



INTERNATIONAL  
HELLENIC  
UNIVERSITY

# Extending IHU's Web pages with semantically- rich metadata

**Michael-Diamandis Hickson**

SID: 3301120002

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

*Master of Science (MSc) in Information and Communication Systems*

OCTOBER 2013  
THESSALONIKI – GREECE



INTERNATIONAL  
HELLENIC  
UNIVERSITY

# Extending IHU's Web pages with semantically- rich metadata

**Michael-Diamandis Hickson**

SID: 3301120002

Supervisor:

Prof. Nick Bassiliades

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

*Master of Science (MSc) in Information and Communication Systems*

OCTOBER 2013  
THESSALONIKI – GREECE

# Abstract

This dissertation was written as a part of the MSc in ICT Systems at the International Hellenic University.

Semantic web and linked open data (LOD) is an area of great research interest nowadays and it is rapidly expanding. There is already a great amount of datasets published over the LOD, interlinked with each other, concerning various domains of interest. The purpose of this dissertation is to enrich International Hellenic University's (IHU's) web site with semantically-rich metadata, following the LOD principles. This metadata will be openly available for anyone to explore and/or consume, either through a user friendly web interface or via a SPARQL-endpoint. Moreover, a test case is presented of how metadata snippets, could be embedded into IHU's webpages through the Content Management System (CMS) it runs on.

At this stage, I would like to acknowledge the help and successful collaboration between my supervisor Dr. Nick Bassiliades and myself. I would also like to thank Anastasia Varitimou, whose work [1] and guidance during the development of the application were really crucial.

Michael-Diamandis Hickson

6/11/2013

# Contents

<b>ABSTRACT.....</b>	<b>III</b>
<b>CONTENTS.....</b>	<b>IV</b>
<b>1 INTRODUCTION.....</b>	<b>6</b>
1.1 THE PROBLEM.....	7
1.2 THE IMPACT.....	8
1.3 SEMANTIC WEB: THE SOLUTION.....	8
1.4 APPLICATION CONTEXT.....	11
1.5 OUTLINE.....	13
<b>2 LITERATURE REVIEW.....</b>	<b>13</b>
2.1 THE WEB OF LINKED DATA.....	13
2.2 LINKED DATA DESIGN ISSUES.....	15
2.3 LINKED DATA IN THE EDUCATIONAL SECTOR.....	20
2.4 SIMILAR APPLICATIONS.....	28
2.4.1 <i>OU Linked Open Data</i> .....	28
2.4.2 <i>Linked Open Data University of Münster</i> .....	31
2.4.3 <i>Tsinghua University Open Data</i> .....	33
2.4.4 <i>The Data.dcs Project</i> .....	34
2.4.5 <i>The LOD.CS.UNIPA Project</i> .....	35
2.4.6 <i>The AcademIS Ontology</i> .....	36
2.4.7 <i>DBpedia</i> .....	39
2.5 CONCLUSIONS.....	42
<b>3 PROBLEM DEFINITION.....</b>	<b>43</b>
<b>4 IMPLEMENTATION.....</b>	<b>46</b>
4.1 DATA CATEGORIES AND SCHEMA.....	47
4.1.1 <i>The Datasets</i> .....	47
4.1.2 <i>Vocabularies and Extensions</i> .....	48
4.1.3 <i>The Academic Ontology's Instance Level</i> .....	66
4.2 SYSTEM INFRASTRUCTURE AND USER INTERFACES.....	74
4.2.1 <i>The Workflow for Generating Linked Data</i> .....	75
4.2.2 <i>Accessing IHU's Linked Data</i> .....	78

<b>5 CONCLUSIONS.....</b>	<b>83</b>
<b>6 FUTURE WORK.....</b>	<b>84</b>
<b>BIBLIOGRAPHY.....</b>	<b>86</b>
<b>APPENDIX.....</b>	<b>88</b>

# 1 Introduction

The World Wide Web (WWW) as we know it today contains a vast amount of data that is presented to end-users through documents constructed by markup languages, such as the HyperText Markup Language (HTML). These documents are connected to each other through hyperlinks, as shown in Figure 1, creating today's web of interconnected documents.

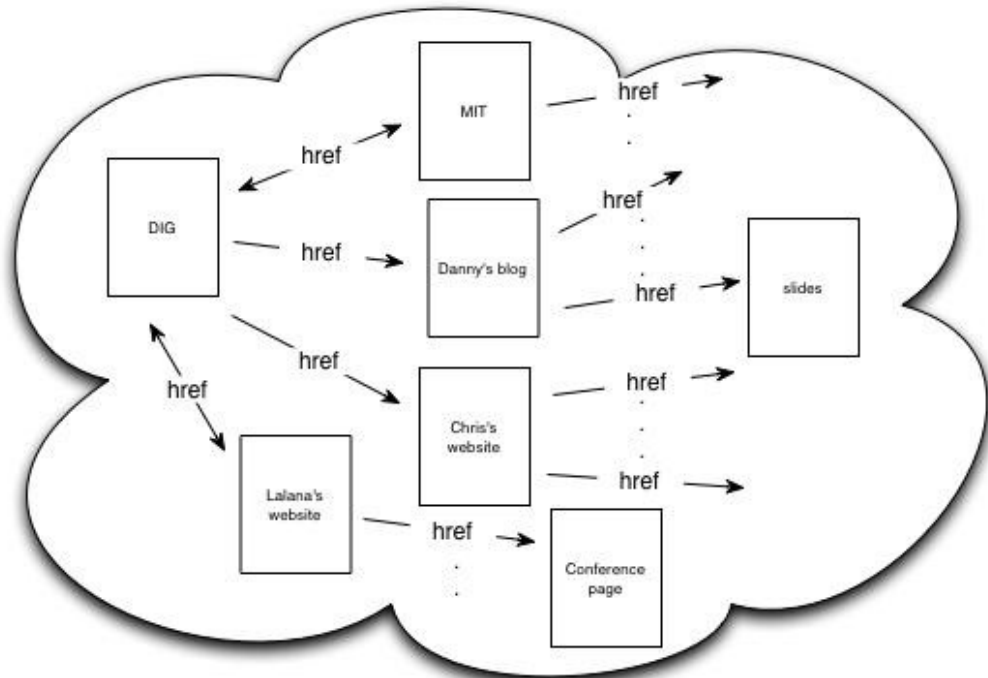


Figure 1: Web 1.0

This architecture provides a good and efficient way of presenting the data in a human readable format. Search engines also appear to have no problem crawling over these documents and indexing their data using textual analysis algorithms.

