

EcoPortal: An Environment for FAIR Semantic Resources in the Ecological Domain

Xeni Kechagioglou¹, Lucia Vaira¹, Pierfrancesco Tommasino², Nicola Fiore¹, Alberto Basset^{1,2,3} and Ilaria Rosati²

¹ LifeWatch ERIC, Service Centre, S.P. Lecce-Monteroni – Ecotekne, Lecce, 73100, Italy

² Institute of Research on Terrestrial Ecosystems (IRET), National Research Council (CNR), Via Salaria Km 29.300, 00015 Monterotondo Stazione, Rome, Italy

³ University of Salento, Department of Biological and Environmental Sciences and Technologies, S.P. Lecce-Monteroni – Ecotekne, Lecce, 73100, Italy

Abstract

Ecological semantic resources are being increasingly required by researchers. To improve their FAIRness, LifeWatch ERIC launched a repository to collect, describe, display, manage and facilitate semantic resource discovery, called EcoPortal. Based on OntoPortal, EcoPortal has achieved the key functionality targets commonly expected by knowledge organisation system repositories. In addition, it allows for resource updating through an integrated editor, VocBench, while published resources may also obtain a Digital Object Identification for citing.

Keywords

Ecology, biodiversity, FAIR, semantic resource, knowledge organisation system, EcoPortal

1. Introduction

Perceiving the necessity for collaboration and integration to overcome fragmented and siloed ecological data [1], research has been investing in creation and exploitation of semantic resources. Formalised knowledge organisation systems (KOS) of a range of intricacy may now be encountered, from simple lists of local species to domain thesauri, all aiming at systematising knowledge and aligning information. With advancements in data collection and the accumulation of vast amounts of data to be consumed, this need for homogeneity has nowadays become even more pressing [2].

To serve the purpose of knowledge integration, targets for data producers and publishers have been summarised in the four main principles of FAIRness: Findability, Accessibility, Interoperability, and Reusability [3]. For semantic resources, these principles have been translated into ten rules for accurate communication across the users and producers of ecological information [4]. Among others, the rules make it explicit that FAIR KOS be available from at least one repository recognised by the target community, with adequate metadata, resource mappings, feedback mechanisms and proper citation.

This paper presents EcoPortal, a repository implemented to host KOS, promote discoverability, render content and metadata accessible, increase interoperability, and facilitate reuse.

2. The EcoPortal environment

S4BioDiv 2021: 3rd International Workshop on Semantics for Biodiversity, held at JOWO 2021: Episode VII The Bolzano Summer of Knowledge, September 11-18, 2021, Bolzano, Italy

EMAIL: xeni.kechagioglou@lifewatch.eu (X. Kechagioglou); pierfrancesco.tommasino@iret.cnr.it (P. Tommasino); lucia.vaira@lifewatch.eu (L. Vaira); nicola.fiore@lifewatch.eu (N. Fiore); alberto.basset@lifewatch.eu (A. Basset); ilaria.rosati@cnr.it (I. Rosati)

ORCID: 0000-0002-5935-6074 (L. Vaira); 0000-0002-9538-2966 (N. Fiore); 0000-0002-3603-9316 (A. Basset); 0000-0003-3422-7230 (I. Rosati)



© 2021 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

The initiative to implement a repository for ecological semantic resources was taken by LifeWatch Italy and the Long-Term Ecosystem Research Network in Europe (LTER-Europe) [5]. The physical realization of EcoPortal (ecoportal.lifewatch.eu) was launched by the European consortium for e-science infrastructure for biodiversity and ecosystem research (LifeWatch ERIC). The domain niche specified for the repository is ecology, particularly biodiversity and ecosystem research, as well as biological and physical environmental data observations. It pertains all fields relevant to the study of ecosystems, including both their biotic and abiotic components and their interactions.

Biodiversity is also a theme of interest to AgroPortal [6], although there it is intended as auxiliary to the main domains: agronomy and plants. Similarly, the GFBio Terminology Service [7] lists largely relevant KOS, although its thematic scope goes beyond that of EcoPortal, to include e.g., geosciences and chemistry. The EcoPortal niche laps slightly over also with the repository of biomedical ontologies BioPortal [8], whose collection includes a few KOS on phenotype, biotic and abiotic stress, and organism diversity. This overlap of fields of interest is not a drawback. On the contrary, it may serve as incentive for resource harmonization, development of cross-portal services and even detection of novel affinities. Nor may it be foreseen to shrink; if anything, one may expect niche expansion for EcoPortal to encompass new directions, as demonstrate indicative cases such as species conservation [9].

