D2.2 - ONTOLOGICAL ENGINEERING TECHNOLOGIES

Co-funded by the Horizon 2020
Framework Programme of the European Union

DELIVERABLE NUMBER	D2.2
DELIVERABLE TITLE	Ontological Engineering Technologies
RESPONSIBLE AUTHOR	Agroknow

GRANT AGREEMENT N.	731001
PROJECT ACRONYM	AGINFRA PLUS
PROJECT FULL NAME	Accelerating user-driven e-infrastructure innovation in Food & Agriculture
STARTING DATE (DUR.)	01/01/2017 (36 months)
ENDING DATE	31/12/2019
PROJECT WEBSITE	plus.aginfra.eu
COORDINATOR	Nikos Manouselis
ADDRESS	17 Grammou Str., Vrilissia GR15235, Greece
REPLY TO	nikosm@agroknow.com
PHONE	+30 210 6897 905
EU PROJECT OFFICER	Mrs. Pilar Ocon-Garces

WORKPACKAGE N. | TITLE	WP2 | Data and Semantics Layer
WORKPACKAGE LEADER	Agroknow
DELIVERABLE N. | TITLE	D2.2 | Ontological Engineering Technologies
RESPONSIBLE AUTHOR	Antonis Koukourikos
REPLY TO	akukurik@agroknow.com
DOCUMENT URL	http://www.plus.aginfra.eu/sites/plus_deliverables/D2.2.pdf
DATE OF DELIVERY (CONTRACTUAL)	30/09/2017 (M9)
DATE OF DELIVERY (SUBMITTED)	29/09/2017 (M9)
VERSION | STATUS	v1.0 | First submission to the EC
NATURE	P (Prototype)
DISSEMINATION LEVEL	PU (Public)
AUTHORS (PARTNER)	Pythagoras Karampiperis, Antonis Koukourikos (Agroknow)
REVIEWERS	Teodor Georgiev (PENSOFT)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Preliminary Tools review	31/05/2017	Agroknow
0.3	Harmonization with Requirements	31/07/2017	Agroknow
0.5	Tool deployment	07/09/2017	Agroknow
0.6	Report setup	15/09/2017	Agroknow
0.7	Report draft finalization	22/09/2017	Agroknow
0.8	Deliverable Review	27/09/2017	PENSOFT
0.9	Deliverable finalization	28/09/2017	Agroknow
1.0	Submission to the EC	29/09/2017	Agroknow

PARTICIPANTS		CONTACT
Agro-Know IKE (Agroknow, Greece)	Agroknow	Nikos Manouselis Email: nikosm@agroknow.com
Stichting Wageningen Research (DLO, The Netherlands)	ALTERRA WAGENINGEN UR	Rob Lokers Email: rob.lokers@wur.nl
Institut National de la Recherché Agronomique (INRA, France)	INRA SCIENCE & IMPACT	Pascal Neveu Email: pascal.neveu@inra.fr
Bundesinstitut für Risikobewertung (BFR, Germany)	BfR Bundesinstitut für Risikobewertung	Matthias Filter Email: matthias.filter@bfr.bund.de
Consiglio Nazionale Delle Richerche (CNR, Italy)		Leonardo Candela Email: leonardo.candela@isti.cnr.it
University of Athens (UoA, Greece)	National and Kapodistrian University of Athens	George Kakaletris Email: gkakas@di.uoa.gr
Stichting EGI (EGI.eu, The Netherlands)	egi	Tiziana Ferrari Email: tiziana.ferrari@egi.eu
Pensoft Publishers Ltd (PENSOFT, Bulgaria)	PENSOFT	Lyubomir Penev Email: penev@pensoft.net

ACRONYMS LIST

CPU	Central Processing Unit
GB	Giga Byte
OWL	Web Ontology Language
RAM	Random Access Memory
RDF	Resource Description Framework
REST	Representational state transfer
SKOS	Simple Knowledge Organisation System
SPARQL	SPARQL Protocol and RDF Query Language
UI	User Interface
VM	Virtual Machine
VRE	Virtual Research Environment

EXECUTIVE SUMMARY

The present report is the first submitted iteration of a living document that will describe progress and evolvement of the AGINFRA PLUS ontology engineering components, i.e. the services that will be incorporated in the overall AGINFRA PLUS architecture and serve the conceptualization design and requirements of the involved research communities.

