The ODS Vocabulary Bank[*]

Anastasios Koutoumanos and Elina Megalou

Computer Technology Institute & Press "Diophantus", Greece

Abstract—**The ODS Vocabulary Bank is a component of the Open Discovery Space socially-powered federated architecture for repositories of learning resources. It implements a single point of reference for all authoritative sets of terms, concepts and named entities as well as the network of relations between them, using an innovative, semantic web approach. This paper provides an overview of the ODS Vocabulary Bank, the software requirements, the research of the state of the art, the selected implementation approach and some considerations regarding future work towards the evolution of the system.**

Index Terms— **ODS, semantic web, taxonomy, vocabulary**

I. INTRODUCTION

The ODS Vocabulary Bank is a web-based application that implements a single point of reference for authoritative sets of terms, concepts and named entities used for the metadata descriptions of learning resources, as well as the network of semantic relations between them. It is an essential component of the system's architecture of the Open Discovery Space (ODS) initiative, which implements a community-oriented social platform enabling teachers, pupils, and parents to discover, acquire, exchange, and adapt learning resources from a network of interconnected repositories across Europe. The existence of a single point of reference has been identified as crucial from the early phases of the design of the ODS architecture, to accomplish a unified experience for all users interacting with the ever-increasing wealth of resources, allowing the use of a common language for describing, locating, and browsing for these resources.

The paper is structured in six sections: following this introduction, section 2 provides a short overview of the ODS initiative; section 3 describes the essential requirements of the ODS Vocabulary Bank, based on which an analysis of the state of the art for similar systems is presented in section 4; the implementation of the system is documented in section 5; and section 6 concludes the paper with a discussion of the main outcomes from system's usage and planned future work.

II. THE ODS FEDERATED ARCHITECTURE

The ODS infrastructure and its overall system architecture

is described in the following paragraphs, to make apparent the importance and need for a vocabulary bank as a central reference point, enabling the management and usage of common terminology across the different components and facilitating multilingualism for all related services.

The Open Discovery Space is a prominent European initiative that aims in (a) implementing a socially-powered open federated infrastructure for a super-repository on top of existing repositories of learning resources and related federated infrastructures; and (b) providing social features for building and sustaining web-based educational communities and communities of best teaching practices from 2,000 European schools [1]. One of the main objectives of the ODS infrastructure is, thus, to make available a wealth of learning resources from a large network of repositories across Europe, through a community-oriented social platform (web portal) where teachers, pupils, and parents discover, acquire, discuss, and adapt learning resources on their topics of interest. An essential challenge for ODS is, therefore, to achieve interoperability and information sharing between semi-autonomous de-centrally organized repositories. In order to facilitate discovery of resources, all their descriptions need to be aligned to a uniform metadata schema, which includes a set of predefined terms, concepts, and thematic taxonomies for each specific educational discipline. Users can therefore search against selected terms in specific fields or browse for resources in their preferred disciplines.

In order to address this challenge, the metadata of all interconnected repositories are collected, processed, and stored by the central component of the ODS infrastructure, the ODS harvester. This component also executes all necessary transformations to map information from each metadata schema to the native metadata schema. Thus, all information becomes readily accessible through the ODS web portal, which provides the front-end user interface and services. The portal serves an international, multilingual audience of teachers, learners, and parents and facilitates searching, browsing, and accessing the wealth of learning resources across all connected repositories, removing linguistic and cultural barriers. Furthermore, a set of peripheral components allows the interconnection of the ODS infrastructure to external systems, using established protocols and standards.

The ODS Vocabulary Bank provides a specific, yet essential "glue" service to the infrastructure: it implements a central point of reference for all associated vocabulary terms and taxonomies, their descriptions, and their translations. The next section provides an enumeration of the key software requirements and the associated expected benefits from the incorporation of the ODS Vocabulary Bank in the overall infrastructure.

III. FUNCTIONAL REQUIREMENTS

This section presents the key functional requirements for the ODS Vocabulary Bank, to enable the collaborative editing and management of terms using a linked open data approach, while taking into consideration the multilingualism requisites and, at the same time, enabling its integration with other parts of the ODS infrastructure as well as external systems.