To meet the functional scope commonly expected by KOS repositories, and thus differentiate from KOS registries that merely store and retrieve metadata about semantic assets [10], EcoPortal adheres to a set of key functionalities specified in [11]. The main components of the EcoPortal environment are the repository deployment infrastructure OntoPortal and the VocBench platform for semantic resource development, both coming with compatibility with a Resource Description Framework (RDF) store.

2.1. Basic Functionalities

The basic functionalities of EcoPortal are based on OntoPortal (deployed version 2.5), a VMware Virtual Appliance infrastructure [12] currently maintained by the Stanford Center for Biomedical Informatics Research (BMIR). It is advanced by and for the collaborative OntoPortal Alliance [13], of which LifeWatch ERIC is an active member. Its basic tools cater for browsing, searching, publishing, annotating, mapping, and receiving recommendations for semantic assets.

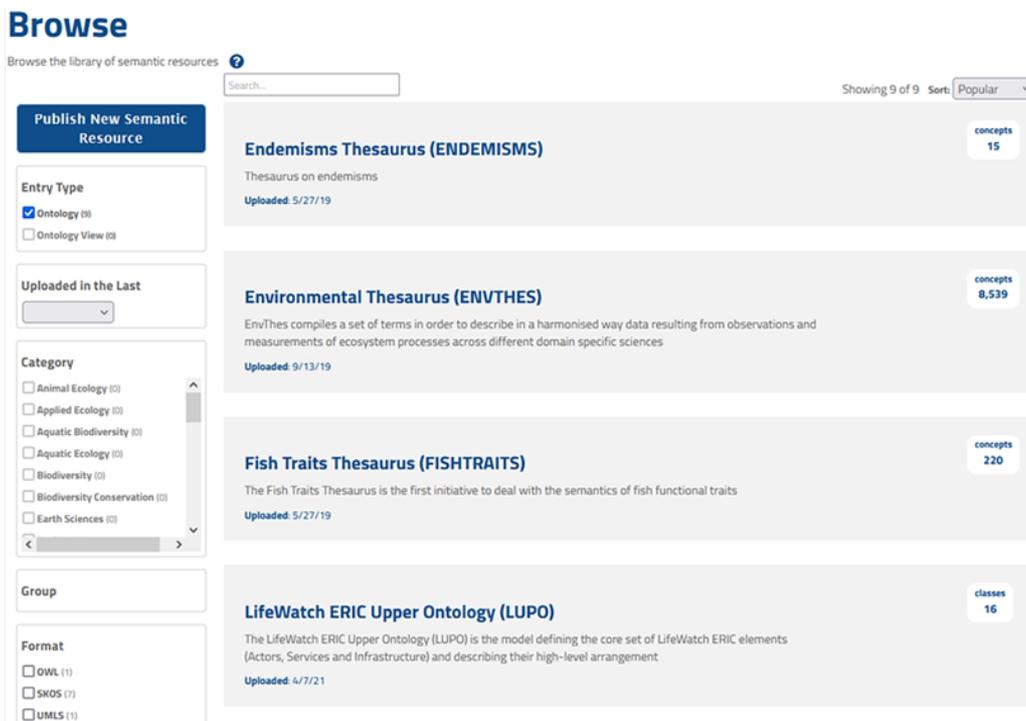


Figure 1: The browsing interface of EcoPortal. Central to the view is the list of published KOS with an indication of their size. Users may restrict their search by ecological categories (among others), and they may also query the list by text.

Browse

Browsing (Figure 1) is assisted by matching user-inserted text to the title or the description of the KOS. Filtering is accomplished by means of EcoPortal-specific categories (Table 1), currently mostly coinciding with GEMET (GENERAL MULTILINGUAL ENVIRONMENTAL THESAURUS) concepts [14] and intended to be updated according to user needs. Upon location of a KOS of interest, users may view its metadata, as well as visualise its hierarchical and overall structure and content.