The current version of the deliverable focuses on the installation and deployment of prominent ontology engineering frameworks that will serve as the baseline for producing the final AGINFRA PLUS ontology engineering components, tailored to the core needs reported by the user groups active within the project and adhering to the requirements of the specified AGINFRA PLUS use cases.

It is expected that, as the use cases are refined and executed, the components will be accordingly updated and extended. Additionally, general developments on the used baseline tools will be monitored and adopted if suitable for the purposes of AGINFRA PLUS. To this end, the report is treated as a living document, with regular submission to the EC of versions that report on significant changes in the ontology engineering prototypes.

TABLE OF CONTENTS

TABLE OF FIGURES

1 INTRODUCTION

The present document provides details on the role and functionality of Ontology Engineering components within the overall AGINFRA PLUS architecture, reports on the initial selection of the technologies to be used and summarizes the extension and training actions foreseen for the following period.

In its general sense, ontology engineering refers to the process of building formal representations of conceptualizations for a given subject / knowledge domain. Ontology engineering tools, thus, are called to provide the means for experts in the respective domain to express accurately and intuitively their intended definitions for the concepts and relations between them.

In the context of AGINFRA PLUS, ontology engineering tools face further, specialized requirements in order to be applicable to the research communities addressed by the project and beyond, after its expiration, and be in line with the modus operandi of the modern information and knowledge exchange practices.

The following sections present the core requirements for the ontology engineering components that will serve AGINFRA PLUS, and report on the currently deployed tools to be tested and extended before being integrated into the overall AGINFRA PLUS architecture.

Furthermore, the report analyses the steps for ensuring that the modifications on the baseline technologies will follow the needs of the served scientific communities as well as, the need for strengthening cooperation, and information and knowledge exchange between these different disciplines.

2 REQUIREMENTS AND CANDIDATE TECHNOLOGIES

As analysed in the Technical Specifications report document (D2.1, submitted at M7 of the project), the Data and Semantics layer incorporates two main components pertaining to the ontological engineering aspects of the platform; namely, the *Ontology Authoring Service* and the *Vocabulary Authoring Service*.

