International Journal of Semantic Computing
Vol. 9, No. 4 (2015) 459–482

DOI: 10.1142/S1793351X15500099

Herding Linked Data: Semantic Search and Navigation Among Scholarly Datasets

Dimitrios A. Koutsomitropoulos*, Georgia D. Solomou† and Aikaterini K. Kalou‡

High Performance Information Systems Laboratory (HPCLab)
Computer Engineering and Informatics Department
University of Patras, Patras 26500, Greece
**kotsomit@hpclab.ceid.upatras.gr*
†solomou@hpclab.ceid.upatras.gr
‡kaloukat@hpclab.ceid.upatras.gr

Linked Data seem to play a seminal role in the establishment of the Semantic Web as the next-generation Web. This is even more important for digital object collections and educational institutions that aim not only at promoting and disseminating their content but also at aiding its discoverability and contextualization. In this paper we show how repository metadata can be exposed as Linked Data, thus enhancing their machine understandability and contributing to the LOD cloud. We use a popular digital repository system, namely DSpace, as our deployment platform. Without requiring additional annotations that would harden the curation task, educational resources are semantically enhanced by reusing and transforming existing metadata values. Our effort comes complete with an updated UI that allows for reasoning-based search and navigation between linked resources within and outside the scope of the digital repository. Therefore ontological descriptions of resources can now be accessed from within the repository's core context, linked from outside datasets, link to external datasets and get discovered by semantic search.

Keywords: LOD; ontologies; entity extraction; reasoning; learning object repositories.

1. Introduction

Linked and Open Data (LOD) [10] appear to be the "silver-bullet" in the forming Semantic Web ecosystem, which promise to breathe new life to the latter's benefits for real-world web applications. This is often combined with lightweight semantics [11] so that known scalability problems and reasoning inefficiencies could be side-stepped while still retaining some essence of the knowledge discovery capabilities of ontologies. However, tried-and-true systems like digital repositories for educational and other institutions need a little more incentive to embark on such a migration and to get tempted to adopt this new paradigm.

In this paper we show how such institutions can take advantage of Linked Data for their scholarly resources and what they can expect by their exploitation and use. This is exemplified by our work for publishing Linked Data and navigating among resources of a popular digital repository system, DSpace. Resources in DSpace are known as *items* and their descriptions are based on well-defined metadata sets based on standards, such as Dublin Core and Learning Object Metadata (LOM) [15]. This approach for exposing metadata as Linked Data can therefore be extended evenly towards any other metadata-based collection of digital assets or other open repository systems with an educational focus, including Fedora Commons, ePrints and Omeka [1].

In addition, it is shown that it is possible to elicit rich semantics from existing metadata in an automated manner, thus avoiding the need for additional, manual-labor annotations. These semantics can increase the value of linked resources, by allowing new search and navigation possibilities. Based on such premises, we explain the technical means and illustrate use cases for the entire life-cycle of Linked Data, from generating semantic descriptions, to linking with external datasets, to linking to from outside datasets and integrate within the repository's core.

Semantics play a crucial role and this is exhibited by an OWL 2 inference-based knowledge acquisition mechanism that lies at the core of this implementation, aka *Semantic Search for DSpace* [18]. Further, Linked Data provision requires a careful replication design for existing resource descriptions; a data linking and resolution mechanism; and a content negotiation strategy to serve information both to end-users and machines. To our knowledge, the combination of a reasoning-based mechanism with a Linked Data service that augments repository assets by extracting entities, assigning identifiers and injecting external references has not been proposed before.

Next, in Sec. 2 we review the adoption of Linked Data principles for exposing educational content on the cloud. In Sec. 3 we describe the process of extracting and publishing Linked Data out of DSpace resources; Sec. 4 presents Semantic Search for DSpace and outlines its main features; Sec. 5 discusses the Linked Data Provider and how we provide access to the well-known DBpedia dataset; Sec. 6 describes alternative scenarios for acquiring access to internal and external resources through our educational repository and comments on dataset metrics; and Sec. 7 summarizes our conclusions and future work.

Semantic Search is hosted and maintained as a GitHub project[a] and is listed as an official DSpace add-on.[b] A working demo of this implementation is also available.[c]

2. Linked Data for Educational Content

Linked Data have been extensively used during the recent few years as a means to leverage the vision and accomplish the scope of the Semantic Web. In summary, the concept of Linked Data focuses on the creation of typed links between different data sources by using the Web. Linked Data, from the technical aspect, are data with

[a] https://github.com/swigroup/dspace-semantic-search/wiki
[b] https://wiki.duraspace.org/display/DSPACE/Extensions+and+Addons+Work#Extensionsand-AddonsWork-SemanticSearchforDSpacev2.0
[c] http://apollo.hpclab.ceid.upatras.gr:8000/dspace-ld

explicitly defined meaning, published on the Web in a machine-readable format (data in RDF or the Web Ontology Language, OWL). These data are linked to other external datasets and can also be linked to from external datasets. The result of this connection is the Web of Linked Data. 'Linked Data principles', a set of best practices by Berners-Lee, define how to publish and connect data on the Web, so that all published data becomes part of a single global data space [5].

The adoption of the Linked Data principles, especially in educational repositories and information sources of relevance to learning and teaching, would make educational information easier to address, aggregate and reuse for various purposes [1]. The Linked Universities alliance[d] is an attempt to gather and list all universities, schools and research centers that adopt the Linked Data model. Members of this list are, among others, the National Research Council (CNR, Italy) and the UK Open University.

Besides, there are several prominent projects that are moving towards this direction. For example, LinkedUp (Linking Web data for education) [3] is a European project that pushes forward the exploitation and adoption of public, open data available on the Web, in particular by educational organizations and institutions. In order to aid the use of open and Linked Data, the project has created the LinkedUp Dataset Catalog (or Linked Education Cloud)[e] which collects and makes available all sorts of data sources of relevance to education.

The mEducator [8] is another project which implements and critically evaluates existing standards in order to enable specialized state-of-the-art medical educational content to be shared and re-used across European academic institutions. A particular outcome of this attempt is a Linked Data-compliant dataset (mEducator Linked Educational Resources[f]) describing educational resources according to a well-defined RDF schema (mEducator Resource RDF schema [20]). Next, the Organic.Edunet portal [19] is a central point of access to the resources stored by a federation of repositories hosting thousands of digital learning resources about organic agriculture and agroecology.

The aforementioned projects' datasets are either directly related to educational institutions or they can be used in teaching and learning scenarios. Nevertheless, it is common to use DBpedia or other big datasets as "linking hubs". Following this approach, one can ensure that such datasets are commonly used by other datasets, which automatically leads to a plurality of indirect links [7].

A more specific attempt to enhance a digital repository system — typically used for building institutional and educational repositories — with the Linked Data facility appears with the recently released version of DSpace (v5.0) [4]. Nevertheless, this approach lacks the Semantic Search capability as well as the content negotiation procedure when providing the requested content. As a result, resources are exposed as Linked Data only with their standard RDF format, something which makes difficult for repository users to further search and navigate among interlinked information.

[d] http://linkeduniversities.org/
[e] http://data.linkededucation.org/linkedup/catalog/
[f] http://www.meducator.net

3. Linking Data for DSpace

3.1. *Publication and linking of entities*

The Linked Data principles foster the idea of an interlinked 'Web of Data'. In our context this means that resource URIs need to be dereferenceable, to provide meaningful information for users and services alike and to give references (or links) to other related entities whenever possible.

In DSpace the main unit of information is the 'item', i.e. a publication or learning object that is described with a set of metadata based on Dublin Core (DC) and extended with LOM. For each item, a metadata record is created and maintained in the DSpace system, holding for each piece of its metadata the appropriate value, as it is filled during the submission process. The schematic representation of a fragment of an item's metadata record can be seen in Fig. 1.

