontologies as well as thesauri and domain relevant reference lists. We plan to test NCBO BioPortal technology to accommodate community requested functionalities.

Keywords: Registry · Ontology · Thesaurus · Reference List · Semantics.

1 Introduction

To address today's ecological challenges, it is necessary to use data coming from different disciplines and providers. Thus, discovery and integration of data, especially from the ecological domain, is highly labour-intensive and often ambiguous in semantic terms. To improve the discovery, integration and re-usability of data the use of semantic resources can help to harmonise and enrich the description of datasets and its content. In the last decade research groups and infrastructures focusing in the monitoring and analysis of ecosystem properties have increasingly put effort into the development of semantic resources mainly based on core ontologies such as OBOE or the O&M data model [1].

This paper presents the joint intention of the European networks LifeWatch Italy¹ and LTER-Europe [2] to design a vocabulary repository focused on the ecology and biodiversity research as well as on observation of biological and physicalenvironmental data. This initiative will support the community in the management and integration/alignment of their semantics and subsequently also of their data [3].

In order to increase interoperability between different domains and institutions, LifeWatch Italy and LTER-Europe² developed ontologies (LifeWatch Ontology³ and

¹ http://www.lifewatchitaly.eu

² http://www.lter-europe.net/

³ http://semantic.lifewatchitaly.eu

SERONTO⁴) as a semantic framework for integration of monitoring [4] and biodiversity data and common vocabularies for harmonised data annotation (LifeWatch-Italy Thesauri⁵, concerning functional traits, and EnvThes⁶ - Environmental Thesaurus [5]).

LifeWatch Italy and LTER-Europe are collaborating in order to improve and extend the existing thesauri and trace the semantic relations between them. In this context, the lack of a common semantic repository for the ecological domain became evident. We envisage to build a semantic platform for the domain to support not only the joint work done by the infrastructures, but to be a robust and stable reference repository for the European ecological community.

2 State of the art

Scientific communities are using an increasing number of ontologies or controlled vocabularies to disambiguate the description of data. To make these vocabularies discoverable and usable by software or by the community, different approaches exist:

- Distributed RDF stores with SPARQL endpoints allowing to access vocabularies using SPARQL queries. The presence of such endpoints does not solve the issue of discoverability, as you must already be aware of the semantic resources by other means to make use of them. To a certain extent they facilitate interoperability between different semantic resources but this requires high familiarity with the data representation schema and the granularity of each federated source.
- Semantic repositories (known also as ontology libraries [6]) are centralised access points providing both discoverability and access to semantic resources.

The latter, on which this paper focuses on, are collections of ontologies and thesauri with the primary purpose of enabling users to find and use them. They should be distinguished from ontology search engines, such as Swoogle⁷, which automatically crawl the Web to index ontologies rather than collect them. Also, we want to exclude here collections on data, such as the Linked Open Data collection of datasets⁸.

According to the targeted user set approach of d'Aquin and Noy [6] different types of repositories can be identified, although they often exist in a mixed form: curated directories, registries and application platforms.

Many repositories already offer additional services, the most prominent ones are BioPortal⁹ and EBI OLS¹⁰.

⁴http://www.umweltbundesamt.at/fileadmin/site/daten/Ontologien/SERONTO/SERO NTOCore20090205.owl

⁵ thesauri.lifewatchitaly.eu/

⁶ http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn

⁷ http://swoogle.umbc.edu/2006/

⁸ https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

⁹ http://bioportal.bioontology.org/

¹⁰ https://www.ebi.ac.uk/ols/index

We would also like to emphasise that most of the repositories are dedicated solely to ontologies, some only contain thesauri like Finto¹¹ and LusTRE¹² and only a few seem to offer the place to publish both ontologies and thesauri like AgroPortal¹³. The inclusion of thesauri is important in our considerations because they are essential sources of harmonised knowledge (not only) in the ecology domain.

3 EcoPortal

3.1 Requirement Elicitation: Purpose and Coverage

The main goal of the EcoPortal initiative is to provide a central registry for semantic resources (e.g. vocabularies) used in the ecological and biodiversity domain allowing users to identify and select semantic resources for specific tasks, as well as offering generic services to exploit them in search, annotation or other scientific data management processes.