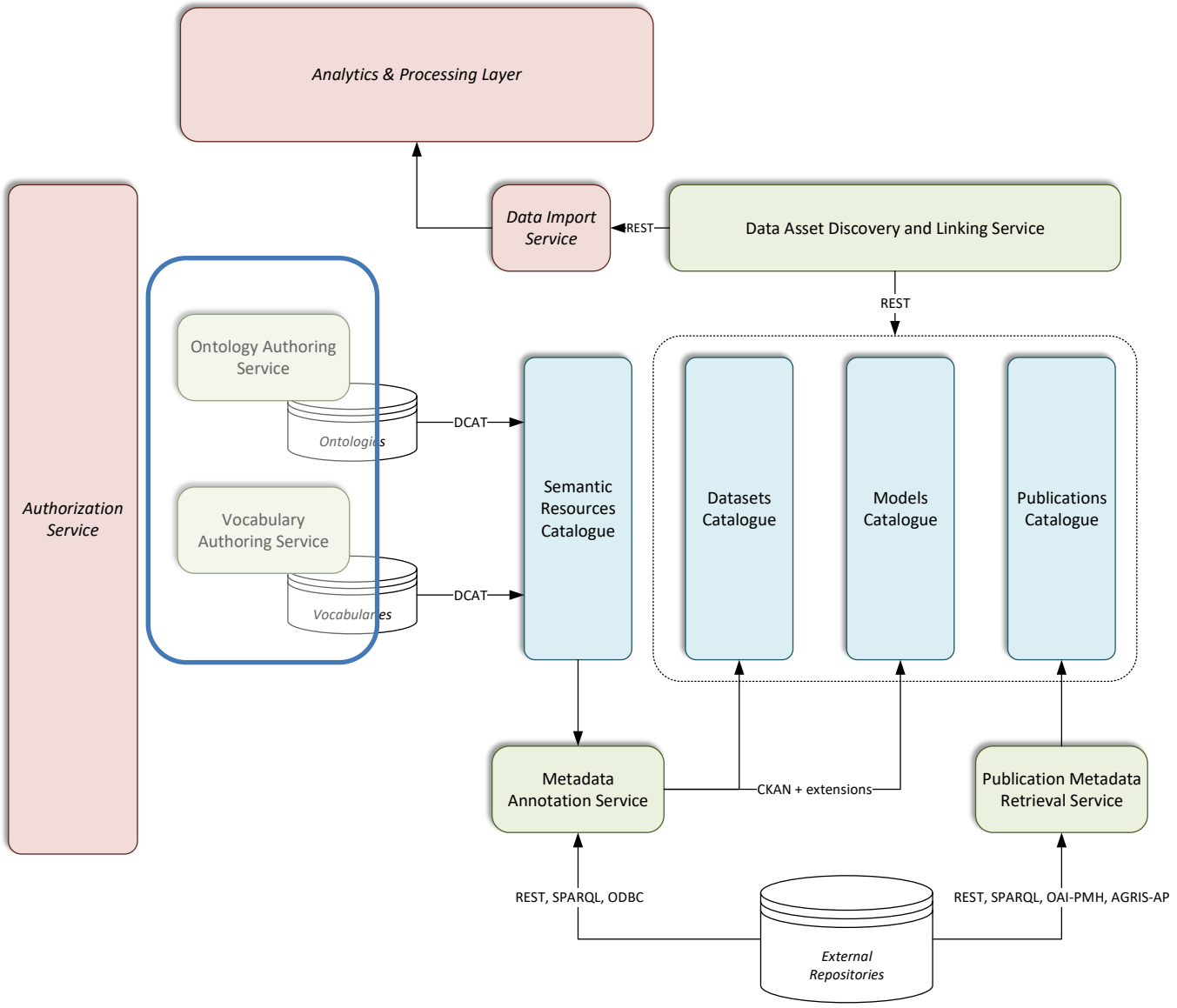

Figure 1: Ontology Engineering Technologies within the overall AGINFRA PLUS Architecture

For identifying technologies suitable for incorporation in AGINFRA PLUS and determining the aspects where extensions or modifications will be required, we use the following set of functional and operational requirements for the services.

Requirement ID	Requirement	Description
Functional Requirements		
C1	Functionality	Degree to which the system supports the basic operations required by the different workflows
C2	Versioning	Degree to which the system's versioning mechanism supports transparency, integrity and consistency

C3	Storage and Export	Support for different storage mechanisms and ability to export in different formats/standards
C4	Mapping	Support for declaring / maintaining associations between different specifications
C5	Multilinguality	Support for labels / comments / descriptions in more than one language
Operational Requirements		
C6	Cooperation	Role support, support for concurrent editing, propagation of updates
C7	Complexity	Ease of deployment / usage / system maintenance
C8	Robustness	Stability of the system during workflow execution
C9	Help System	Degree to which the system provides suggestions / additional information and guidelines during editing / mapping processes
C10	Sustainability/Support	Ease of acquisition, pricing, licensing, support

To cover the functional and operational requirements for the AGINFRA+ ontology engineering services, we will build on top of two major open sourced environments, VocBench[1] and WebProtégé[2].

The following subsections provide a brief overview of these frameworks that will serve the base ground for the AGINFRA PLUS ontology engineering services.

2.1 VOCABULARY AUTHORING SERVICE - VOCBENCH

VocBench is a vocabulary management web-based tool. Its original purpose was the management of the AgroVoc thesaurus. It is actively maintained by the ART group of University of Rome Tor Vergata. Its latest stable version is 2.4.4. An update to VocBench 3 was planned for the end of July 2017, though this has not been provided at the time of the present deliverable's preparation.

VocBench is an actively maintained open-source project. Its installation package is available through the system's website, while the source code is available from the system's Git page[3]. It requires some third-party software components, namely the Apache Tomcat application server[4], a MySQL server[5], and optionally the Ontotext GraphDB[6] triple store installed on an RDF4J server[7]. All the external components are open-source or free-to-use software.

VocBench is a full-feature thesaurus management platform, allowing the creation, import, update and merge of vocabularies. The adopted UI layout is the standard for systems of this nature, with a navigation panel to the right where the available concepts and properties are presented in a tree structure and a main panel where the edit/updated operations are performed.