While it could be possible to manually manage the set of terms and taxonomies that are used across the ODS infrastructure and the network of interconnected repositories of learning resources, this would be a very demanding task. Curators from each repository would be required to invest many human resources, engage in extensive communication and error-prone processes to exchange information, align terms, provide translations, updates, and so on. It has thus deemed important that a software system is built in order to facilitate this task – this system is the ODS Vocabulary Bank.

The main goal of the ODS Vocabulary Bank is therefore to implement an open vocabulary server, which allows collaborative editing, management, evolution, and publishing of vocabularies, thesauri, and taxonomies. The key requirements for the development of the ODS Vocabulary Bank have been identified as following:

1) In order to be able to harvest resources from all repositories that are harvested within the ODS federated infrastructure, curators from each repository need to provide a mapping of the vocabulary terms and thematic taxonomy used in their repository to the terms and taxonomy of the ODS metadata schema.

2) The ODS central repository needs to expose the metadata terms and taxonomies as well-defined ontologies, using a linked open data approach, allowing all learning resources to be linked and described against them.

3) A set of different formats is used for the exchange of terminologies (terms and taxonomies) from related repositories: (a) Spreadsheet: a simple spreadsheet file using different columns to categorize terms' classifications; (b) OWL: a file following the Web Ontology Language specification; (c) SKOS: the W3C recommendation for representation of thesauri, taxonomies, or any other type of structured controlled vocabulary (Simple Knowledge Organization System). It has been identified as essential for the ODS Vocabulary Bank to be able to import terminologies from files that follow each of these three formats.

4) The ODS Vocabulary Bank needs to address the multilingual requirements of ODS, by facilitating translations of all managed terms and taxonomies. To this end, all necessary relations between relevant terms in different languages need to be easily identified, managed and exposed to the other components of the ODS infrastructure. Moreover, the system itself ought to provide a multilingual user interface, to allow curators from repositories across Europe to use it in their preferred language.

5) As a single point of reference, the ODS Vocabulary Bank needs to provide access to all managed terms and their relationships through a SPARQL endpoint, following the Linked Open Data approach and enabling integration with other applications and services of the semantic web.

6) The system should feature a user-friendly, multilingual user interface that facilitates an intuitive overview and search through the managed vocabularies and a specific web page with a well-defined URL holding the detailed description of each term and all its relations to other terms.

IV. RELATED WORK

After the initial requirements specification, the development team engaged in a research of the state of the art in terms of technology, standards, and protocols, as well as the associated choices for the implementation of the vocabulary server. The results of this activity are briefly described in the following paragraphs.

The *Vocabulary Bank for Education* (VBE) [2] is a terminology management system for publishing vocabularies, concept schemes, data definitions, taxonomies, and thesauri. The VBE is designed to support a range of exchange formats and exposes all historical information about the terminology. The VBE is an outcome of the ASPECT (Adopting Standards and Specifications for Educational Content) EU-funded project and is further developed by Knowledge Integration with the brand name "Lexaurus Bank" [3].
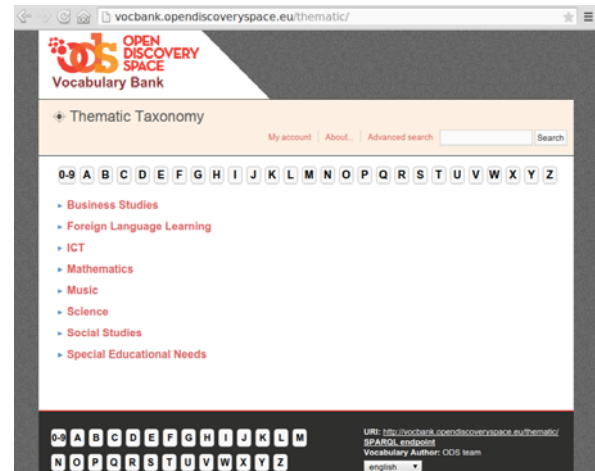