To reach this objective the user-centred, structured and systematic approach AWARE (Analysis of WebApplication Requirements) has been adopted [7].

Following the AWARE guidelines, the following main stakeholders (i.e. user profiles to be considered for the Web application) have been identified:

- Domain Expert, is the user of the portal and expert of the ecological domain. One high-level goal of this kind of user is to explore the semantic world in the ecological domain to understand how to annotate experimental data to enable interpretation, comparison, and discovery across databases. For this kind of user it is necessary to offer very user-friendly tools and services.
- *Semantic Author*, is a domain expert user that creates and shares a specific vocabulary/ontology and is responsible to maintain it updated.
- Semantic Engineer, is a type of user with semantic technology skills, who aims to design new tools/services for the domain expert.
- System Owner, who creates and manages EcoPortal and its services.

Figure 1 shows parts of the requirements analysis made for the stakeholder *Domain Expert*. For each stakeholder we have identified goals and tasks (i.e. high-level user activities on the site) and in the refined process they have been recompiled into requirements. We can classify and synthesise the main requirements of the EcoPortal in the following categories.

• Content Requirements

The focus of the portal will be on the ecology, ecosystem and biodiversity domains. Not only ontologies but also thesauri will be collected and managed. Each semantic resource will be described by metadata (i.e. Structure Content

 $^{^{11}\} https://www.kansalliskirjasto.fi/en/services/system-platform-services/finto$

¹² http://linkeddata.ge.imati.cnr.it:2020/

¹³ http://agroportal.lirmm.fr/

Requirements in AWARE). The need of a common metadata set has been identified by several initiatives like OBO Foundry [8], LOV¹⁴ and AgroPortal.

- Access Path and Navigation Requirements
 - Different search paths should be supported fitting the general requirements: search within and across ontologies/thesauri, structured search via a SPARQL query engine and advanced search will be developed. In the scenario in Figure 1 the *Domain Expert* needs to perform a search for "equivalent terms" and to navigate from a term to the related one.
 - To facilitate the semantic resource discoverability, we want to use categories (ecology, observation, etc.) as used in AgroPortal.
 - Browsing functionality will be offered including different types of visualisation of the content. So, we aim to foresee automatic translations for single terms wherever the vocabularies provide multilingual labels for them. Access services will be also provided for all the resources, including ontologies/thesauri metadata and the mappings between them.
 - We also intend to collect reference lists codified in SKOS used to define permissible values in certain data fields, providing information needed to make other data meaningful and interpretable in an unambiguous way. Translating from one reference list to another within the same domain is an essential need for ecologists.



Fig. 1. Requirements for Domain Expert

- System and User Operation Requirements
 - The portal should enable automatic (based on exact matching labels) as well as manual mappings between semantic resources (private and/or public accessible -

¹⁴ http://lov.okfn.org/dataset/lov/

storing also metadata on mappings), and it should allow upload of mappings created elsewhere.

- For collecting resources, EcoPortal should use a hybrid approach: apart from the administrators (ensuring to host the newest version of the resource also published in other portals) also users should be enabled to submit their resources to the collection through a dedicated user interface.
- As far as belongs to gatekeeping, we envisage a two-step approach: after uploading, the semantic resource is validated by a quality committee, after that it is published in the catalogue. But before validation the resource should already be visible to the users labelled as not yet validated. Quality requirements should include metadata description, syntactically correctness and thematic relevance.
- The portal should be able to automatically compute ontology metrics.
- It should enable social interaction, allowing comments on ontologies and components (at class level).
- Instead of ranking ontologies by their relevance, we would prefer an exchange information platform between supplier and user where it should become clear for which use cases the resources were originally developed and then used. This concept of semantic marketplace has been introduced at the EUDAT Semantic Workshop¹⁵. We want to encourage developers to publish their vocabularies in our Portal in an early stage of their development taking advantage of the domain community.