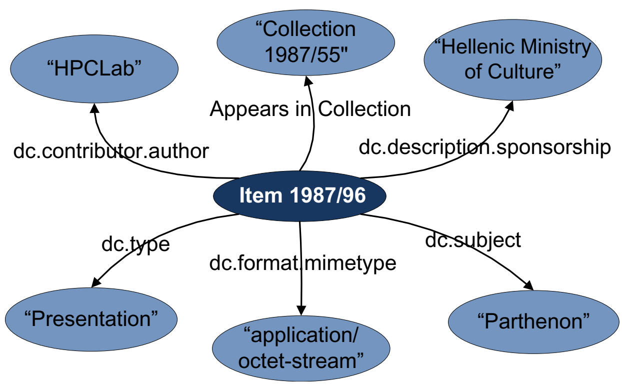


Fig. 1. Example repository item and its metadata representation in the DSpace system.

The extraction and mapping of these metadata to a machine understandable format, such as OWL 2 [12] would permit the identification of additional implicit entities and would allow us to assign resolvable URIs to them too. This process is described in detail in Secs. 3.2 and 3.3. Further, these entities can be linked together or refer to other external datasets like DBpedia (see Sec. 5.2). Figure 2 illustrates a sample instance of the resulting DSpace ontology and the way this instance gets interlinked with other entities and/or datasets.

Using the Jersey framework,[g] the reference implementation of the Java API for RESTful services, both HTML as well as RDF/XML representations are accommodated and the following URI pattern has been established:

- `http://{repositoryURL}/semantic-search/resource/{entity-id}`
- `http://{repositoryURL}/semantic-search/page/{entity-id}(HTML)`
- `http://{repositoryURL}/semantic-search/data/{entity-id}(RDF)`

[g] https://jersey.java.net/

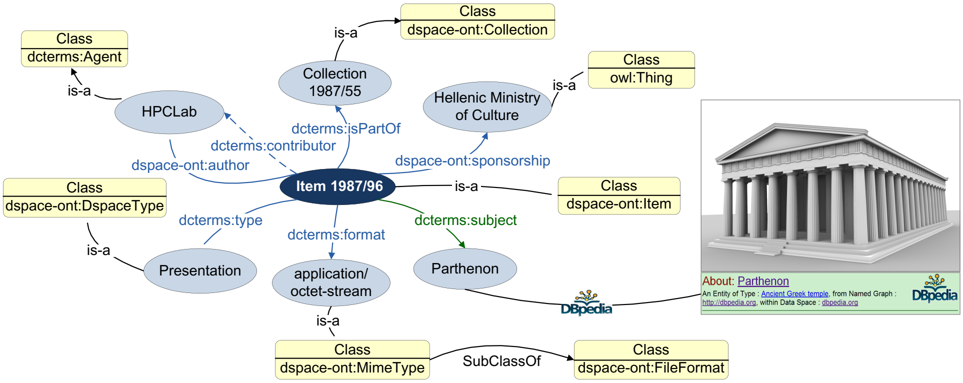


Fig. 2. Example repository item and its relationships to other entities.

Moreover, these URIs are dereferenceable by using Jersey's content negotiation capability and performing HTTP *303 See Other* redirects.

3.2. *The DSpace ontology*

The DSpace ontology is the knowledge base of the Semantic Search plugin. The main idea is to enable export of as much available DSpace metadata as possible and then to translate them into full-fledged RDF/OWL triples. In this process, certain implicit entities (like items, collections, authors, sponsors) are reified and become individual nodes themselves instead of mere literals.

This reification is performed on-the-fly, based on the interoperable system's mechanisms for exporting resources' metadata through OAI-PMH [17]. In essence, metadata are first harvested and then transformed by means of an XSLT to the required OWL/RDF representation. The vocabulary used in these triples is specified by the DSpace ontology.

The DSpace ontology is based on the qualified Dublin Core (QDC) RDF schema and is in fact comprised of several ontology documents that import each other, thus forming an incremental semantic application profile. All these documents are kept and maintained separately from the application's code base and are available at: `http://swig.hpclab.ceid.upatras.gr/dspace-ont/` which serves also as the main vocabulary namespace.

Since DSpace allows the extension and configuration of its metadata schema, we have also extended it with several LOM metadata. At least for the part used in the implementation, a reference LOM ontology has also been devised. As a result, the translation to OWL takes also into consideration LOM metadata and their mapping to QDC.

The main idea behind this mapping is to consider first those LOM elements having an explicit counterpart in the DC elements set and map them accordingly. For example, LOM elements like *Title*, *Language* and *Contribute* were directly mapped to `dc.title, dc.language and dc.contributor` respectively. The exact correspondence between these elements can also be found in the IEEE LOM specification [15]. Remaining LOM elements, for which no such explicit correspondence exists, are carefully mapped to those QDC properties that seem as a best match for their interpretation. The decision was made after considering both the LOM specification and the work suggested by IEEE LTSC [16], which proposes a potential LOM to DC Abstract Model (DCAM) mapping.

3.3. *Exposing educational metadata into OWL*

The replication strategy we follow in order to expose educational metadata into OWL is comprised of two phases: *Phase A* contains the metadata extraction process whereas *Phase B* includes the translation of the ingested metadata into OWL.

During Phase A, the first step towards providing Linked Data is to unleash resource information and metadata that are hidden within databases. The point here is to export as much information as possible and especially to tackle additional elements and not just the limited set of defaults. To maximize the semantic value of exported metadata as well as to maintain interoperability, a careful and elaborate process has to be conducted. This process includes an exhaustive mapping of the inherent DSpace metadata application profile to the DC metadata terms (DCTerms). What is more, because our educational repository has been enhanced with the IEEE LOM metadata in order to effectively characterize scholarly resources, a mapping of the adopted LOM profile to the DCTerms has also been applied. Therefore, the implemented mapping (shown in Table 1) provides for the following:

- Other than the default, additional elements are exported.
- Map everything under the DCTerms namespace, rather than mixing DC with DCTerms, which is generally not advisable [6].

Table 1. The implemented mapping of DSpace internal metadata to DCTerms.

DSpace internal metadata representation	Provided mapping	Notes
`dc.contibutor.author`	`dspace-ont:author`	'author' is not compatible with QDC. However it is a subproperty of 'contributor'
	`dcterms:contributor`	
`dc.language`	`dcterms.language`	
`dc.publisher`	`dcterms.publisher`	
`dc.subject`	`dcterms.subject`	We use the DCTerms namespace
`dc.title`	`dcterms.title`	
`dc.type`	`dcterms.type`	
`dc.identifier.other`	`dcterms:identifier`	

Table 1. (*Continued*)

DSpace internal metadata representation	Provided mapping	Notes
`dc.identifier.uri`	`dcterms:identifier type= http://www.w3.org/2001/XMLSchema#anyURI`	Literals get typed when possible
`dc.description`	`dcterms:description`	
`dc.title.alternative`	`dcterms:alternative`	Not exposed by default
`dc.description.provenance`	`dcterms:provenance`	
Bitstream metadata	`dcterms:format` `dcterms:extent`	Not exposed by default
`dc.relation.ispartofseries`	`dcterms:isPartOf`	Expresses part of series
`dc.description.sponsorship`	`dspace-ont:sponsorship` `dcterms:description`	
`lom.annotation`	`dcterms:description`	
`lom.context`	`dcterms:educationLevel type="lom:Context"`	
`lom.difficulty`	`dcterms:type type="lom:Difficulty"`	
`lom.intendedenduserrole`	`dcterms:audience type="lom:IntendedEndUserRole"`	
`lom.interactivitytype`	`dcterms:instructionalMethod type="lom:InteractivityType"`	LOM specific mapping
`lom.learningresourcetype`	`dcterms:type type="lom:LearningResourceType"`	
`lom.typicallearningtime`	`dcterms:extent type="lom:TypicalLearningTime"`	
`lom.status`	`dcterms:type type="lom:Status"`	
`lom.version`	`dcterms:hasVersion`	

– Provide additional Learning Object mappings to DCTerms (in case LOM metadata exist).
– Assign types to non-literal values.

Mapped metadata can then be exposed and harvested through the OAI-PMH interface.

Having exported as much available DSpace metadata as possible, we move to Phase B of our replication strategy. Our next step is to translate the exposed metadata into full-fledged RDF/OWL triples. To this end, an appropriately constructed XSL transformation is triggered, giving the desired result.

During this semantic translation, certain implicit entities (Table 2) are reified and become individual nodes themselves instead of mere literals, thus resulting into an OWL 2 DSpace ontology. Most of these entities, as shown in Table 2, are assigned resolvable identifiers, so that it would be easy for them to get dereferenced within the application.