[1] http://vocbench.uniroma2.it/

[2] https://webprotege.stanford.edu

[3] https://bitbucket.org/art-uniroma2/vocbench2

[4] http://tomcat.apache.org/

[5] https://dev.mysql.com/downloads/mysql/

[6] https://ontotext.com/products/graphdb/

[7] http://rdf4j.org/

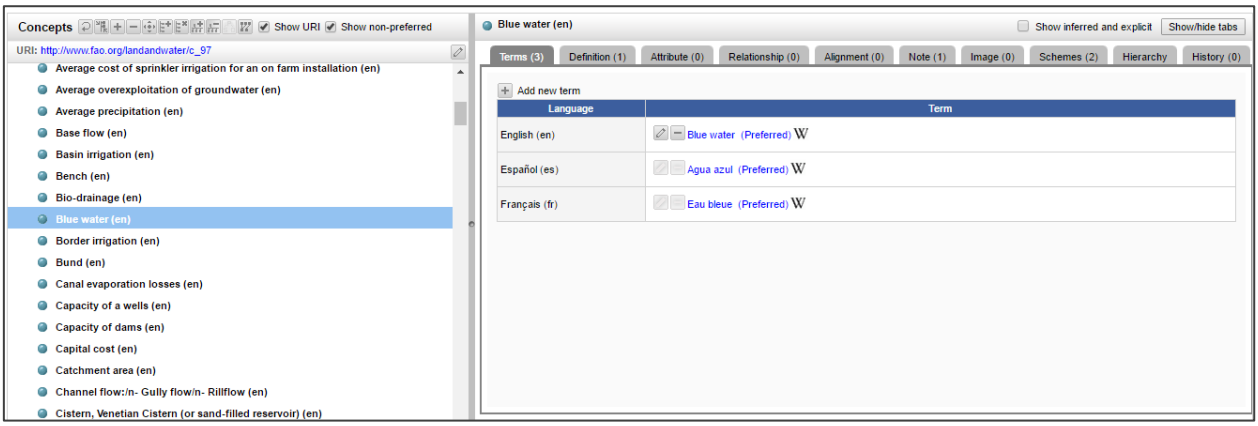

Figure 2: VocBench Concept Editor

VocBench 2 currently supports only the handling of thesauri expressed in SKOS or SKOS-XL. As a consequence, the ontology import functionality injected into the system handles solely the constructs that were related to a SKOS definition (a direct instantiation, a subsumed concept, etc.). However, VocBench 3 will support also the authoring of OWL 2 ontologies.

Querying is performed via a SPARQL editor. There is no functionality for adding predefined queries.

VocBench provides a detailed change history for each project, schema and schema entity managed within the platform. However, there is no rollback functionality provided.

For large-scale projects, VocBench recommends the usage of the Ontotext GraphDB triple store as the underlying repository. The storage mechanism is thus centralized, in the sense that all web clients access and update the same repository.

Regarding the export options available via VocBench, the only supported option at the moment is the export of the created Vocabularies in SKOS or SKOS-XL, in a variety of formats. In addition to the export of an entire project, the user can export specific schemas or even specific concepts included in the active project.

The alignment functionality within VocBench allows the definition of different relations between defined concepts and properties, using the relation levels defined by SKOS. The user is able to search among the schemas imported in the project and select the concept to be aligned and the type of match with the source concept.

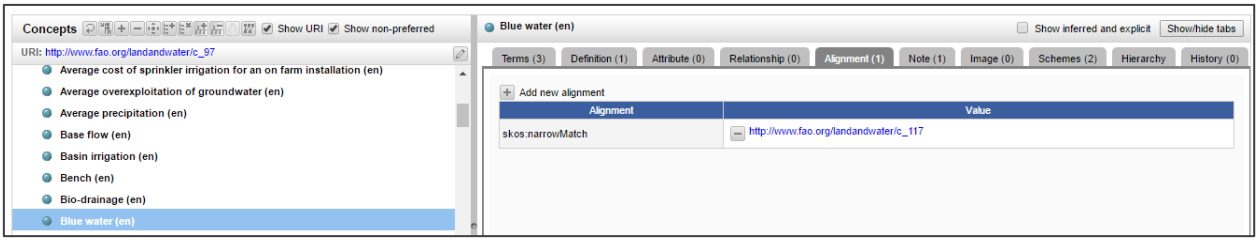

Figure 3: Alignment of concepts in VocBench

The VocBench environment supports four languages (English, Spanish, Dutch, and Thai). Regarding the creation of terms and labels, VocBench supports 44 languages in total. Each user, upon their creation, is associated with their languages of expertise. They are consequently able to create content solely on those languages.

VocBench relies on the Groups/Users structure to organize its users and their rights (permissions). There are multiple predefined user groups as presented in the following figure.