3.2 Expected Contents.

A first inventory of the appropriate and relevant ontologies, thesauri and reference lists to be hosted in the repository can be accessed online¹⁶. This list will be extended by community contributions as collaborative and open process.

3.3 Conclusions and Future Work

The paper briefly introduces the ongoing work of LifeWatch Italy and LTER-Europe in order to develop EcoPortal, a semantic repository focused on the ecosystem and biodiversity research as well as on observation of the ecosystem. A common domain specific repository of semantic resources allows their better integration into the work-flows of metadata annotation (e.g. DEIMS-SDR¹⁷) and discovery. This fosters the semantic interoperability not only on the metadata but also on the data level.

¹⁵ https://www.eudat.eu/events/trainings/co-located-eudat-semantic-working-group-workshop-9th-rda-plenary-barcelona-3-4

¹⁶ http://www.servicecentrelifewatch.eu/web/ecoportal/wiki/-/wiki/Main/EcoPortal+semantic+resources?_36_redirect=http%3A%2F%2Fwww.servicece ntrelifewatch.eu%2Fweb%2Fecoportal%2Fwiki%2F-%2Fwiki%2FMain%2Fall_pages%3Fp_r_p_185834411_title%3DEcoPortal%2Bsemantic% 2Bresources

¹⁷ https://data.lter-europe.net/deims/

A first prototype in line with the described architecture is planned to be online by October 2017. In the initial phase, we will test the NCBO BioPortal technology to accommodate community-requested functionalities with semantic resources of European networks. Considering the importance of such tools in the ecological field, we expect a broad adoption of the EcoPortal in the community in the long run. Furthermore, LifeWatch as ERIC will be able to assure the long-term product sustainability.

The most pressing issues still to be addressed are the ability to manage and search across different types of semantic resources like OWL ontologies and SKOS thesauri as well as the use of a minimal metadata set and of a vocabulary marketplace considering the ongoing discussions in the RDA VSIG¹⁸.

4 Acknowledgments

The work is funded using resources from the ENVRIPlus (H2020, Nr. 654182), ECOPOTENTIAL (H2020, Nr. 641762) and eLTER (H2020, Nr. 654359) projects.

References

- 1. Cox, S.J.: Ontology for observations and sampling features, with alignment to existing models. Semantic Web, 8(3), 453-470 (2017).
- Mirtl M.: Introducing the Next Generation of Ecosystem Research in Europe: LTER-Europe's Multi-Functional and Multi-Scale Approach. In: Müller F., Baessler C., Schubert H., Klotz S. (eds) Long-Term Ecological Research. Springer, Dordrecht (2010). doi: 10.1007/978-90-481-8782-9_6
- Oggioni, A., Carrara, P., Kliment, T., Peterseil, J. & Schentz, H.: Monitoring of Environmental Status through Long Term Series: Data Management System in the EnvEurope Project. In: Proceedings EnviroInfo 2012, pp. 293-301. Sahker Verlag, Aachen (2012).
- Van der Werf, B., Adamescu, M., Ayromlou, M., Bertrand, N., Borovec, J., Boussard, H., et al.: SERONTO: A Socio-Ecological Research and observation oNTOlogy. In: Weitzman, A. L. & Belbin L. (eds.) Proceedings of TDWG. Fremantle, Australia (2008).
- Schentz, H., Peterseil, J., Bertrand, N.: EnvThes interlinked thesaurus for long term ecological research, monitoring, and experiments. In: Proceedings EnviroInfo 2013: Environmental Informatics and Renewable Energies, pp. 824-832. Shaker Verlag, Aachen (2013).
- d'Aquin, M., Noy, N.: Where to Publish and Find Ontologies? A Survey of Ontology Libraries. In Web Semant 11: pp. 96–111 (2012). doi: 10.1016/j.websem.2011.08.005.
- Bolchini, D., Paolini, P.: Goal-driven requirements analysis for hypermedia-intensive Web applications. In: Requirements Eng 9: pp. 85–103 (2004). doi: 10.1007/s00766-004-0188-2
- Smith, B., Ashburner, M., Rosse, et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. In: Nature Biotechnology, 25 (11), 1251 (2007).

¹⁸ https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html