## 1.1 The problem

Although this appears to be sufficient there is a major problem with the architecture of today's World Wide Web. The data that lies on the web today is not so easily accessible to machines after all, since as we mentioned above they can only understand it and extract knowledge from it by using sophisticated textual analysis algorithms. This means that there are no semantics associated with the data, so that machines, such as search engines, can read the data and actually understand its meaning. However, there have been efforts to bypass this barrier by using other more machine friendly ways of serving data on the web, through the use of markup languages such as the Extensible Markup Language (XML). XML was designed in order to describe data, provide meaning to it, so that machines can understand it, and in general to separate content from display. Although this provided some data interoperability between different machines it did not manage to solve the problem, since it actually didn't provide any formal semantics, so that information from various documents could be conceptually merged. In order for the machines to understand the meaning of the data they parsed, they would have to follow a convention of a common vocabulary, but even if that was the case, the vocabulary concepts did not carry any semantics. In spite of the semantic interoperability matter though, another concern is the way the data are linked with each other. Nowadays data are embedded into semantically marked up documents which are then linked to each other. This means that there is no actual interconnection between each data concept lying in each individual document, but only between the documents that carry them. This also makes it difficult for machines and even humans to discover relevant information, as they have to traverse each individual document to find what they are looking for. Last but not least, most of today's data lies hidden and isolated behind distributed databases that can only be accessed through individual interfaces, limiting the interoperability and accessibility of this data. This information that is buried down on these databases, constructs today's "hidden web" as it cannot be discovered and indexed by search engines. There is a need to describe all of this data in a commonly agreed data representation model, held in a global database, and to have it semantically interconnected to each other so that anyone or anything can have access to it through an individual interface.

## 1.2 The impact

These problems have a great impact on the evolution of the World Wide Web as it limits its capabilities in numerous ways. The most important amongst them is that it does not provide to machines, such as search engines, the means to really understand the semantics of the information they crawl over. This really limits the relevance of the results they return over queries users make. It also does not allow relevant information to be connected to each other, so that users and machines can explore and navigate over a common interconnected database, which holds all of the information in a commonly agreed data representation model. Also an openly accessible global database, would mean that various applications could be developed that could navigate over this information, extract knowledge from it and provide useful services to the end users.

## 1.3 Semantic web: The solution

In order to overcome these problems the Semantic Web, also known as “The Web of Data” came to replace today’s web, also known as the “Web of Documents”. The concept behind the Semantic Web is that data should be represented in a commonly agreed structure and be openly available, through a common interface, so that it can be shared and reused across applications and enterprises. Furthermore this data, which may come from different sources, should be connected to each other with links to allow discovery of related information. This concept is shown in Figure 2.

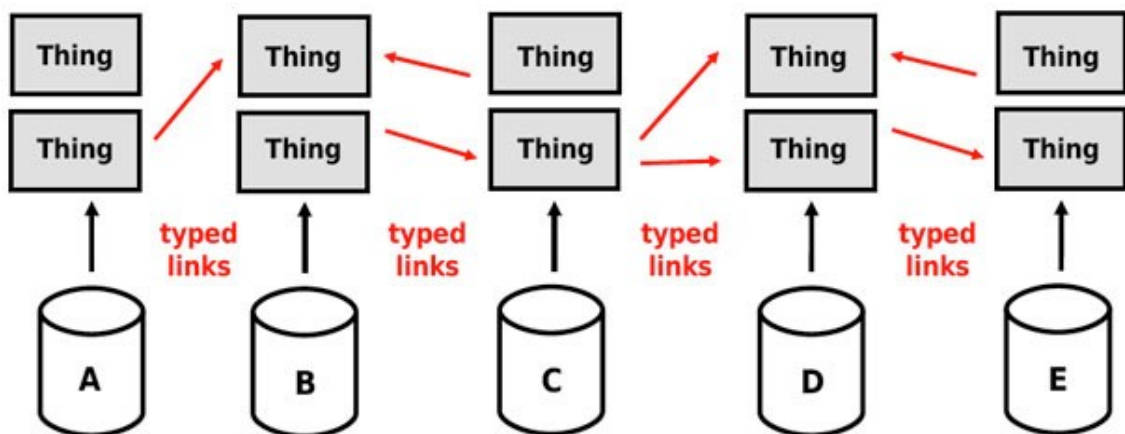


Figure 2

Another main concept of the Semantic web is that this data should not just carry plain information, but it should have meaning. In order to achieve this several layers of



representational structures are needed. The basic layers, which are also presented in Figure 3, are the following:

- the XML layer, which represents the structure of data
- the RDF layer, which represents the meaning of data
- the Ontology layer, which represents the formal common agreement about meaning of data
- the Logic layer, which enables intelligent reasoning with meaningful data

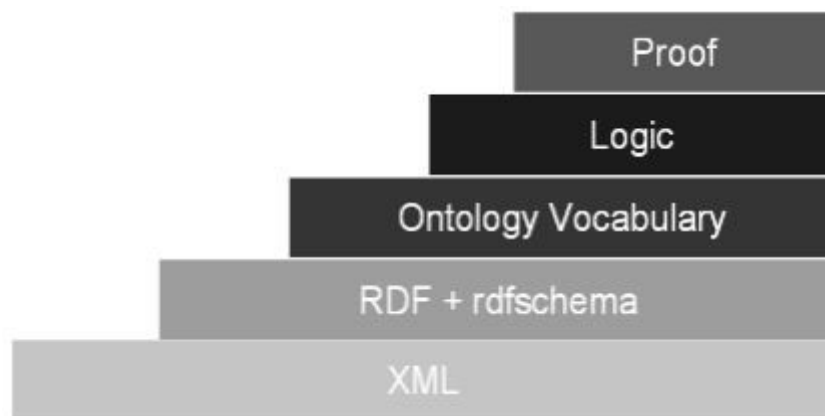


Figure 3: Basic layers of the Semantic Web architecture

It is worth to note that the real power of the Semantic Web will be realized when people create many systems that collect web content from diverse sources, process the information and exchange the results with other human or machine agents. Thereby, the effectiveness of the Semantic Web will increase drastically as more machine-readable web content and automated services become available. This level of inter-agent communication will require the exchange of "proofs".