In addition, this is when URLs to DBpedia are injected, in order to enrich reified entities such as authors, contributors, sponsors and item types. `foaf:page` links are

Table 2. Reified entities and their identifier.

Reified entity	Identifier
Item	`http:// {repositoryURL}//semantic-search/resource/{item_handle}` *e.g. http://apollo.hpclab.ceid.upatras.gr:8000/dspace-ld/semantic-search/resource/123456789/71*
Collection	http:// {repositoryURL}//semantic-search/resource/{collection_handle} *e.g. http://apollo.hpclab.ceid.upatras.gr:8000/dspace-ld/semantic-search/resource/123456789/69*
Community	`http:// {repositoryURL}//semantic-search/resource/{community_handle}` *e.g. http://apollo.hpclab.ceid.upatras.gr:8000/dspace-ld/semantic-search/resource/123456789/54*
Author	`http:// {repositoryURL}//semantic-search/resource/{author}` *e.g. http://swig.hpclab.ceid.upatras.gr/dspace-ont/Solomou_Georgia*
Contributor	`http:// {repositoryURL}//semantic-search/resource/{contributor}` *e.g. http://swig.hpclab.ceid.upatras.gr/dspace-ont/Solomou_Georgia*
Sponsor	`http:// {repositoryURL}//semantic-search/resource/{sponsor}` *e.g. http://swig.hpclab.ceid.upatras.gr/dspace-ont/European_Commission*
LOM specific entity	`http://ltsc.ieee.org/xsd/LOM/{entity}` *e.g. http://ltsc.ieee.org/xsd/LOM/Expositive*
DCTerms entity	`http://purl.org/dc/terms/{entity}` *e.g. http://purl.org/dc/terms/DDC*
DSpace Item type	`http://swig.hpclab.ceid.upatras.gr/dspace-ont/{DSpace_item_type}` *e.g. http://swig.hpclab.ceid.upatras.gr/dspace-ont/Article*

Table 3. Injected URLs and their linking to DBpedia.

Linked entity	Link to DBpedia	Injected during transformation	Call to DBpedia Lookup Service
Contributor	Property: `foaf:page` Value: `http://www.dbpedia.org/resource/{contributor}`	✓	✓
Author	Property: `foaf:page` Value: `http://www.dbpedia.org/resource/{author}`	✓	✓
Sponsor	Property: `foaf:page` Value: `http://www.dbpedia.org/resource/{sponsor}`	✓	✓
DSpace Item type	Property: `foaf:page` Value: `http://dbpedia.org/resource/{type}`	✓	✓
Publisher	—	—	✓
Subject	—	—	✓

explicitly constructed by appending the name label of the entity, during the ontology creation process. To enhance discoverability, a call to the DBpedia Lookup Service[h] is made during runtime, i.e. during user browsing. These entities are summarized in Table 3.

[h] http://wiki.dbpedia.org/Lookup

4. Semantic Search and Navigation for DSpace

In the following Sec. 4.1 we make an overview of the Semantic Search architecture (Fig. 3). We explain the design principles followed and the technology exploited for the implementation of each component. In Sec. 4.2 we outline the main idea behind the Semantic Search interface design, and present its latest features and deployed facilities appearing in the current version 3.0 and onwards.

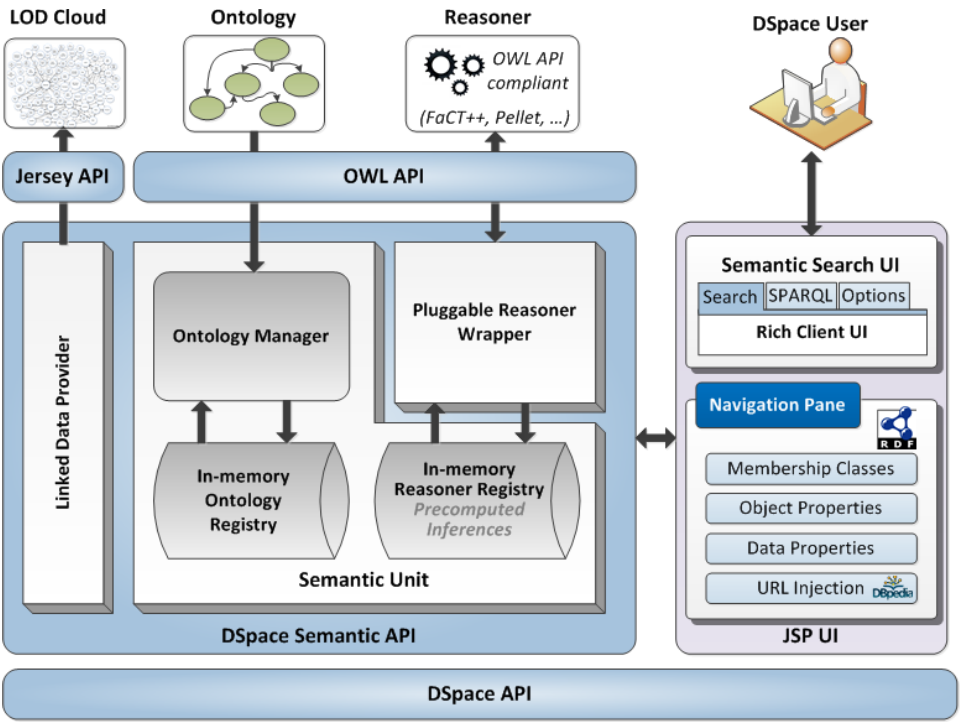


Fig. 3. The Semantic Search architecture and interface.

4.1. *Architectural overview*

One of the design principles of Semantic Search is extensibility and support for different reasoners and ontologies. The interface is backed by the *DSpace Semantic API* that uses OWL API v3 [12] which provides advanced handling for reasoners and ontologies. The DSpace Semantic API is designed to remain independent of the DSpace business logic and to be agnostic to the rest of the user interface or even to the underlying ontology. One of its features is that it gives users the ability to "hot-swap" between reasoners dynamically by implementing a *pluggable reasoner wrapper*. For the time being, any OWL API compliant reasoner can be supported, including out-of-the-box support for Pellet [22], FaCT++ [23] and HermiT [9].

Semantic Search runs by default on top of the DSpace ontology. Nevertheless, it can work against any OWL ontology accessible through the Web. Loaded ontologies can be reused, thus avoiding the overhead of reloading and parsing the whole ontology definition. In particular, when a user asks for a new ontology, this is loaded and stored only once in an *internal registry*. When another user asks for the same ontology, no re-parsing is needed, and the ontology is served from the registry.

Just like ontologies, *reasoner* objects among different users can also be cached in an internal registry and reused. Combined with inference precomputation, this can also improve average performance of multiple individual requests on the same ontology. Precomputation is an initialization phase where the reasoner classifies the ontology, and computes in advance certain reasoning operations, such as instance checking. Precomputation results are then readily available in-memory and, instead of engaging the inference engine each and every time, new queries can take advantage of them, resulting in a significant performance increase. A 'reload' button, provided through the 'Options' tab of the Semantic Search interface, clears registry and reloads both the ontology and the reasoner anew.

The *Linked Data Provider*, which has been deployed on top of the Jersey API, is responsible for allowing uniform access to repository resources. Depending on whether a request is made by a human or a service, the provider produces an HTML page or its RDF analog for the corresponding entity, respectively. In either case, the requester acquires access to an even greater amount of interlinked information leading either to internal resources or to other external datasets.

4.2. *An interface for semantic querying educational resources*

The main idea used by our Semantic Search interface lies behind the deconstruction of a semantic query into smaller building parts (query *atoms*) that are assigned to different fields of a dynamic UI. Query crumbs that are provided through these fields are then assembled by the underlying mechanism to create valid Manchester Syntax expressions [14] (see Fig. 4). Each such expression is an atomic or anonymous class in OWL [21] and its (both inferred and asserted) members are the answers to the query. Search results are presented as a browsable list of linkable entities/resources. Effective querying of the knowledge base is accomplished by interconnecting to an appropriate inference engine, capable of reasoning with OWL 2 ontologies.

When an entity is selected, the corresponding individual's *navigation pane* is produced on the fly, gathering its ontological information which are arranged in *Classes*, *Object Properties* and *Data Properties* (see Fig. 5).