Table 1
EcoPortal categories for filtering the resources during browsing

Category Name	Source	Category Name	Source	Category Name	Source
Ecology	GEMET	Biodiversity	GEMET	Earth sciences	EcoPortal
Applied ecology	GEMET	Biodiversity conservation	GEMET	Environmental sciences	EcoPortal
Forest ecology	GEMET	Terrestrial biodiversity	GEMET	Oceanography	EcoPortal
Landscape ecology	GEMET	Aquatic biodiversity	GEMET		
Plant ecology	GEMET	Urban biodiversity	GEMET		
Population ecology	GEMET	Soil biodiversity	GEMET		
Synecology	GEMET	Forest biodiversity	GEMET		
Urban ecology	GEMET	Species diversity	EcoPortal		
Human ecology	GEMET	Ecosystem diversity	EcoPortal		
Animal ecology	GEMET	Functional diversity	EcoPortal		
Aquatic ecology	GEMET	Genetic diversity	GEMET		
Palaeoecology	GEMET				

Search, Mappings, Annotator, Recommender

Searching for specific concepts in the entire list of KOS also uses the above ecological categories to filter results. Mapping of similar concepts across EcoPortal resources is based on text-matching concept labels, while text analysis for annotation operates by string matching and by leveraging the KOS hierarchical structure and mappings [15]. Equally important for resource reuse, the Recommender service evaluates relevance of each EcoPortal KOS with regard to keywords or free text inserted by the user. The tool builds on the annotation service, as well as on statistical and quality metrics of the resources and of the annotation output, before assigning criteria weights and computing scores [16].

Publish

Publishing a new KOS on EcoPortal is at the heart of the repository's existence. The resources have metadata associated with them, submitted to the system through an online form upon uploading. The required elements follow a schema (Table 2) that, having started with the BioPortal Metadata ontology [17], was adjusted with elements from the extended AgroPortal schema [18] and was completed with fields from the DataCite Metadata Schema (v.4.3) [19]. Albeit all elements were selected to serve FAIRness, the DataCite ones in particular are meant for Digital Object Identifier (DOI) assignment to the uploaded KOS.

Uploading of a KOS is not a prerequisite to its indexing in EcoPortal. In fact, the cataloguing aspect of the repository is independent of whether the resource resides in it or not, as long as sufficient information for successful redirection is provided. Uploaded resources may further be marked as public or private, imposing restrictions on their accessibility, while persons indicated as contacts may optionally receive feedback in the form of comments and proposals from users of their resource.

Table 2

The main metadata elements for KOS description in EcoPortal

Metadata element	Description	Source
Name	Name of the KOS	BioPortal
Acronym	Acronym of the KOS	BioPortal
Description	Description of the KOS	BioPortal
Status	Release status, e.g., alpha	BioPortal
Version	KOS version	AgroPortal
Format	KOS format	AgroPortal
Visibility	Viewing restrictions (public/private)	AgroPortal
Contact	Name and e-mail of contact person	AgroPortal
Homepage	Homepage URL	AgroPortal
Publications	URL for bibliographic reference	AgroPortal
Documentation	URL for further documentation	AgroPortal
Categories	Controlled list with categories	AgroPortal
Groups	Controlled list with groups	AgroPortal
Release date	Date of upload to the EcoPortal	AgroPortal
Identifier	Unique string ID (e.g., DOI)	DataCite
Identifier type	Type of identifier	DataCite
Creators	Actors involved in KOS creation	DataCite
Titles	Alternative names of the KOS	DataCite
Publisher	Publishing entity	DataCite
Resource type	KOS type, e.g., ontology	DataCite
Resource type general	Dataset or other	DataCite

2.2. EcoPortal extended functionalities

VocBench

Recognising the need for editing and updating EcoPortal resources with as few steps as possible, the VocBench 3 (VB3) development platform has been integrated. VB3 - deployed version 8.02 - is an open-source editing environment for ontologies, thesauri and RDF datasets. Initially created to support the needs of the AGROVOC thesaurus, it offers itself as a web-based platform [20].