The Semantic Web movement has led to the emergence of the Linked Open Data (LOD) cloud. The LOD is the actual global database that is day by day growing, by initiatives made from various individuals and organizations to publish [2] structured data on the web and make it available to everyone in order to explore and reuse it. This data is represented by using technologies such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL) and is interconnected with each other by using Uniform resource identifiers (URIs). Also the published data should follow the LOD principles [3] which are the following:

1. Use URIs to identify things.

2. Use HTTP URIs so that these things can be referred to and looked up ("dereference") by people and user agents.
3. Provide useful information about the "thing" when its URI is dereferenced, using standard formats such as RDF/XML.
4. Include links to other, related URIs in the exposed data to improve discovery of other related information on the Web.

These principles provide a framework for publishing data on the Web that is being followed by many data providers. The principles entail to share a common data format based on URIs and RDF, as well as to use SPARQL as a common language to manipulate the data. In addition, the fourth principle encourages data providers to federate their datasets to others in the Web of Data by explicitly stating the relationships between the data they publish and the data already published by third-parties.

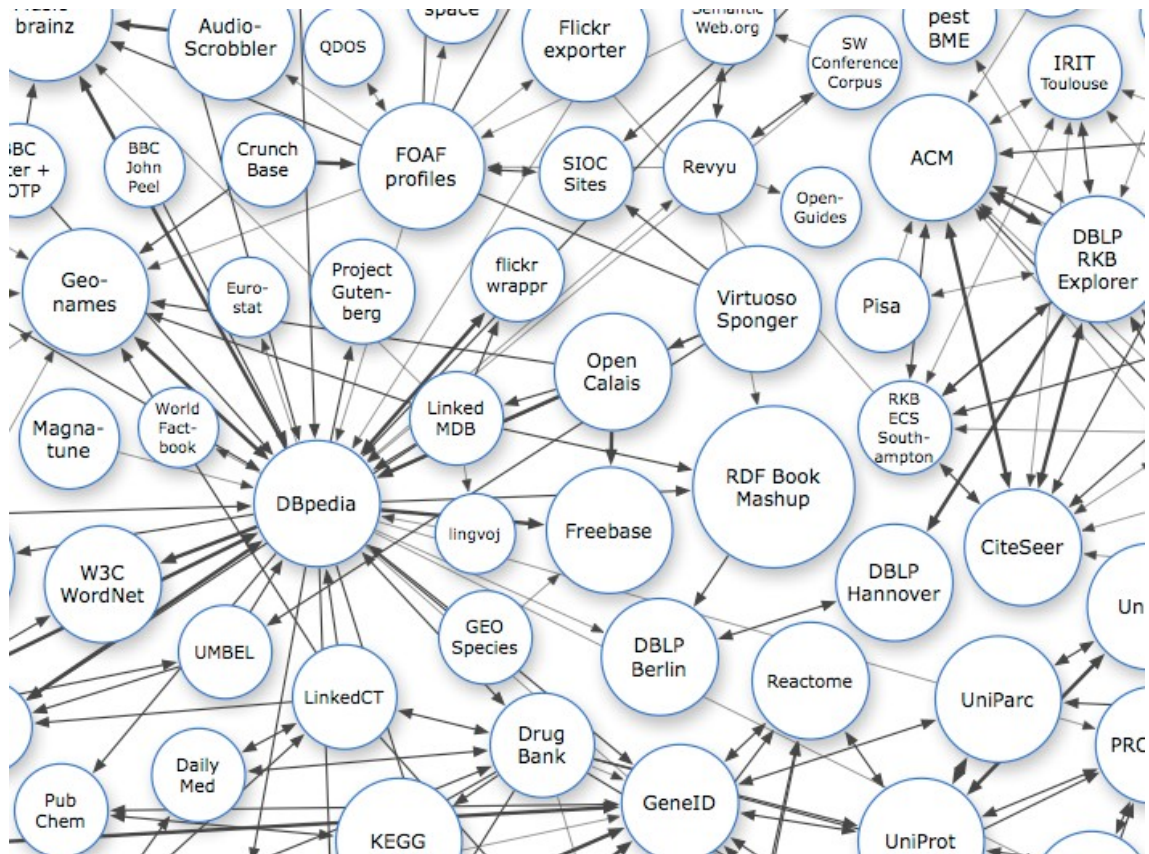


Figure 4: The linked open data cloud

In Figure 4 we can see part of today's LOD cloud, and the various datasets that have been currently published. Each dataset represents a certain domain of knowledge, while the size of each circle depicted in the figure shows the amount of information (triples) it holds. We can also see that information from various datasets is connected with relevant

information from other datasets, thereby creating this huge interconnected graph. These connections refer to the fourth LOD principle. It is said that a picture is worth a thousand words, and in this case this is absolutely correct, since through this image we can see the true meaning of this movement that has started towards the creation of the new web, the “Web of Data”.

## **1.4 Application Context**

The context of this dissertation is to make use of all the modern semantic web technologies in order to develop an application over the LOD cloud. The purpose of this application is to publish and make available over the LOD cloud information that concerns various aspects of the university such as offered courses, people working in it, its structural organization etc. and the relationships with each other. Since this information will be contained in a separate subdomain, a workflow of how this information can directly be embedded into IHU’s web pages is also going to be presented. Although these two approaches serve the same purpose, which is to publish the university’s data, and are based on the same components, they are implemented in different ways.