New features in the current version of the Semantic Search interface are described below. These features intend not only to provide advanced searching capabilities to an educational repository but also to help users in the process of constructing semantic queries.

Syntax highlighting for Manchester Syntax: Each word used to compose a semantic expression, as long as it corresponds to a valid query atom, is emphasized

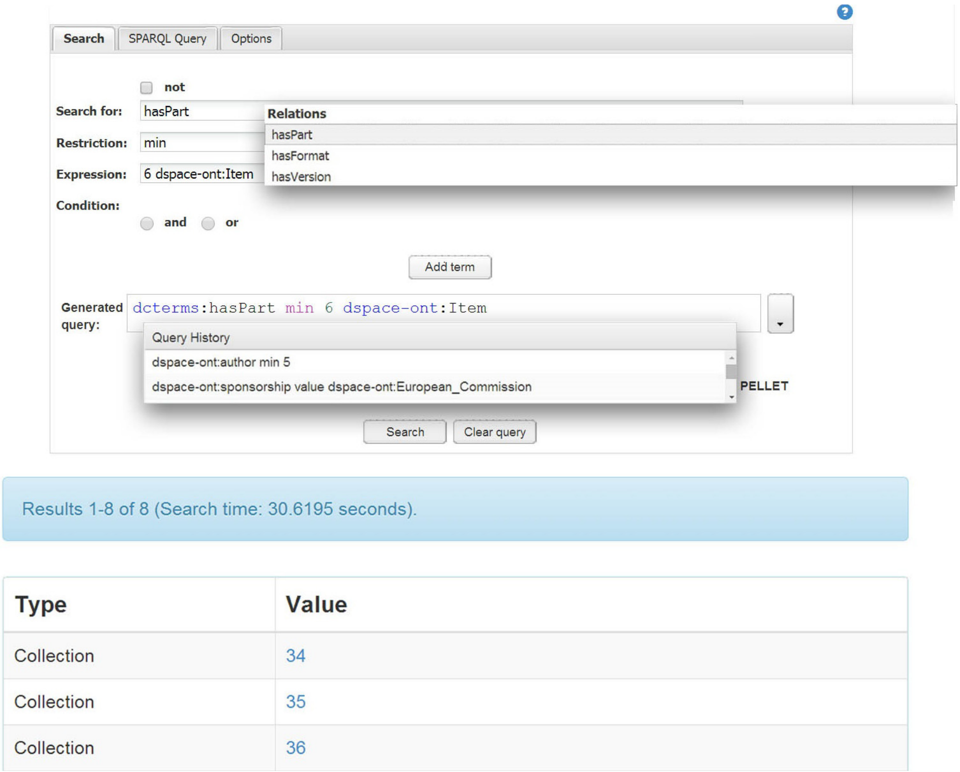
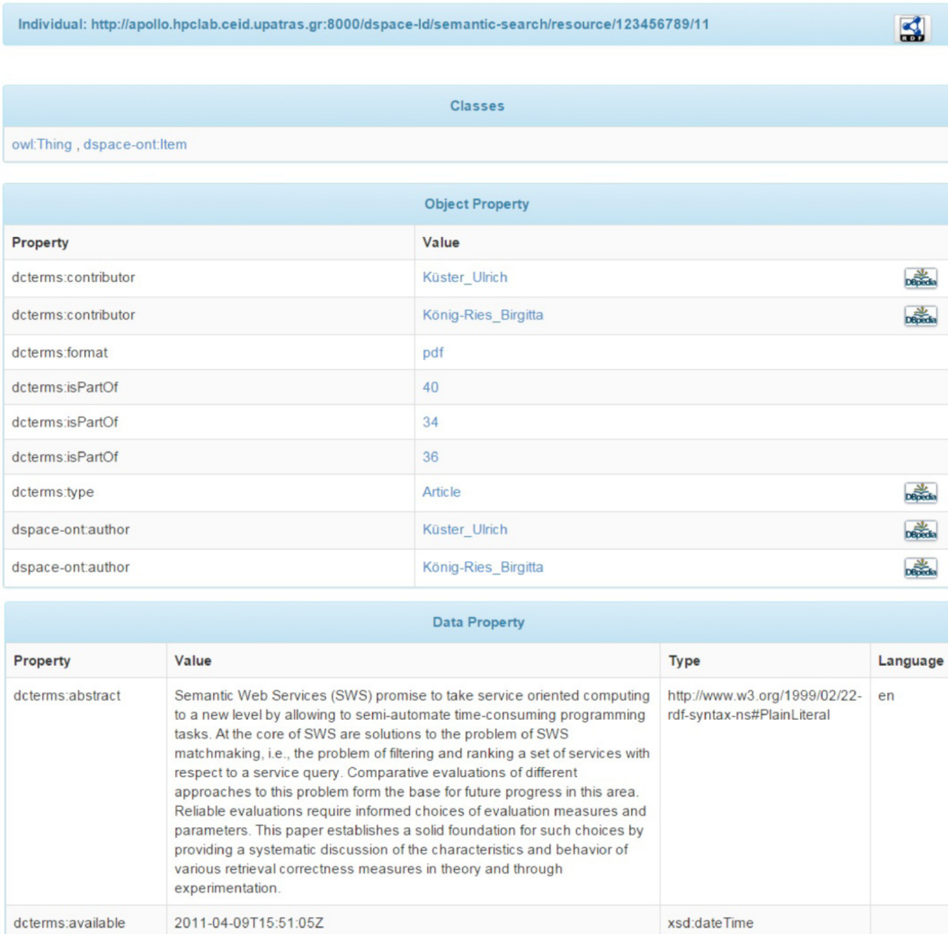


Fig. 4. The query construction mechanisms of the Semantic Search interface.

automatically based on a particular color scheme: Manchester Syntax's reserved keywords (e.g. *some*, *min*, *max*, etc.) are considered a separate colored category, prefixes of the DSpace ontology (e.g. `dcterms`, `dspace-ont`) are assigned a different tone, whereas another color is picked for the Boolean operands *and/or/not* (see Fig. 6). The color scheme applied tends to approximate the colors used by the well-known Protégé ontology editor, which is very familiar to ontology experts. The syntax highlighting facility has been implemented with Codemirror,[i] a Javascript-based editor that is designed to color code-like contents for various languages. This color highlighting feature works as a useful guidance in the process of constructing semantic queries. It facilitates users in providing ontology entities correctly and helps them in avoiding language specific misspellings.

A history subsystem: Every syntactically correct query posed to the Semantic Search interface is stored in a list, designed to keep track of the current user's last ten inputs. The history subsystem discards duplicates and updates the list every time a new successful query is submitted. The history list is scrollable (see Fig. 7) and every time someone wants to resubmit a recently evaluated query or even a slight

[i] http://codemirror.net



Individual: http://apollo.hpclab.ceid.upatras.gr:8000/dspace-ld/semantic-search/resource/123456789/11

Classes
owl:Thing , dspace-ont:Item

Object Property

Property	Value
dcterms:contributor	Küster_Ulrich
dcterms:contributor	König-Ries_Birgitta
dcterms:format	pdf
dcterms:isPartOf	40
dcterms:isPartOf	34
dcterms:isPartOf	36
dcterms:type	Article
dspace-ont:author	Küster_Ulrich
dspace-ont:author	König-Ries_Birgitta

Data Property

Property	Value	Type	Language
dcterms:abstract	Semantic Web Services (SWS) promise to take service oriented computing to a new level by allowing to semi-automate time-consuming programming tasks. At the core of SWS are solutions to the problem of SWS matchmaking, i.e., the problem of filtering and ranking a set of services with respect to a service query. Comparative evaluations of different approaches to this problem form the base for future progress in this area. Reliable evaluations require informed choices of evaluation measures and parameters. This paper establishes a solid foundation for such choices by providing a systematic discussion of the characteristics and behavior of various retrieval correctness measures in theory and through experimentation.	http://www.w3.org/1999/02/22-rdf-syntax-ns#PlainLiteral	en
dcterms:available	2011-04-09T15:51:05Z	xsd:dateTime	

Fig. 5. The Navigation Pane.