Fig. 1. The main page of a taxonomy managed by the ODS Vocabulary Bank

VocBench [4] is a web-based, multilingual, editing tool that manages thesauri, authority lists, and glossaries using SKOS. It is developed by the Food Agriculture Organization (FAO) of the United Nations (UN) and its partners. VocBench supports a growing set of communities, including the distributed group of terminologists who manage AGROVOC, a multi-lingual thesaurus of 40,000 terms in over 20 languages. Developed in Java, VocBench's latest production release has undergone a major rewriting, rewiring the RDF backend on Semantic Turkey, an RDF framework for Knowledge Acquisition and Management developed by the ART team of the University of Rome Tor Vergata.

iQvoc [5] is a vocabulary management tool that combines easy-to-use graphical user interfaces with semantic web interoperability. It supports a number of features including: import of existing vocabularies from a SKOS representation;

multilingual browser-based display and navigation; editorial features for registered users; publishing in the semantic web. iQvoc is implemented with the Ruby on Rails web framework and is licensed under the Apache License, Version 2.0.

TemaTres [6] is an open source vocabulary server to manage, maintain, and exploit formal representations of knowledge. It is able to manage relations between terms across different vocabularies. It supports the lifecycle of vocabulary management and terms can be marked as candidate, accepted or rejected. TemaTres enables export of the vocabularies and classifications in RDF format (SKOS-Core) and XML format (Zthes, TopicMaps, Dublin Core, and more). TemaTres is built with the PHP web server scripting language, uses the MySQL database engine and is licensed under the GNU GPL.

MoKi [7] is a collaborative tool for modeling ontological and procedural knowledge in an integrated manner, based on the MediaWiki open source wiki package that was originally written for Wikipedia. The main idea behind MoKi is to associate a wiki page, containing both unstructured and structured information, to each entity of an ontology and process model. MoKi's units have been implemented upon a LAMP (Linux-Apache-MySQL-PHP) platform, with the addition of the Java language that has been used for implementing a part of the Ontology Import unit.

The *PoolParty Thesaurus Server* [8] is a tool for the creation of taxonomies, thesauri, and knowledge graphs on the market. It combines methods based on linked data, text mining, and knowledge engineering. It is available in three versions, all of which are offered under a commercial license with an increasing, four-figured price tag.

After our state of the art research and taking into consideration the requirement for open source software as the basis for the implementation of the ODS Vocabulary Bank, it was clear that we need to rule out the options of the VBE, as well as the PoolParty Thesaurus Server. MoKi has been successfully used in the EU-funded "Organic.Edunet" and "Organic.Lingua" projects [9], but the source code was not distributed at the time of our research and it was not directly reusable for implementing the planned functionality. The same held true for the Java-based VocBench. TemaTres and iQvoc both offer similar functionality and their source code is freely available. Our final choice has been TemaTres, based on the following criteria: (i) it uses the PHP scripting language, which was more familiar to our development team than the Ruby language; (ii) it has been successfully deployed in many projects; (iii) it offers responsive feedback from its developer community; and (iv) it features a well-documented source code.

V. IMPLEMENTATION APPROACH

The ODS Vocabulary Bank is based on the TemaTres open source software and implements a web-based system that serves all tiers and workflows for the management of the terms and taxonomies of the metadata schemes related to the ODS infrastructure. Meeting the requirements set in section 3, it facilitates authorized users for editing and organizing the terms, their descriptions, translations and relations. Additionally, it assists in linking terms between associated vocabularies and supports Semantic Web technologies for exporting the ODS terms in Linked Data formats. Through the ODS Vocabulary Bank, all terms of relevant vocabularies are publicly available to every interested party, readily accessible with any modern web browser, as well as via well-documented web services. Furthermore, the ODS Vocabulary Bank enables a complete export of the vocabularies and taxonomies using well-established specifications and file formats (SKOS-Core, Zthes, TopicMaps, Dublin Core, etc.).
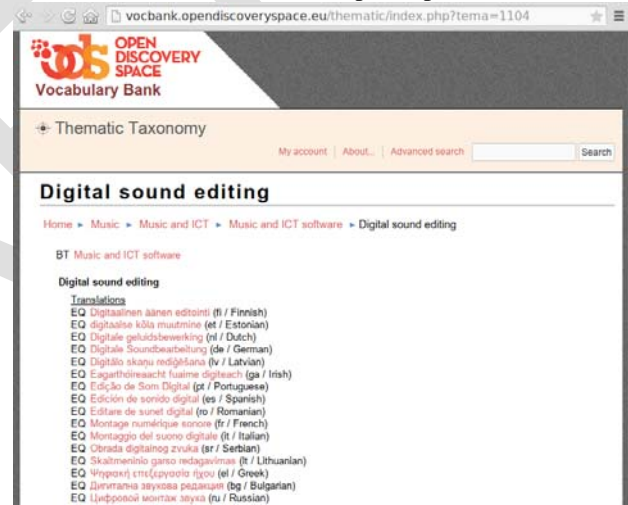