At this point, we should state that mainly information concerning the university’s School of Science and Technology was taken into consideration in order to produce a test case. Our first objective was to create a semantically enhanced dataset, which includes information extracted from the university’s website. In the conceptual level this dataset was described with the use of already published and widely used ontologies, while the representation languages that were used to achieve the description of these concepts were the Web Ontology Language, also known as OWL, and the Resource Description Framework Schema (RDFS). After the selection of the appropriate ontologies, where each one of them efficiently described parts of the domain of our concern, they were combined in order to coherently describe the main part of the domain. In cases where the descriptions were inefficient, extensions and new concepts were introduced in order to cover the whole domain of interest. This led to the creation of the Academic Ontology (see Appendix), which is also openly available on the web for third parties to use/reuse in order to describe concepts concerning the Academic community. In the instance level the Resource Description Framework (RDF) was used in order to describe our data in an appropriate format. After the data was constructed we

were able to publish it on the LOD cloud following the LOD principles. In order for the data to be openly available and accessible for consumption, a SPARQL endpoint interface was also provided. Also a more user friendly web interface is provided, so that the published information can more easily be explored and read by humans. All of the above mentioned, are publicly available as they are currently running on public domains, which can be accessed through a web browser. Finally, part of the data that is published in RDF format was converted into RDFa [4] format in order to provide a showcase of how someone could embed these semantically enhanced snippets of information into IHU's web pages through the CMS it is runs on. In Figure 5 we can see in a layered approach the different components that take part in our application. For terms of simplicity, we do not include the whole picture of the application's functionality, as it will be discussed in more detail later on.

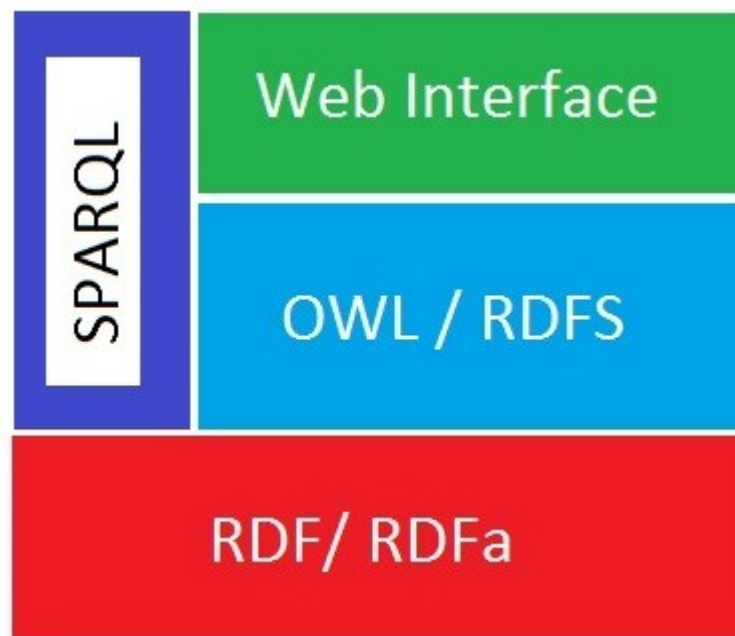


Figure 5: Application layers

In a nutshell, the purpose of the dissertation is twofold. Our first goal is to publish IHU's data on the LOD cloud, following the LOD principles, so that it is made available for anyone to explore and consume, either directly from the web pages of the university or via a SPARQL-endpoint. Our second goal is to provide a brief guide of how someone could easily enrich IHU's web pages with semantically-rich metadata, so that search engines can discover and classify the relevant information more accurately. This will provide a more open and machine readable version of the website's data, leading to all the benefits that were mentioned in the previous subchapters.

## 1.5 Outline

The rest of this dissertation is organized as follows. Chapter 2 contains the related work that has been done in the area of semantic web in general, while it also presents similar applications, to the currently discussed, that have been developed by other educational institutions. The problem is defined in chapter 3. In chapter 4 the context and implementation of the application is explained in detail. Chapter 5 describes the conclusions that we have drawn from this work and chapter 6 identifies issues for further research and future work that can be done in order to extend the functionality of the presented application.

# 2 Literature Review

In general, with the rapid development of the Semantic Web, there has been an increased interest from various organizations to publish their data on the web, as Linked Open Data, in order to benefit from the advantages this new technology offers. In this chapter we provide a review of the related work that has been done in the literature. There are several related works on this field but we are mainly going to focus on the ones related to the educational sector, since this topic relates to the context of our application. Below we list and comment on the most related works. On the first section we are going to provide a brief review of the literature concerning the Semantic Web and the LOD in general, while in the rest sections we will present the most significant and more relevant to our application.

## 2.1 The Web of Linked Data

[5] presents in a very descriptive way the web of linked data. It presents the benefits of the transition from Web 1.0 to Web 3.0, also known as the Semantic Web. It also provides a guide that explains the technologies and the LOD principles that have to be followed in order for someone to publish data on the LOD cloud. In a few words, the main concepts someone should have in mind when publishing his data is that it should firstly use standard technologies such as RDF/RDFs and OWL to form his data and that this data should also include RDF links to other information, so that people can navigate

the Web of Data as a whole and discover other relevant information. On top of that HTTP URIs should be used as names for the information someone wants to present, so that people can easily look it up. It is also emphasized that it is a good practice to form the data by using already published and well known ontologies, in order to make it easier for the LOD applications to parse it. Another important point is the use of properties, such as the owl:SameAs property, in order to link similar concepts that lie in different datasets. This not only benefits client side applications but the LOD movement in general, since we would like to present information about things as whole entities and not as entities that are fragmented around the cloud. These are the general guidelines we are also going to follow in order to develop a valid dataset that can be published on the LOD cloud, so that it can be available for reuse by semantic applications. RDFa is also briefly presented, as a Linked Data serialization format that can be used so that RDF triples can be embedded into HTML pages. We will use this serialization in order to enhance similar to IHU's HTML web pages with semantic information. Finally, some of the research challenges concerning the Semantic Web are discussed among of which we are going to present the most significant. One major concern is schema mapping and data fusion. So far, in order to map concepts described with the use of OWL and RDF Schema, that are similar but are defined in different ontologies, data publishers had to make use of properties, such as owl:equivalentClass, owl:equivalentProperty, rdfs:subClassOf, rdfs:subPropertyOf. This is actually the methodology we are using in order to map similar classes from different ontologies. But this is not always the case since many problems, such as structural heterogeneity and value transformations, may occur, and this is what makes schema mapping an open research challenge. Data fusion refers to the problem we mentioned before, where we have situations where an object may have different values from different sources for the same property. This is a problem, since the whole point of the semantic web is to present objects with a single and consistent representation and not as a fragmented entity. Other important issues that are discussed are link maintenance, since new interlinked information is continuously added and removed from the cloud, and user browsing experience, since the way this information is going to be well presented to the user has not yet been realized.