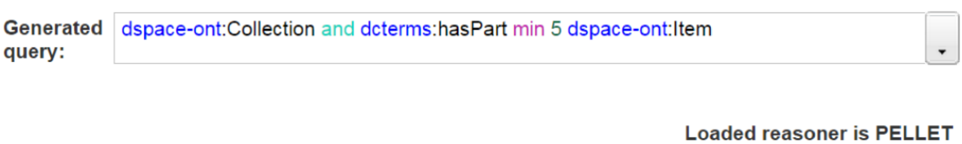


Fig. 6. Syntax highlighting for Manchester Syntax queries.

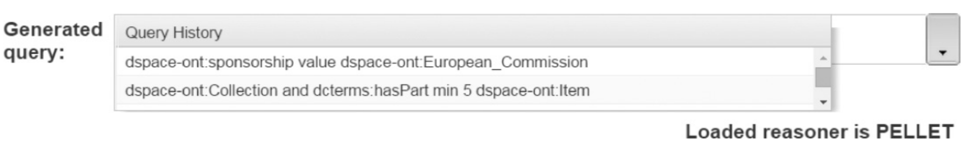


Fig. 7. The history sub-system.

modification of it, they can click on the desired expression, avoiding the process of re-constructing the same query from scratch.

A RESTful Linked Data Provider: It is responsible for exposing resources' metadata as resolvable entities.

A DBpedia URL injection facility: It is responsible for injecting external links, either during the ontology creation phase or during browsing.

Integrated access to Linked Data: By the click of a button, users can get access to an item's Linked Data from within DSpace's native interface.

The Linked Data services are detailed in the following Sec. 5.

5. Linked Data Services

In what follows, we present in more detail the Linked Data Provider (5.1) and then we discuss how certain DSpace entities are linked to the DBpedia dataset (5.2). Next, we describe the newly added functionality within the DSpace core item view page, which provides easy access to the implemented Linked Data facility (5.3).

5.1. *Linked data provider*

The access point of our Semantic Search system to the LOD cloud is the Linked Data Provider. When an entity is selected, the corresponding individual's navigation pane that gathers the resource's semantic metadata is produced on the fly. This ontological information is formed as Linked Data and is presented in either an HTML or an RDF format, depending on whether the request was made by a person or a service, respectively (Fig. 8).

It is important to notice that this holds not only for DSpace items themselves, but also for every other (implicit) entity in the repository model that gets reified during the semantic translation phase.

The main objective of the navigation pane is not only to give a detailed reference to the resources' ontological information (semantic metadata) but also to allow users to further explore and navigate among interlinked information in the LOD cloud. To achieve this, semantic information is structured in the form of resolvable URIs, as much as possible.

All non-literal metadata values are denoted as URIs, which can be dereferenced on the Web. In particular, each class redirects back to the Semantic Search page with the specific class already predefined and its members appearing on the result list. In the case of object properties, the corresponding values are resolvable entities that lead to the particular entity's navigation pane. And even for data properties, where mere text values mostly apply, a datatype of `xsd:AnyURI` is rendered as a resolvable link. This is useful for example to maintain context with the original DSpace item view or with external references, such as the DBpedia Lookup service.

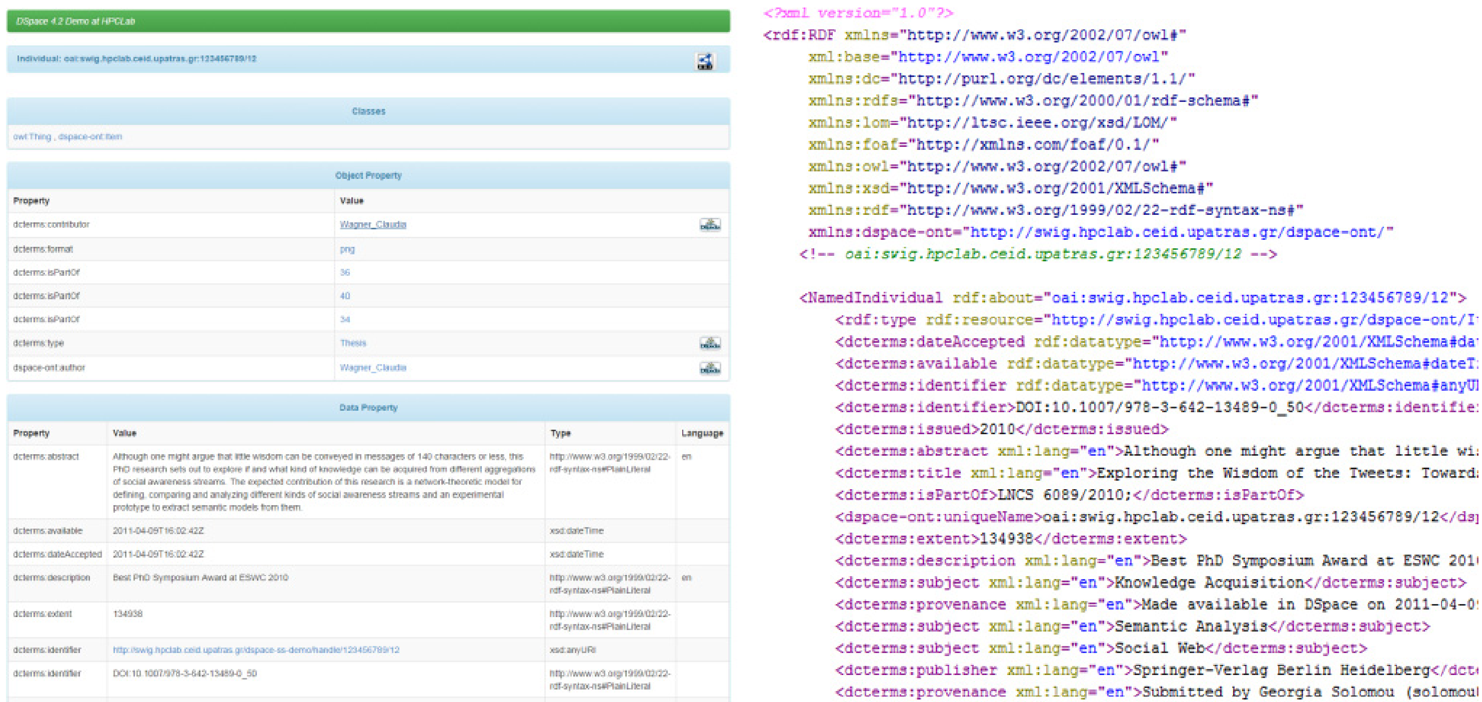


Fig. 8. The HTML (left) and RDF (right) format of an item's semantic information.

Table 4. URI resolution scheme (Linked Data Provider).

ru: the request uri *id*: the resource uri parameter passed to the Linked Data Provider service *local_context*: the base URL path of the application *request_accept_header*: the request accept header for content negotiation
ru → "http://{*local_context*}/semantic-search/resource/{*id*}" **if** ***id*** **is_not_absolute or** *id* **is_not_valid_uri** **then** *id* → "http://{*local_context*}/semantic-search/resource/{*id*}" **if** *request_accept_header* **is** "`text/html`" **then** *ru* → "http://{*local_context*}/semantic-search/page/{*id*}" **request_redirect_303** (*ru*) **if** *request_accept_header* **is** "application/xml+rdf" **then** *ru* → "http://{*local_context*}/semantic-search/data/{*id*}" **request_redirect_303** (*ru*)

The Linked Data Provider implements an algorithm, which ensures that entities will be dereferenced evenly within the navigation pane, while retaining their context at the same time (Table 4).

So for example a request URI of the form: `http://{local_context}/semantic-search/resource/123456789/94`, where 123456789/94 is the DSpace item or collection handle, would resolve to itself (content negotiation aside). This particular form is actually followed when minting item, collection, community, author, contributor, sponsor (reified entities') IRIs during the ontology creation process. What this implies is that such entities pertaining to a specific repository installation would get their own, application-dependent namespace, and be resolvable within the application context. Other ontology IRIs that happen to have their own absolute namespace would also be resolved normally within our context, but maintain their independent namespace, like for example the DSpace types or formats: `http://{local_context}/semantic-search/resource/dspace-ont:Article` where `dspace-ont` amounts to `http://swig.hpclab.ceid.upatras.gr/dspace-ont`.