Connection between VocBench and EcoPortal user accounts is established via a personal API key assigned to the user upon registration to EcoPortal. Users may update the resource in VocBench when and as deemed suitable; once the new version is ready to be submitted in EcoPortal, users may do so from VocBench, through a form (Figure 2) that contains the same metadata fields as the EcoPortal Publish tool does; in this way, KOS deployment through VocBench automatically updates EcoPortal.

Beyond employing a straightforward process for new KOS versions, advantages to using VocBench include better development control. Users assigned to a collaborative KOS creation project may be delegated roles and get authorisation to perform specific tasks only, while all edits undergo a validation process. Last, VocBench comes with a SPARQL endpoint for querying KOS structures and linked data.

Digital Object Identifier

LifeWatch ERIC is a direct member of DataCite and has the authorisation to provide a DOI to ecological semantic resources that fulfil certain quality criteria, in order to be cited in a reliable and constant way. KOS owners may request this upon publication of their resource on the EcoPortal repository. DOI assignment is free of charge and is deemed fundamental to the FAIRness of resources.

The screenshot shows the EcoPortal metadata form for KOS description. At the top, there is a warning message: "The API key is stored without encryption on the server. Be aware that the system administration could be able to see it." Below this, the form contains several fields with labels and asterisks indicating mandatory fields: API Base URL, API Key *, Acronym *, Description *, Version, Format *, Status *, Release date *, Contacts *, Homepage, Documentation, Publications *, Creators *, and Titles *. The Titles * field is expanded to show a sub-field for Title * (with a small icon) and a dropdown for Title type * (set to "Other"). There are also several "+" and "-" icons next to the fields, and an "Ok" button and a "Cancel" button at the bottom right. A legend at the bottom left indicates "(*) Mandatory field".

Figure 2: The EcoPortal metadata schema for KOS description is respected whether resource submission occurs through EcoPortal or through VocBench.

3. Conclusion and outlook

The EcoPortal environment for FAIR semantic resources in the ecological domain currently fulfils functionalities commonly required by KOS repositories, while it also offers unique capabilities that increase reuse of the hosted assets. Its realisation is incorporated in the broader mission of LifeWatch ERIC to support with e-science infrastructure the scientific community in biodiversity and ecosystem research. Next steps in its establishment involve cases of utilisation by the end users to guide implementation of new repository features, as well as component upgrade and consolidation in a new LifeWatch ERIC platform.

End users play in fact a key role in EcoPortal advancement, as identification of needs and design for improvement may only be achieved through stepping-up the engagement of the ecological community and the exchange of feedback. In this way, the EcoPortal environment may serve its purpose as the entry point for KOS discovery, with the aim to use them in annotation of Digital Objects, in enrichment of data, in reuse of information and services, and in FAIR maintenance of resources and components.