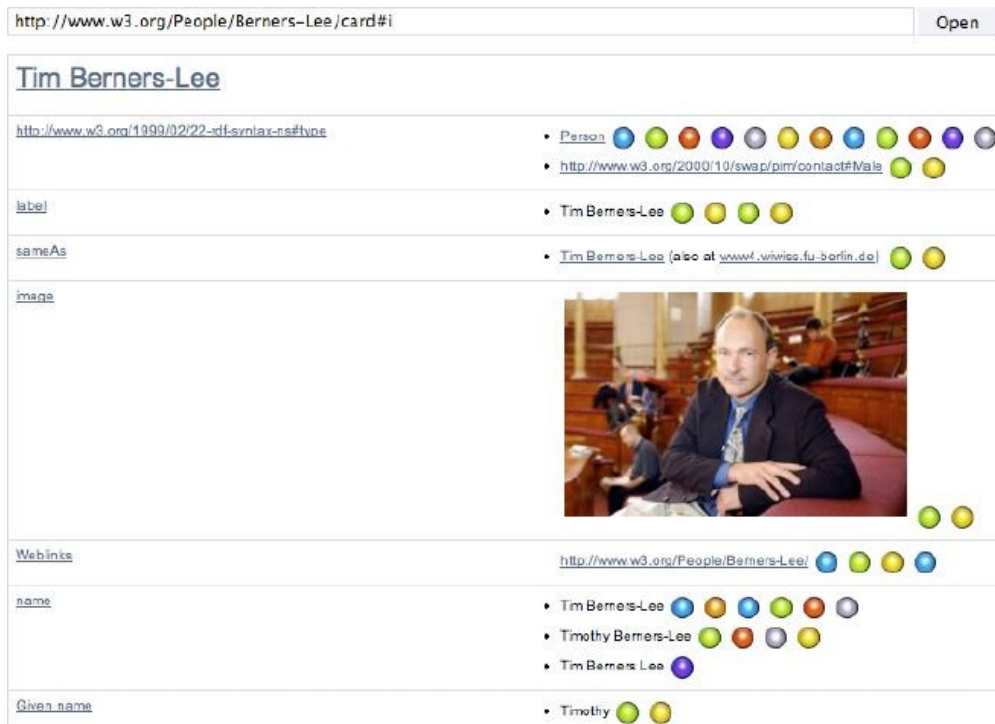


Figure 6: Linked Data Browsers

For instance, Figure 6 shows us the way today's linked data browsers usually present information. Here we can see how a linked data browser presents information in RDF format, information about the inventor of the internet-Tim Berners Lee- but only from a certain dataset, while similar instances from other datasets are stated with the owl:SameAs property. Their drawback is that they provide scattered information about objects depending on the dataset they are currently navigating through, while they should provide a more entity-centric view of the information they present and be able to keep it up to date.

## 2.2 Linked Data Design Issues

[6] explains the importance of building linked data not just for machines, but also for humans. The construction of the *riese*<sup>1</sup> dataset, which contains statistical data about the European Union, is taken as an example to demonstrate how a human and machine friendly dataset could be constructed. In order to achieve this they used RDFa as their main deployment tool, since this methodology provides both a machine and a human readable format of their information. They also deployed it as a user friendly semantic web application with a rich interface, since they support that this should be done in

<sup>1</sup> <http://riese.joanneum.at>

every dataset in order to be considered human friendly. On top of that they provided a feature called “User Contributed Interlinking”, in order to create interlinks with external datasets. This feature is mainly based on the community effect, which is also used in the successful creation of the DBpedia dataset, where it is believed that the human community will successfully take care of the creation and maintenance of interlinks. They used this approach because they strongly support that humans are good in making associations and therefore the quality of links created by humans would outreach the quality of the ones created by automated algorithms. Figure 7 below shows the final outcome of a human friendly view of the dataset, while Figure 8 shows the machine friendly view of the same dataset. In Figure 7 we can see the RDFa information presented in a rich user interface enabling the user to easily navigate through the various levels of the information hierarchy. On the other hand Figure 8 shows the plain RDFa triples that are fed to the machine parser, which is not really interested in its presentation, but only in the information that it holds. This is achieved through the combinational use of RDFa+XHTML as we mentioned before, where RDFa information is embedded into XHTML pages for proper presentation to the end user.