5.2. *DBpedia linking*

For certain non-literal metadata values, a DBpedia URL correspondence is produced and finally injected in the underlying DSpace ontology, using the `foaf:page` property (see Table 3). This URL is constructed using the term expressed by the corresponding metadata value and is actually a link to a DBpedia resource. In addition, a DBpedia icon is automatically placed next to these entities and next to the object property values (within an item's navigation pane) that correspond to these entities (see Figs. 9 and 10). When the icon is clicked, the DBpedia Lookup service is triggered for this entity, leading to a keyword-based search against DBpedia. This label matching process inevitably includes a certain extent of ambiguity. In order to

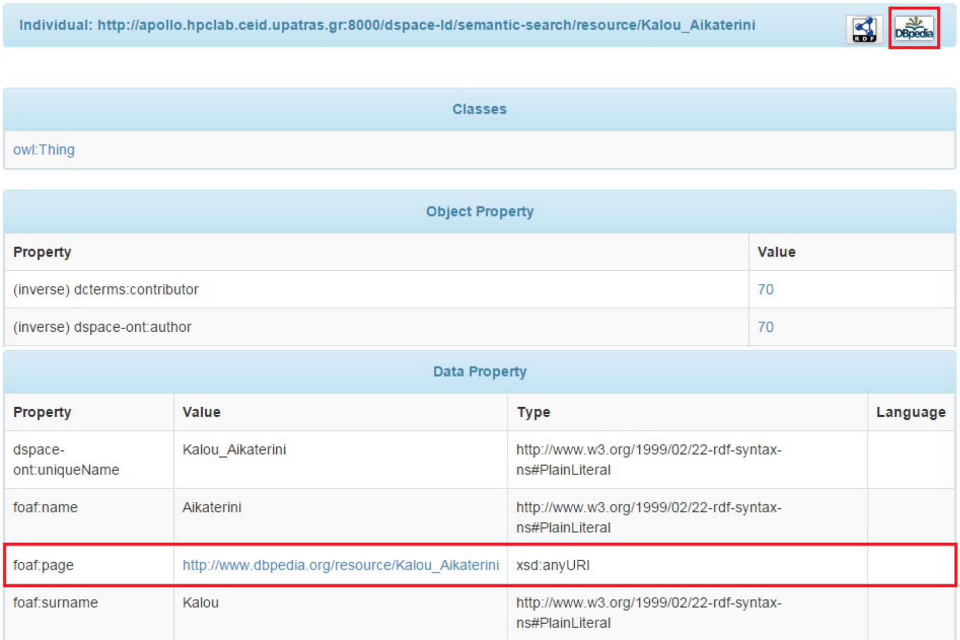


Fig. 9. The navigation pane for an entity of type `dspace-ont`:author.

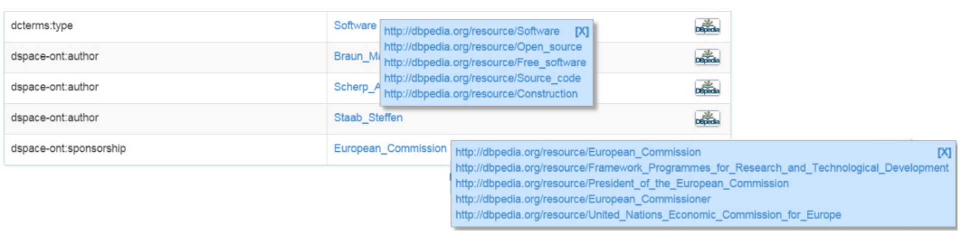


Fig. 10. Dynamic tooltips — The DBpedia Lookup service is triggered for type 'Software' and sponsor 'European_Commission'.

resolve this, a dynamic tooltip is presented to the user, including up to five matching DBpedia resource URIs (see Fig. 10), thus injecting additional URLs during runtime. Moreover, the `dcterms:subject` and `dcterms:publisher` values, albeit literals, are also linked to DBpedia in the same way.

5.3. *Facilitating access from the repository's core context*

In DSpace, the item view page works as the rendering of the items inner metadata record in human readable format. When navigating to this page, someone can either review a snippet of an item's metadata information (like title, authors, etc.) or be presented with more detailed information about its elements and corresponding

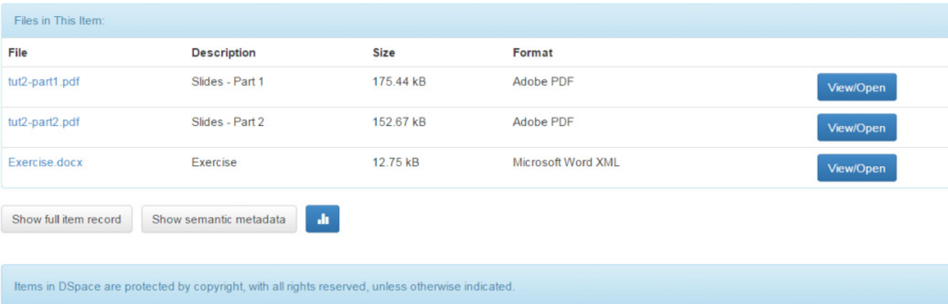


Fig. 11. The 'Show semantic metadata' button directs to the item's navigation pane.

values. Users can interchange between the item's short and full metadata record view, by selecting a button placed at the end of the page.

Following the DSpace design logic, a new button (*Show semantic metadata*) working as a direct link to the item's navigation pane, has been created and placed next to the ones already existing in the view page. By taking advantage of this newly added functionality, users acquire direct access to the item's navigation pane and hence to the repository's Linked Data facility. As shown in Fig. 11, this button is similar to the one used in order to interchange between the two different views of an item's metadata record (short or full). Consequently, access to Linked Data comes to the typical repository user as intuitively as the process of browsing among repository resources.

6. Use Cases and Dataset Metrics

The Linked Data facility, which we have enhanced our educational repository with, is targeted both to a typical DSpace user, as well as to someone who is more familiar with Semantic Web concepts. In this section, we present various usage scenarios that demonstrate how resources' ontological descriptions can be accessed from within the repository's core context, interlinked from outside datasets, link to external datasets and get discovered by semantic search. Finally, we discuss some characteristics of our test dataset and the knowledge gain that comes as a consequence of external linking.

6.1. *Browsing the item view page*

Let us consider the following scenario where a scholar visits our repository with the view to research more information about the Semantic Web, semantic query-answering, reasoning and rules. With this particular request in mind, he will probably choose to browse items within the 'RR 2014' Community. As denoted by the Community's short description, its Collections contain a selective set of abstracts coming from the articles presented at the 8th International Conference on Web Reasoning and Rule Systems.[j]

[j] http://rr2014.di.uoa.gr/

As a first step, the user needs either to search the contents of the Collection that interests him (using the *Search DSpace* textbox) or alternatively to use one of the DSpace standard browsing capabilities, like e.g. *Browse by title*. The user will be presented with a result table, arranging the retrieved articles by title and containing the contributors' names as well. When he eventually hits upon an article of interest, he can further investigate it and proceed with reviewing its metadata, just by clicking on its title and opening the item's view page. At this point, and while inspecting more thoroughly the item's metadata (provided in either their short or full form), the scholar comes upon the option to also give a look at the item's semantic information. By pressing the '*Show semantic metadata*' button, added for this purpose at the end of the item's view page, the user acquires access to the article's navigation pane where its ontological information is organized in a human readable format. What is actually presented to the user is not only mere textual information, but also a number of resolvable URIs which link to either our internal or to other external resources. This linkable information may regard not only the item itself, but also its authors, publishers or even sponsors, its subject keywords as well as the learning resource type of the submitted article.

6.2. *Exploiting the semantic search interface*

The same ontological information — result of the first scenario described above — could also be reached by performing a semantic-based search. This facility is offered as an additional capability to the users of the repository. Provided that the scholar described earlier (searching for reasoning, semantic querying and rules related information) is a bit familiar with the Semantic Web notions, they could be able to exploit our Semantic Search mechanism and perform a Manchester Syntax query, appropriately constructed so as to express their research interests. For example, to search for articles related to the topics under question which have been authored by '*Koutsomitropoulos*' the following query can be posed through the Semantic Search interface:

```
dspace-ont:author some (foaf:surname value "Koutsomitropoulos")
```

After evaluating this query, the results presented in Fig. 12 are returned. By hitting any of these items, the user is directed to the selected individual's navigation pane.