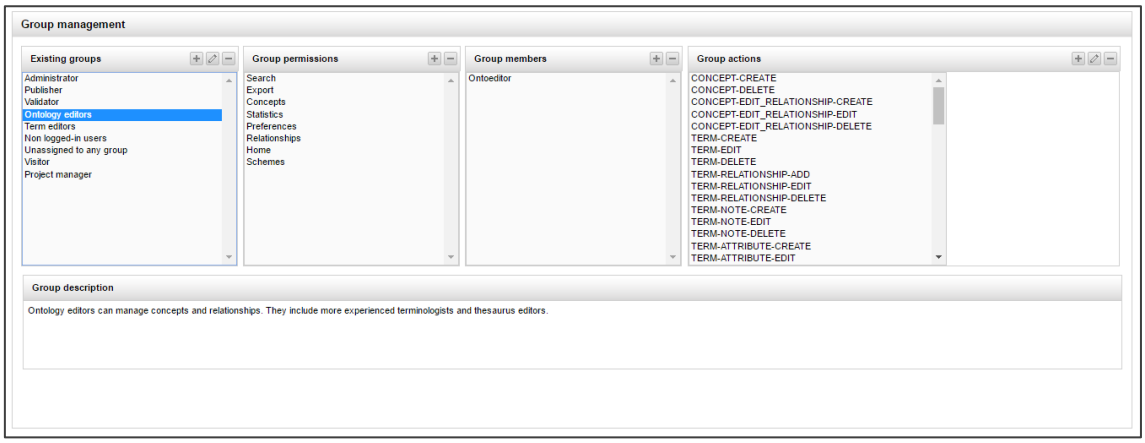

Figure 4: VocBench Group Management

The rights of each user group are defined explicitly by the administrator, and the users are subsequently and mandatorily assigned to one or more groups and one or more projects.

Regarding collaboration, the users assigned to a project view the same repository (in compliance with their rights), so any changes are propagated on-the-fly. There is also an e-mail notification feature included in VocBench (the mail server serving the notifications is external and configurable by the administrator of the platform).

2.2 ONTOLOGY AUTHORING SERVICE - WEBPROTÉGÉ

WebProtégé is a web-based ontology development environment based on the popular Protégé editor. As with VocBench, its web nature allows the provision of fairly extensive collaboration features. It also incorporates tracking and revision history functionality and multilinguality features. WebProtégé supports most of the functionalities of the classic desktop Protégé (and is also cross compatible with the latter), which culminates to a very powerful and complete system. However, it is primarily targeted to expert users with extensive knowledge of the OWL 2 specification and the overall RDF ecosystem.