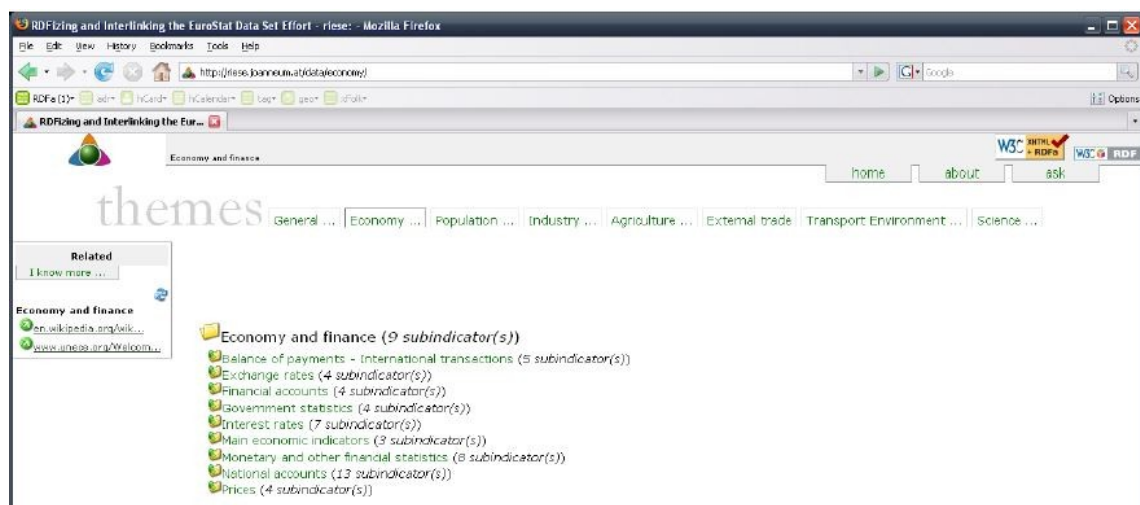


Figure 7: Human friendly view of RDFa



```

1 <body
2   about="http://riese.joanneum.at/
3     data/economy"
4   instanceof="riese:Dataset">
5   ...
6 <div id="main-ind">
7   ...
8   <a href="http://riese.joanneum.at/
9     data/bop"
10    rel="skos:narrower">
11     Balance of payments -
12     International transactions
13   </a>
14 </div>

```

Figure 8: Machine friendly view of RDFa

Since we agree with the opinions stated above, we take these facts into consideration when we construct our own semantic application. In our case though we do not only provide a static RDFa+XHTML view of the data, but we also provide a separate user friendly interface, fed by information that is dynamically embedded in our web pages through run time queries made through the provided SPARQL<sup>2</sup> endpoint.

The ontology we built in order to describe the various concepts concerning a university was constructed by combining and integrating already published and widely used ontologies, since ontology reuse is strongly encouraged by the LOD movement. The integration process that we followed is very similar to the one presented in [7]. More specifically [7] presents the ontology integration process and the problems that a publisher may face during this process, while it also proposes an integration methodology that should be followed in order to avoid this kind of implications. We are mainly going to focus on the integration process rather than the proposed methodology, since we would like to point out the integration process and the various problems that may occur rather than providing a solution for them. Integration process is defined as the process during which a resulting ontology arises from the integration of one or more ontologies, which are actually the ontologies that are being reused in order to create a more fitting to our domain ontology. As in our case, the integrated concepts are either used as they are or are modified and adapted to the domain's specific needs, resulting usually in a more specific ontology. The advantages of reusing ontologies are firstly that we do not reinvent the wheel, since we use knowledge that that has already been defined, and secondly that usually this knowledge reflects a consensus among research communities. After the ontologies that are to be integrated have been chosen though, a

<sup>2</sup> <http://www.w3.org/TR/rdf-sparql-query/>

number of problems such as consistency of the resulting ontology and level of detail throughout the whole ontology, need to be dealt with. In order to minimize this kind of problems, ontologies that are going to be reused should be evaluated by domain experts and assessed by ontologists, even from the conceptualization phase, in order to know and understand their faults and strong points. After the adequate ontologies are chosen, the various integration operations have to be applied. These operations, which can be viewed as composing, combining or assembling operations, specify how the knowledge resulting from the integrated ontologies is going to correctly be combined, so that the resulting ontology is consistent and makes sense. The last step of the integration process is that the resulting ontology should be analyzed and evaluated. To sum up, integrating ontologies in order to construct a unified and consistent resulting ontology is not trivial, since many factors should be taken into consideration and many problems may arise during the integration process. In many cases, if the integration operations that have to be applied, in order to have the expected outcome, are too many, then the publisher should consider constructing a new ontology from scratch, since this would mean that the already provided knowledge does not reflect his needs.

Similarly to [7] but from a more cost-benefit estimation point of view, [8] analyzes the challenges related to the reuse process on the basis of two scenarios in the domains of e-Recruitment and medicine, which aim at building domain ontologies by reusing existing knowledge sources. More specifically, by using these two scenarios as test cases they try to enumerate the most significant cost drivers and benefits arising from ontology reuse. They state that in many cases the costs of ontology reuse may outweigh the benefits, since reusing ontologies is not a trivial task and the appropriate technologies have not yet been developed to support this process. One of the major costs is finding, getting familiar with, adapting and updating the necessary modules in the new ontology. Another important cost is the task of trying to translate the various modules that are scattered in a number of ontologies, from the representation language they currently reside to the specific language of the application. On top of that, schema mapping is another challenge, since each module of information existing in different ontologies may represent the same concept but with a different schema representation, meaning that a mapping between these concepts needs to be made in order to state that these two modules describe the same information. The need of appropriate tools to be developed is also emphasized, in order to automate these tedious tasks, which increase the costs of ontology reuse and decrease profitability. For example, a tool that would

perform automated schema mapping or language translation would vastly decrease the costs of ontology reuse. As [7] also mentions, the benefits of ontology reuse are really important and critical since application interoperability is increased, both on the syntactic and on the semantic level. This means that both humans and machines have a common understanding of the concepts described in the ontologies and thus define and use domain concepts in the same way. Concerning the two case studies that were analyzed, the one about the human resources ontology creation showed that ontology reuse was very time saving and the benefits by far outweighed the costs. They calculated that about 15% of the total engineering time was spent on gathering the relevant sources, about 35% was spent on customizing the selected source ontologies and 40% was spent to translate these sources to the target representation language (which was OWL). They state that the customization and integration of the selected sources was quite difficult due to the fact of the wide range of classifications available. Also the translation of these sources from their initial representation language to OWL was a tedious task, because it had to be manually done, while if an automated tool existed it would have saved them a lot of time. On the contrast, in the case of the medicine ontology the benefits of ontology reuse were outweighed by their costs, because of the difficulties related to the evaluation and technical management of large scale ontologies (40% of the overall engineering time) and because of the costs of the subsequent refinement phase (45% of the overall engineering time). The complexity of the ontologies describing various concepts of the medicine domain, due to their often ambiguous modeling decisions, and the lack of a common representation format, such as OWL, which supports sharing and reuse made the task of ontology reuse really tedious. Overall, we can conclude that although ontology reuse is highly recommended the appropriate methodologies and technologies that are going to make this task easier are far from being realized. This leaves the decision of reusing ontologies or creating a new one from scratch to each individual, depending on the cost estimation calculations they make in order for the final outcome to be profitable. In our case, the benefits of reusing already published ontologies far outweighed the costs, since a large part of the concepts that were required to describe the university data was already available. But the creation of the final ontology was not trivial for us either, since an integration of the various participating ontologies into our own ontology had to be made.