Nevertheless, because the list of items returned corresponds to articles presented at other conferences as well, the scholar could narrow its search by constructing a more refined query (taking for example into account a subject keyword of the requested article). The extent of this refinement is up to each user's familiarity with Semantic Web concepts and queries. What is important, though, is the fact that the scholar who visited our educational repository searching for '*RR 2014*' articles is offered alternative paths to acquire access to the items' inner semantic information as well as to the external interlinked resources, as depicted in Fig. 13.

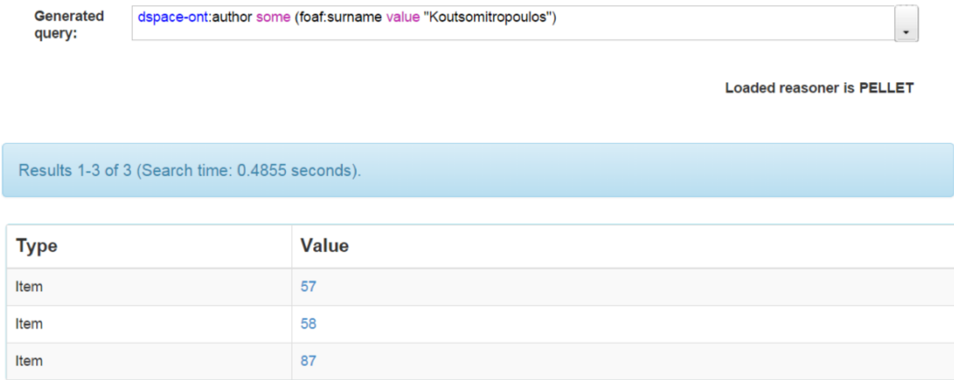


Fig. 12. Accessing semantic information using the Semantic Search interface.

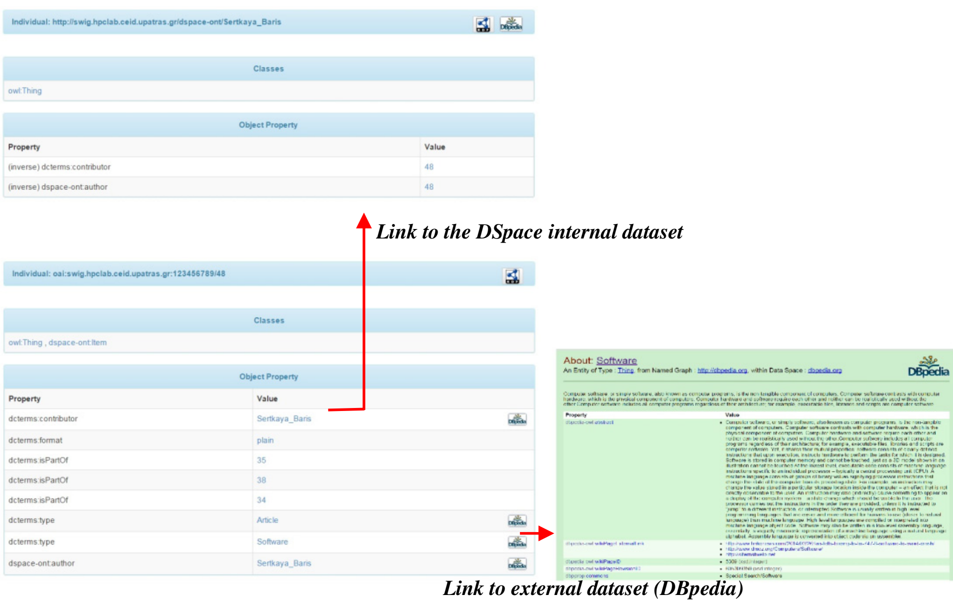


Fig. 13. The Navigation Pane as an access point to both internal and external interlinked datasets.

6.3. *Access from external resources*

In both cases described above, someone using this enhanced snapshot of the DSpace system comes across the Linked Data facility and is able to access other scholarly datasets across the Web. But this is not the only option offered by this system. Apart from those users that choose to visit and explore our repository, because they are aware of its existence, other individuals, services or mechanisms — externally maintained — can also reach and navigate among its interlinked information.

Entities being assigned resolvable URIs means that their semantic representations can now be accessed from the outside world. The URI resolution scheme implemented by the Linked Data Provider (see Sec. 5.1) ensures that application dependent entities, i.e. resources that are specific to a particular repository installation, including items and author entities, will be dereferenced within the application. Generic vocabulary terms, such as DSpace types, LOM specific fillers and Dublin Core schemes will instead maintain their own namespace and be resolved (or not) accordingly.

External access to repository data can be enabled by outside data management services, like for example datahub.io[k] where a sample DSpace dataset has already been registered.[l] Additionally, RDF-enabled browsers and data exploration services can now process the Linked Data entities maintained by the repository (see Fig. 14). This, combined with the external references to resources injected in the data would result in highly navigable Linked Data graphs (Fig. 15). Thanks to content negotiation, both human readable (HTML) as well as machine readable (RDF) representations can be requested and accommodated.

1 records (27 triples, 18 properties) match selected filters.
To enable grouping, order records by some value.

78

dcterms:extent	28855
dcterms:title	Biological Names and Taxonomies on the Semantic Web - Managing the Change in Scientific Conception
wdrs:describedby	proxy:entity/http/apollo.hpcla...b.ceid.upatras.gr:123456789/78
dcterms:identifier	http://apollo.hpclab.ceid.upat...pui42-demo/handle/123456789/78
dcterms:format	http://swig.hpclab.ceid.upatras.gr/dspace-ont/pdf
dcterms:language	en
dcterms:type	http://swig.hpclab.ceid.upatras.gr/dspace-ont/Working_Paper
dcterms:abstract	Biodiversity management requires the usage of heterogeneous biological information from multiple sources. Indexing, aggregating, and finding such information is based on names and taxonomic knowledge of organisms. However, taxonomies change in time due to new scientific findings, opinions of authorities, and changes in our conception about life forms. Furthermore, organism names and their meaning change in time, different authorities use different scientific names for the same taxon in different times, and various vernacular names are in use in different languages. This makes data integration and information retrieval difficult without detailed biological information. This paper introduces a meta-ontology for managing the names and taxonomies of organisms, and presents three applications for it: 1) publishing biological species lists as ontology services (ca. 20 taxonomies including more than 80,000 names), 2) collaborative management of the vernacular names of vascular plants (ca. 26,000 taxa), and 3) management of individual scientific name changes based on research results, covering a group of beetles. The applications are based on the databases of the Finnish Museum of Natural History and are used in a living lab environment on the web.
dcterms:contributor	http://swig.hpclab.ceid.upatras.gr/dspace-ont/Hyvönen_Eero http://swig.hpclab.ceid.upatras.gr/dspace-ont/Laurenne_Nina http://swig.hpclab.ceid.upatras.gr/dspace-ont/Tuominen_Jouni
dcterms:isPartOf	http://apollo.hpclab.ceid.upat...pui42-demo/handle/123456789/54 http://apollo.hpclab.ceid.upat...pui42-demo/handle/123456789/56 http://apollo.hpclab.ceid.upat...pui42-demo/handle/123456789/34
dcterms:issued	2011-05-12
dcterms:subject	ONKI Biological Ontologies
dcterms:available	2011-05-12T15:09:07Z
dcterms:dateAccepted	2011-05-12T15:09:07Z
dcterms:provenance	Made available in DSpace on 2011-05-12T15:09:07Z (GMT). No. of bitstreams: 1 tuominen-et-al-taxmeon-2011.pdf: 28855 bytes, checksum: fa8da2d5e1b27a747f754799d85c5e44 (MD5) Submitted by Georgia Solomou (solomou@hpclab.ceid.upatras.gr) on 2011-05-12T15:09:07Z
author	http://swig.hpclab.ceid.upatras.gr/dspace-ont/Tuominen_Jouni http://swig.hpclab.ceid.upatras.gr/dspace-ont/Hyvönen_Eero http://swig.hpclab.ceid.upatras.gr/dspace-ont/Laurenne_Nina
uniqueName	oai:apollo.hpclab.ceid.upatras.gr:123456789/78

Fig. 14. OpenLink Data Explorer view of a repository item.