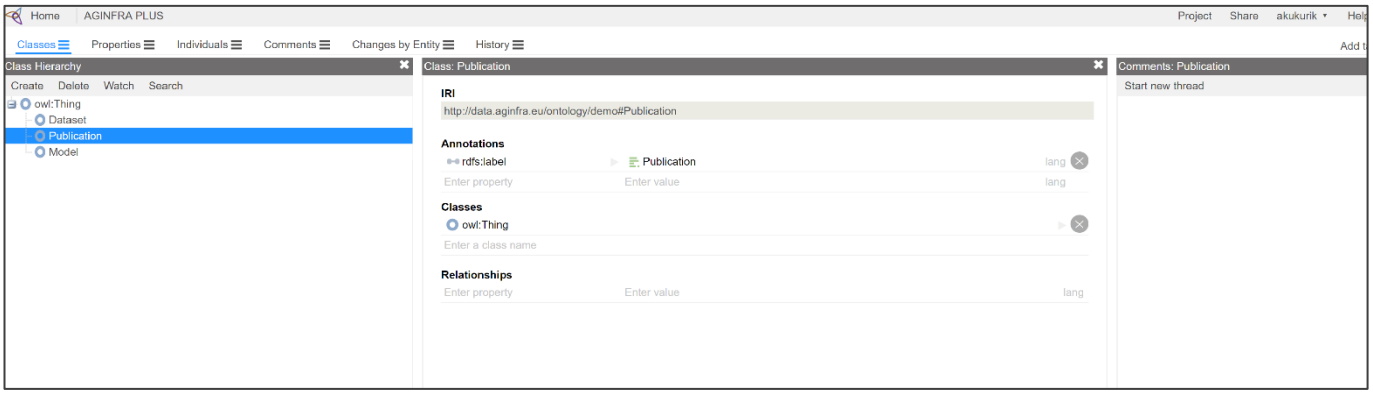

Figure 5 WebProtégé Ontology Creation

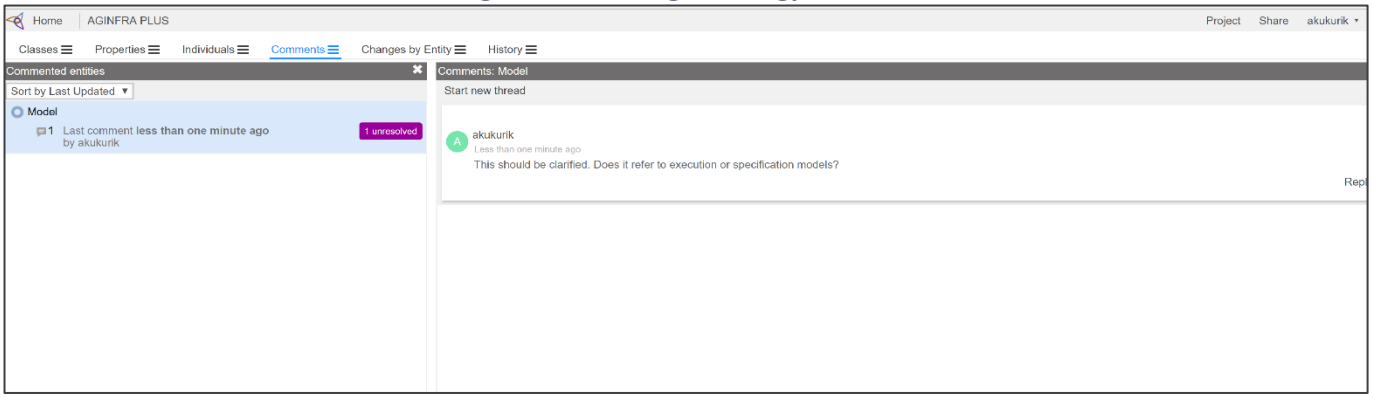

Figure 6: WebProtégé Comment Section

As WebProtégé adheres to the Semantic Web interpretation of what constitutes an ontology – as opposed to the knowledge modelling and philosophical views – it also allows the introduction of instances within the authoring environment, thus allowing the direct and explicit incorporation of data under the designed conceptualization.

From the technical point of view, WebProtégé also relies on some third-party frameworks for deployment and operation. Namely, it also runs on top of an Apache Tomcat container (also supporting alternative containers like GlassFish), and uses MongoDB [8] for preserving some configuration and execution parameters separately for its actual data store.

[8] https://www.mongodb.com/

3 CURRENT DEPLOYMENT

3.1 DEPLOYMENT STATUS

For the first version of the deliverable, stand-alone deployments of the aforementioned frameworks have been installed and configured. The two installations, for the timespan that they will be used as stand-alone services, will be hosted in a dedicated VM hosted by the Greek Research and Technology Network OKEANOS cloud services, with 8 CPU cores and 8 GB of RAM.
While there are no foreseen performance issues for the services, as the technologies are mature and the workload for the purposes of AGINFRA PLUS is considered moderate in terms of number of users and simultaneous access to the services, the resource usage of the two systems will be monitored and cross-checked with their logs in order to identify possibly problematic operations that need to be taken care of in a full deployment setting.
The VocBench installation can be found at: http://83.212.101.14:8080/vocbench-2.4.4/
The WebProtégé installation can be found at: http://83.212.101.14:8080/webprotege/

3.2 THIRD PARTY COMPONENTS

As stated in the previous section, both tools utilize third party infrastructure components for them to be operational at a large scale. However, all third-party software is open-source or offering free, community editions. Specifically, the AGINFRA PLUS deployment uses the free edition of the Ontotext GraphDB server and the community edition of MongoDB. All other used frameworks are available under open source licences, accompanying their distribution packages.

4 NEXT STEPS & ACTION PLAN

The two ontology engineering components that aim to cover the respective requirements from the AGINFRA PLUS platform will undergo usability testing, in order to identify possible further needs of the user communities and apply the required modifications / extensions on their interfaces and functionality. To this end, a set of usage/testing scenarios will be produced, in order to act as tutorials for the components and corresponding digital questionnaires will be shared with the representatives of the end user communities in order to receive feedback on the appropriateness and completeness of the tools.

The immediate next step is the application of the modifications perceived as necessary to cover the needs of the targeted user groups. To this end, existing mining, alignment and classification modules that were the outcomes of related projects like SemaGrow, OpenMinted and the original AGINFRA project will be used as extensions of the baseline frameworks.

Following that, work on the integration of the ontological engineering tools with the VRE will begin, on the principles reported in deliverable D2.1.