## 2.3 Linked Data in the Educational Sector

Much research work in the area of education and learning in general has focused on creating standards to facilitate the exchange and reuse of educational resources in different contexts. While there are still many issues in relation to the adoption of such standards, linked data and semantic web technologies have made their goal look more achievable. Accordingly, several institutions and initiatives have started adopting these approaches to expose open data concerning the educational domain. Among them are universities, schools and research centers<sup>3</sup>, government agencies<sup>4</sup>, and projects around specific domains<sup>5</sup>.

As mentioned before our ontology was constructed by reusing already published and well-known ontologies, in order to follow a common conceptual model with the rest of the educational linked data network. [9] gives an overview of the educational linked data landscape by presenting a preliminary study of available web datasets relevant to the educational domain, investigating how well they connect, not necessarily through explicit links between them, but through the way they reuse common vocabularies, contain resources of similar types and therefore could be jointly used to form a globally addressable network of resources for education. By applying a network analysis on the educational relevant datasets they present how the participating datasets reuse various vocabularies. They found out that 144 different vocabularies were used in the education related datasets, but the interesting thing is that the most popular vocabularies were not specific to education, but were used to represent general concepts and relations, such as resource metadata (Dublin Core), people (FOAF), topics (SKOS), organizations (OpenOrg), time (W3C Time Ontology), etc. However, quite a few educational relevant vocabularies were identified, among them being the Bibliographic Ontology (BIBO), the Academic Institution Internal Structure Ontology (AIISO) and the Model of Learning Opportunities (MLO).

---

<sup>3</sup> <http://www.linkeduniversities.org>

<sup>4</sup> <http://www.education.data.gov.uk>

<sup>5</sup> <http://www.meducator.net>

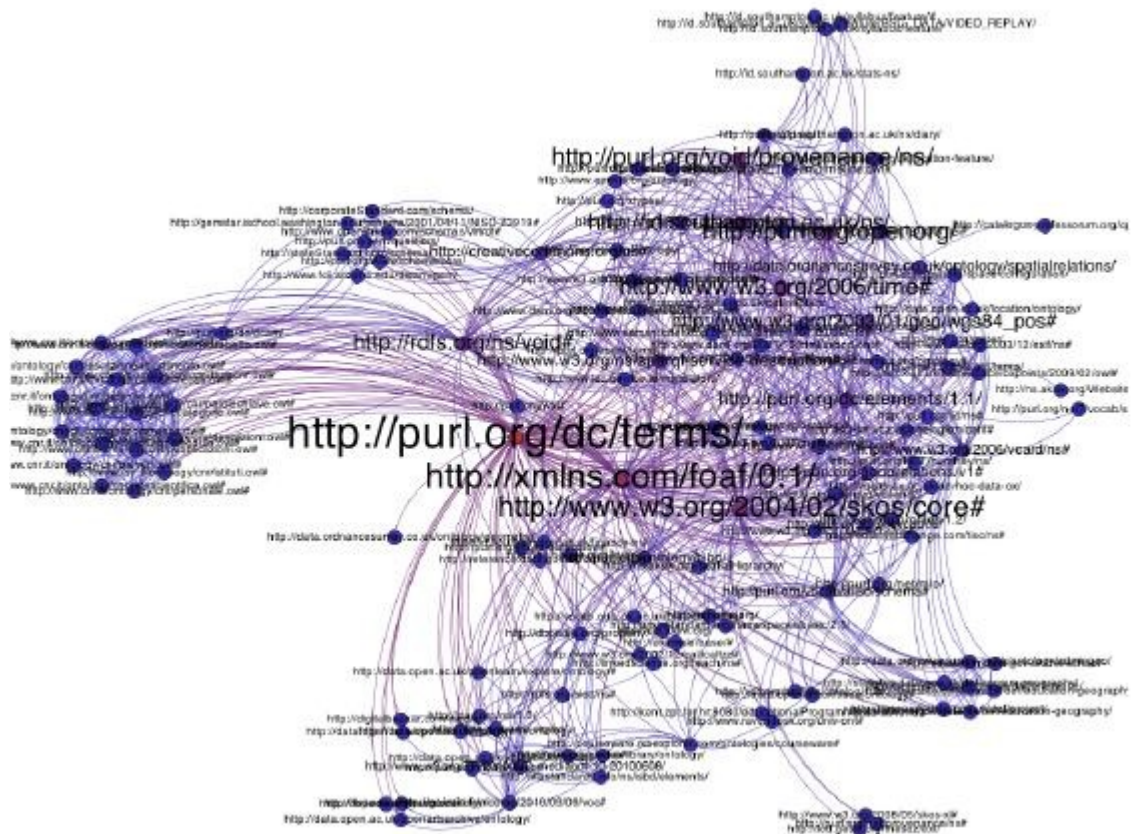


Figure 9: The educational network of vocabularies

Figure 9 shows graphically the network of vocabularies used in the educational sector. Node colors indicate the connectivity (red=most connected, blue=least connected), while label sizes indicate the number of datasets using each vocabulary. As mentioned above non relevant to education vocabularies, such as the Dublin Core or the FOAF, are the most widely used and even stand out as hubs connecting even the most isolated datasets. Furthermore, they present the types of resources that are included in the network, and show how trivial data integration effort, of mapping a couple of hundred types, can significantly improve the overall cohesion of the network. This was done by mapping 201 of the most frequent types in the datasets to 79 classes of pre-selected vocabularies. All this, shows the need for a coordinated effort to increase the coherence and interoperability of the educational linked data network, so that it can eventually turn into an entity-centric representation of resources and semantic applications can easily exploit it, to provide useful services to the educational community.