Fig. 2. The page of the term "Digital sound editing"

TemaTres has been adapted and extended, including:
- new functionality for the management terms' translations
- enhancement of the user interface for a more user-friendly look & feel and a responsive, mobile-friendly experience;
- support of multilingualism at the front-end to allow usage from an international audience of interested users;
- support export of vocabularies' terms using the JavaScript Object Notation for Linked Data (JSON LD);
- enhancement of the SKOS-Core export;
- enhanced quick navigation by letter and language;
- web services for incremental harvesting of terms, including support for reporting of last updated terms;
- web services for getting translations of any term.

For every vocabulary that needs to be managed, and for each language that needs to be supported, a new instance of the software is being setup. All instances share a common source code and differ only in the configuration of the database, the set of users, and related access rights, and, possibly, the layout of the user interface. In order to indicate translations of terms, authorized users can associate definitions of terms, across different instances of the Vocabulary Bank, using an "equality" relation. It is not an overstatement to describe the overall system as a "federation of vocabularies to support the ODS federated infrastructure".

The front page of the ODS Vocabulary Bank, available at http://vocbank.opendiscoveryspace.eu, presents the index of

```
<rdf:RDF
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
 xmlns:skos="http://www.w3.org/2004/02/skos/core#"
 xmlns:map="http://www.w3c.rl.ac.uk/2003/11/21-skos-
mapping#"
 xmlns:dct="http://purl.org/dc/terms/"
 xmlns:dc="http://purl.org/dc/elements/1.1/">
<skos:ConceptScheme rdf:about="http://vb.ods.eu/thema/">
 <dc:title>Thematic Taxonomy</dc:title>
 <dc:creator>ODS team</dc:creator>
 <dc:description><![CDATA[  ]]></dc:description>
 <dc:date>2013-11-28</dc:date>
 <dct:modified>2014-05-22 18:47:41</dct:modified>
 <dc:language>en</dc:language>
</skos:ConceptScheme>
<skos:Concept rdf:about="http://vb.ods.eu/thematic/1104">
 <skos:prefLabel xml:lang="en">
  Digital sound editing</skos:prefLabel>
 <skos:prefLabel xml:lang="es">
  Edición de sonido digital</skos:prefLabel>
...
 <skos:inScheme rdf:resource="http://vb.ods.eu/thematic/"/>
 <skos:broader
rdf:resource="http://vb.ods.eu/thematic/1103"/>
 <dct:created>2013-11-28 16:32:34</dct:created>
</skos:Concept>
</rdf:RDF>
```

Fig. 3. SKOS-Core export of a selected term from the ODS Vocabulary Bank

all the taxonomies and vocabularies hosted in the Vocabulary Bank. The "Thematic Taxonomy" item holds the set of terms that comprise the thematic, tree-structured taxonomy for describing learning resources (Fig. 1). Other items, include the "Vocabularies", which holds all the sets of controlled vocabularies with the terms that are used for specific elements of the metadata description of learning resources and the "National Taxonomies", which includes all taxonomies from each nation holding the terms for the disciplines of "History", "Religion", and "Language & Literature". The "Repositories' Vocabularies" item contains vocabularies and taxonomies as they are specific for certain repositories aggregated by ODS.