4. References

- [1] J. N. Thompson, O. J. Reichman, P. J. Morin, G. A. Polis, M. E. Power, R. W. Sterner, C. A. Couch, L. Gough, R. Holt, D. U. Hooper, F. Keesing, C. R. Lovell, B. T. Milne, M. C. Molles, D. W. Roberts, S. Y. Strauss, *Frontiers of Ecology, Bioscience* (1999) 15(1) 15-24. doi:10.1641/0006-3568(2001)051[0015:FOE]2.0.CO;2.
- [2] C. Barba-González, J. García-Nieto, M.M. Roldán-García, I. Navas-Delgado, A.J. Nebro, J.F. Aldana-Montes, *BIGOWL: Knowledge centered Big Data analytics, Expert Systems With Applications* 115 (2019) 543-556. doi:10.1016/j.eswa.2018.08.026.
- [3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al., *The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data* (2016) 3(1):160018. doi:doi.org/10.1038/sdata.2016.18.
- [4] S. J. D. Cox, A. N. Gonzalez-Beltran, B. Magagna, M.-C. Marinescu, *Ten simple rules for making a vocabulary FAIR. PLoS Computational Biology* (2021) 17(6):e1009041. doi: 10.1371/journal.pcbi.1009041.
- [5] N. Fiore, B. Magagna, D. Goldfarb, *EcoPortal: A proposition for a semantic repository dedicated to ecology and biodiversity*, in: *Proceedings of the 2nd International Workshop on Semantics for Biodiversity, S4BioDiv '17, CEUR Workshop Proceedings Vol-1933 (CEUR-WS.org), Vienna, Austria, 2017.*
- [6] C. Jonquet, A. Toulet, E. Arnaud, S. Aubind, E. Dzalé Yeumod, V. Emoneta, J. Graybealf, et al., *AgroPortal: A vocabulary and ontology repository for agronomy, Computers and Electronics in Agriculture* 144 (2018) 126–143. doi:10.1016/j.compag.2017.10.012.
- [7] N. Karam, C. Müller-Birn, M. Gleisberg, D. Fichtmüller, R. Tolksdorf, A. Güntsch, *A Terminology Service supporting semantic annotation, integration, discovery and analysis of interdisciplinary research data, Datenbank Spektrum* 16 (2016) 195–205. doi: 10.1007/s13222-016-0231-8.
- [8] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, M. A. Musen, *BioPortal: Enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications, Nucleic acids research*, 39 (Web Server issue) (2011) W541–W545. doi:10.1093/nar/gkr469.
- [9] R. E. Hintzen, M. Papadopoulou, R. Mounce, C. Banks-Leite, R. D. Holt, M. Mills, et al., *Relationship between conservation biology and ecology shown through machine reading of 32,000 articles, Conservation Biology*, 34(3) (2019) 721–732. doi:10.1111/cobi.13435.
- [10] J. Voß, J. M. Agne, U. Balakrishnan, M. Akter, *Terminology registries and services, Poster, The Research Data Alliance Deutschland Treffen, RDA-DE-16, Potsdam, 2016.* doi:10.5281/zenodo.166717.
- [11] M. d'Aquin, N. F. Noy, *Where to publish and find ontologies? A survey of ontology libraries, Web Semantics: Science, Services and Agents on the World Wide Web* 11 (2012) 96–111. doi:10.1016/j.websem.2011.08.005.

- [12] J. Graybeal, A. Skrenchuk, J. Vendetti, M. Dorf, C. Jonquet, N. Fiore, X. Yang, M. A. Musen, OntoPortal Virtual Appliance and Community Alliance, Poster n.74, Research Data Alliance Plenary 14, 2020. URL: <https://ontoportall.org/ontoportall-alliance-at-rda-plenary-14/>.
- [13] J. Graybeal, C. Jonquet, N. Fiore, M. A. Musen, OntoPortal software: Adoption increasing, Poster, Research Data Alliance Plenary 13, 2019. URL: <https://ontoportall.org/ontoportall-alliance-at-rda-plenary-13/>.
- [14] European Environment Information and Observation Network – EIONET, General Multilingual Environmental Thesaurus (GEMET) v.4.2, 2021. URL: <https://www.eionet.europa.eu/gemet/en/about/>.
- [15] N. H. Shah, N. Bhatia, C. Jonquet, D. Rubin, A. P. Chiang, M. A. Musen, Comparison of concept recognizers for building the Open Biomedical Annotator, *BMC Bioinformatics* (2009) 10(Suppl 9):S14. doi: 10.1186/1471-2105-10-S9-S14.
- [16] M. Martínez-Romero, C. Jonquet, M. J. O’Connor, J. Graybeal, A. Pazos, M. A. Musen, NCBO Ontology Recommender 2.0: An enhanced approach for biomedical ontology recommendation, *Journal of Biomedical Semantics* (2017) 8(21) 1-22. doi: 10.1186/s13326-017-0128-y.
- [17] National Center for Biomedical Ontology (NCBO) public wiki, Architecture, 2010. URL: <https://www.bioontology.org/wiki/Architecture>.
- [18] C. Jonquet, A. Toulet, B. Dutta, V. Emonet, Harnessing the power of unified metadata in an ontology repository: The case of AgroPortal. *Journal on Data Semantics* 7 (2018) 191–221. doi: 10.1007/s13740-018-0091-5.
- [19] DataCite Metadata Working Group, DataCite Metadata Schema documentation for the publication and citation of research data, Version 4.3, DataCite e.V., 2019. doi:10.14454/7xq3-zf69.
- [20] A. Stellato, M. Fiorelli, A. Turbati, T. Lorenzetti, W. Gemert, D. Dechandon, et al., VocBench 3: A collaborative Semantic Web editor for ontologies, thesauri and lexicons, *Semantic Web* 5 (2020) 1-27. doi:10.3233/SW-200370.