[k] http://datahub.io/
[l] http://datahub.io/dataset/dspace

Fig. 15. OpenLink Data Explorer graph with item node as root. Linked entities can further be unfolded.

6.4. *About the dataset*

Our test set consists of scholarly publications and other material from Semantic Web conferences, including ESWC and RR. The following Table 5 summarizes some key metrics related to the dataset, highlighting some of its quantitative as well as qualitative aspects.

We notice that URL injection contributes a considerable 19% of additional relation triples to the knowledge base, without taking into account additional URLs injected during page view. Each reified entity is augmented on average by at least one more injected URL and each item is subtly linked to 8 external resources, through its object property relations to other entities.

Table 5. The URI resolution scheme (Linked Data Provider).

Metric	Value
Total DSpace Items	53
Total entities	320
Total entity relations (triples)	2175
OWL dialect	OWL 2 DL
Description Logic	$\mathcal{SROIF}$ (D)
Total injected relations (triples)	415
Average injected triples per item	7.8
Average injected triples per entity	1.3
Percentage of injected triples	19.1%

All added URLs are resolvable; however, a full description may not yet be available within the target dataset (e.g. DBpedia). In this case, the call to the DBpedia Lookup Service during page view lets the user disambiguate and possibly discover additional references. The underlying description logic is highly expressive, giving rise to highly refined semantic queries through the semantic search facility.

7. Conclusions and Future Work

In this paper we have presented a method for enhancing educational resource collections with Semantic Web and Linked Data features. The reference implementation provides a step for researchers and academics towards the ability to navigate among the vast amount of scholarly information that is nowadays available through the LOD cloud and hence to discover additional knowledge that exists beyond the barriers of a single repository. We have shown a process that augments repository assets by extracting entities, assigning identifiers and injecting external references, both inside the ontological model as well as during runtime.

To come across the Linked Data facility is as simple as clicking a button, which is placed under each item's view page and is easily accessible by any typical DSpace user. Moreover, care has been taken to accommodate both human users and external services, by implementing a URI resolution logic for entity dereferencing. Alternatively, the Semantic Search interface can be exploited and eventually give access to the same type of interlinked information by performing appropriately constructed queries.

The latter option, although it may appeal to more Semantic Web-expert users, remains a challenging facility for intelligently searching DSpace. Besides, what it finally retrieves is not mere repository items but a number of resolvable entities, sometimes enhanced with additional information, inferred by the underlying reasoning mechanism. What is more, the Semantic Search interface has been designed and enhanced with a number of features, which aim at overcoming possible deficiencies in Manchester syntax fluency and thus smoothly guide users in the construction of a semantic query.

The combination of a reasoning-based knowledge acquisition mechanism with a Linked Data service can help educational institutions to provide new discovery capabilities for their content and to be part of the greater LOD cloud effortlessly. DBpedia is naturally a nodal point of the latter, but interconnecting with other data sources would also be useful, like for example DBLP. What is more, data from these sources can be brought back into our model, so that we could reason with them and reveal a whole new set of correlations between repository assets and the outside world.

References

[1] M. Andro, E. Asselin and M. Maisonneuve, Digital libraries: Comparison of 10 software, *Library Collections, Acquisitions, and Technical Services* **36**(3/4) (2012) 79–83.

[2] M. d'Aquin, Linked Data for Open and Distance Learning, Report for the Commonwealth of Learning (2014).

[3] M. d'Aquin, H. Drachsler, E. Herder, S. Dietze, M. Guy and E. Parodi, Building the open elements of an open data competition, *D-Lib Magazine* **20**(5/6) (2014).

[4] P.-N. Becker, Bringing DSpace into the Semantic Web, online article, https://wiki.duraspace.org/display/~pbecker/Bringing+DSpace+into+the+Semantic+Web (2014).

[5] C. Bizer, T. Heath and T. Berners-Lee, Linked Data — The story so far, special issue on Linked Data, *International Journal on Semantic Web and Information Systems*, 2009.

[6] DCMI Usage Board, DCMI Metadata Terms, DCMI Recommendation 2008, http://dublincore.org/documents/2008/01/14/dcmi-terms/.

[7] S. Dietze, S. Sanchez-Alonso, H. Ebner, H. Q. Yu, D. Giordano, I. Marenzi and B. Pereira Nunes, Interlinking educational resources and the web of data, *Program* **47**(1) (2013) 60–91.

[8] S. Dietze, H. Q. Yu, D. Giordano, E. Kaldoudi, N. Dovrolis and D. Taibi, Linked education: Interlinking educational resources and the web of data, in *27th ACM Symposium on Applied Computing*, Special Track on Semantic Web and Applications, Trento, Italy, 2012.

[9] B. Glimm, I. Horrocks, B. Motik, G. Stoilos and Z. Wang and T. Hermi, An OWL 2 Reasoner, *Journal of Automated Reasoning* **53**(3) (2014) 245–269.

[10] T. Heath and C. Bizer, Linked data: Evolving the Web into a Global Data Space (1st ed.), *Synthesis Lectures on the Semantic Web: Theory and Technology* **1**(1) (2011) 1–136.

[11] J. Hendler, Why the semantic web will never work, in *Keynote at the 8th Extended Semantic Web Conference*, Heraklion, Greece, 2011.

[12] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider and S. Rudolph (eds.), *OWL 2 Web Ontology Language Primer* (2nd ed.), W3C Recommendation (2012).

[13] M. Horridge and S. Bechhofer, The OWL API: A Java API for OWL Ontologies Semantic Web Journal 2:1, Special Issue on Semantic Web Tools and Systems, (2011) 11–21.

[14] M. Horridge and P. S. Patel-Schneider, Manchester Syntax for OWL 1.1, in *4th OWL Experiences and Directions Workshop*, Gaithersburg, Maryland, 2008.

[15] IEEE LTSC, Draft Standard for Learning Object Metadata, IEEE 1484.12.1-2002 (2002), http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf.

[16] IEEE LTSC, Draft Recommended Practice for Expressing IEEE Learning Object Metadata Instances Using the Dublin Core Abstract Model, IEEE P1484.12.4/D1 (2008) http://dublincore.org/educationwiki/DCMIIEEELTSCTaskforce?action=AttachFile&do=get&target=LOM-DCAM-newdraft.pdf.

[17] C. Lagoze, H. Van de Sompel, M. Nelson and S. Warner, The Open Archive Initiative Protocol for Metadata Harvesting (2002), http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm.

[18] D. A. Koutsomitropoulos, R. Borillo Domenech and G. D. Solomou, A structured semantic query interface for reasoning-based search and retrieval, in *8th Extended Semantic Web Conference*, Part I, 17–31, 2011.

[19] N. Manouselis, K. Kastrantas, S. Sanchez-Alonso, J. Cáceres and H. Ebner, Architecture of the Organic.Edunet Web Portal, in A. Tatnall (ed.), *Web Technologies: Concepts, Methodologies, Tools, and Applications* (2012), pp. 759–773.

[20] E. Mitsopoulou, D. Taibi, D. Giordano, S. Dietze, H. Q. Yu, P. Bamidis, C. Bratsas and L. Woodham, Connecting medical educational resources to the Linked Data cloud: The mEducator RDF Schema, store and API, in *Linked Learning 2011: 1st International Workshop on eLearning Approaches for the Linked Data Age, 8th Extended Semantic Web Conference*, Heraklion, Greece, 2011.

[21] B. Motik, B. Parsia and P. F. Patel-Schneider (eds.), *OWL 2 Web Ontology Language XML Serialization* (2nd ed.), W3C Recommendation, 2012.

[22] E. Sirin, B. Parsia, B. Cuenca Grau, A. Kalyanpur and Y. Katz, Pellet: A practical OWL-DL reasoner, *Web Semantics: Science, Services and Agents on the World Wide Web* **5**(2) (2007) 51–53.

[23] D. Tsarkov and I. Horrocks, FaCT++ description logic reasoner: System description, in *Third International Joint Conference*, LNCS, Vol. 4130, 2006, pp. 292–297.