Selected metadata experts from each associated repository have access to the management back-end, where they can use the provided functionality to manage the terms of their metadata application profile. The description of each term is available through a user-friendly web page (Fig. 2), or can be directly exported in one of the supported formats so that it can be handled and digested by other software systems (Fig. 3). In particular, the ODS Vocabulary Bank offers many alternatives to be integrated with other systems. Each taxonomy and/or vocabulary is accessible through web services over HTTP, it can be made available through batch export in a single file, and a SPARQL endpoint allows for querying the vocabulary bank and requesting for the representation of each term in machine readable format. Regarding the web services, these are listed in the web page "/services.php", which holds the index of the "verbs" of the Application Programming Interface (API), along with their description, the expected input, and the output. They are pretty well self-descriptive and the output can be either in XML or JSON format. In terms of the batch export, the main and fully-supported format is SKOS-Core RDF/XML. This way a single file can hold all the information regarding a specific taxonomy and/or vocabulary. The SPARQL endpoint, provided at "/sparql.php", can respond to complex queries structured in SPARQL syntax and provide specific results. Finally, each term is provided in two

machine readable format, SKOS-Core RDF/XML, and JSON-LD.

VI. CONCLUSIONS AND FUTURE WORK

The usage of the ODS Vocabulary Bank has served well its objective to facilitate the management and maintenance of a large set of vocabularies and taxonomies associated with the metadata schemas of the repositories participating in the ODS federated infrastructure. Moreover it as allowed for additional added-value services to be developed. For example, a visual terminology browser can provide an intuitive overview of the hierarchical structure of a thematic taxonomy, so that users can directly navigate to their topic of interest and browse for resources in each selected topic. Or users can search for resources using keywords and terms in their preferred language, whereas the system can use the information of the ODS Vocabulary Bank to search for resources with equivalent terms in their description and, thus, create a much richer result-set for the original query. A powerful autocomplete service has allowed to step-up the effectiveness of user searches. Last, but not least, other systems can readily use the offered services through the well documented API in order to create new services or enhance existing ones.

In terms of future work, the development team at CTI Diophantus is planning to enhance the system in the following areas: better support for versioning of vocabularies; integration of a service for providing globally addressable persistent unique identifiers; and support for more human-friendly "clean" URLs. Moreover, the source code of the existing and further modifications to the TemaTres source code shall be contributed back to the Free Open Source Software (FOSS) community through a publicly available source code repository, at http://git.dschool.edu.gr.

REFERENCES

[1] N. Athanasiadis, S. Sotiriou, P. Zervas, and D. G. Sampson, "The Open Discovery Space Portal: A Socially-Powered and Open Federated Infrastructure," in *Digital Systems for Open Access to Formal and Informal Learning*, Springer, 2014, pp. 11–23.

[2] R. Tice, "Aspect Vocabulary Bank for Education," in *Digital Libraries: interoperability, Best Practices and Modeling Foundations, 2009.*

[3] "Lexaurus Bank." Knowledge Integration Ltd, 2013. *Accessed online (Feb 2015)* http://www.k-int.com/products/lexaurusbank

[4] A. Stellato, S. Rajbhandari, A. Turbati, C. Caracciolo, M. Fiorelli, J. Keizer, and M. Pazienza, "VocBench 2.0: a Collaborative Web Application for the Development of Thesauri and Concept Schemes," in *LOD2014.*

[5] T. Bandholtz, J. Fock, A. Wolff, and H. Schentz, *"LOD-ready Environmental Terminology with iQvoc,"* 2011.

[6] A. Gonzales-Aguilar, M. Ramirez-Posada, and D. Ferreyra, "TemaTres: software to implement thesauri," *Professional de la Infromaction*, vol. 21, no. 3, pp. 319–325, 2012.

[7] M. Dragoni, C. Ghidini, and A. Bosca, "Multilingual MoKi: How to Manage Multilingual Ontologies in a Wiki," in *The Semantic Web: ESWC 2013 Satellite Events*, vol. 7955, P. Cimiano, M. Fernández, V. Lopez, S. Schlobach, and J. Völker, Eds. Springer, 2013, pp. 162–166.

[8] M. Bastiano, . Blankman, M. Frenzel, U. Grandin, M. Kertesz, and A. Oggioni, *"EnvThes-interlinked thesaurus for long term ecological research, monitoring, and experiments,"* 2013.

[9] A. Bosca, M. Dragoni, C. Di Francescomarino, and C. Ghidini, "Collaborative management of multilingual ontologies," in *Towards the Multilingual Semantic Web*, Springer, 2014, pp. 175–192