In [10, 11] the main challenge towards the vision of a Linked Education, meaning an education which exploits educational web data, is discussed. This challenge has two parts. Firstly, to start adopting linked data principles and vocabularies, and secondly to

start leveraging on existing educational data available on the web that is in not in a linked data compliant format. In order to follow such an approach various tasks have to be taken into concern, which will ensure the desired web-scale interoperability. First of all, a large amount of data is distributed in heterogeneous educational repositories, were each one of them offers a different service/API as a data access medium. This means that this data has to be integrated in order for interoperability to be ensured. Secondly, the continuous change of these web APIs has to be dealt with, in order to avoid further implications. Thirdly, since a large amount of data is usually described by using distinct, often XML based schemas or even with unstructured text, a transformation into an appropriate format, such as RDF, has to be applied, while also a mapping between similar concepts is of equal importance, in order to leverage on already existing educational data. Finally, this data has to be enriched with important RDF structured information and it should also be interlinked with other datasets. In order to overcome these challenges a three-layered architecture is proposed, which is shown in Figure 10 below. The first layer, called *The Educational (Web) data and service layer*, consists of available educational web services and data. The second layer, called *The Educational data and service integration layer*, is the heart of the architectural proposal and is responsible for annotating and interlinking educational services by making use of the linked data principles. *The Educational application and presentation layer*, which is the top layer, uses the APIs provided by the educational data and services integration layer to interact with underlying data and services and provides an interface to end-users. The proposed approach supports a particular methodology consisting of two fundamental steps, which are both facilitated by Linked Data technologies. These are:

1. Educational Services Integration, and
2. Educational Data Integration

The first step, which is also shown in Figure 10, is responsible for integrating educational services and APIs, since as we mentioned above, each API has its own access methods and response message formats, so integrating all these heterogeneities into one compact transparent service is essential. Also this function is responsible for the transformation of the retrieved from the various APIs information, into RDF. The *Educational Data Integration* is responsible for integrating the data that was retrieved from the Educational Services Integration function with the data that is already available in RDF format, from various datasets. This is done by creating interlinks between the

heterogeneous data, by using various methods such as, automated interlinking of datasets or automated clustering and classification.

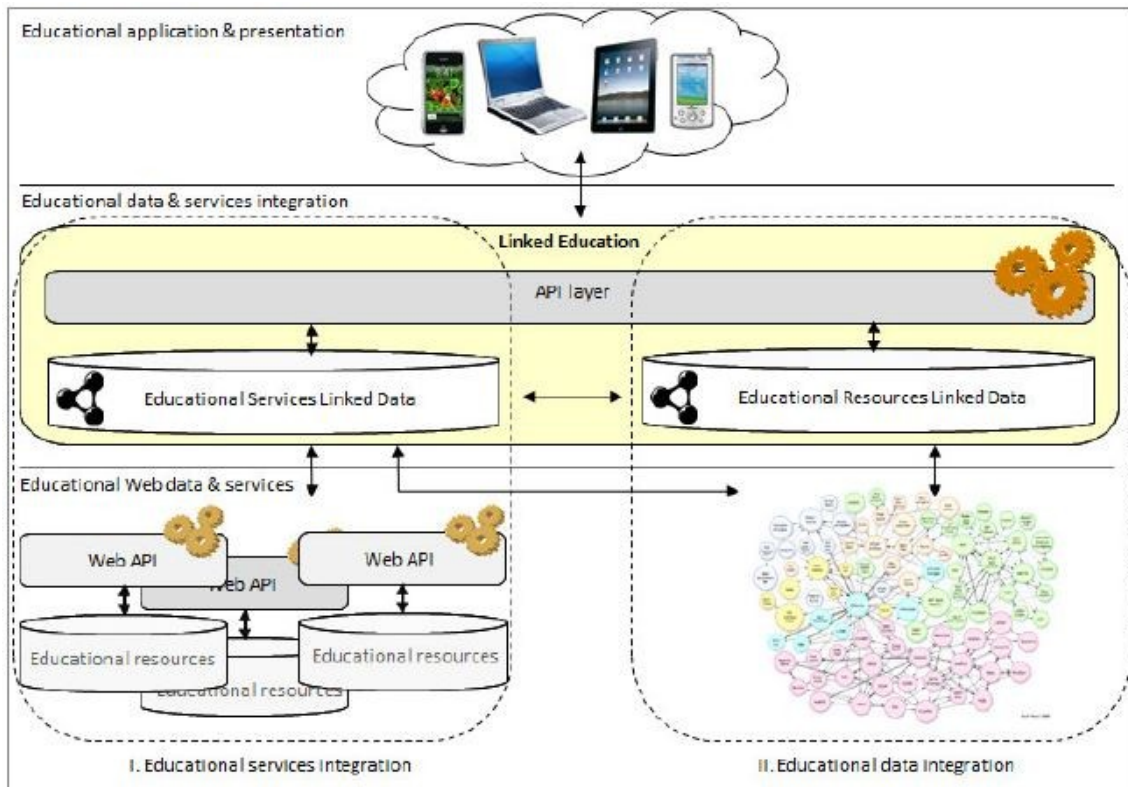


Figure 10: Educational Web data integration architecture

This architecture was applied in the context of the mEducator project [12] where data from a number of open technology-enhanced learning data repositories were integrated, exposed and enriched by following the Linked Data principles. The data and services integration APIs and datasets presented above were fully integrated in Metamorphosis+, a system which merges the paradigms of semantic and social web to produce an environment for sharing linked educational resources. To this end, Metamorphosis + interacts with the APIs provided by the Educational services and data integration layer instead of directly retrieving and processing data from disparate sources, therefore creating a network of educational resources. Overall, the proposed architecture had four major contributions which are:

1. Linked Data-principles were applied to model and expose metadata of both educational resources and educational services and APIs. In this way, not only resources were connected but also services' description and resources were exposed in a standardized and accessible way.

2. Existing heterogeneous and distributed learning repositories were integrated on the fly by reasoning and processing of Linked Data-based service semantics.
3. Metadata retrieved from heterogeneous Web repositories were automatically converted into RDF and exposed as Linked Data.
4. A set of APIs were developed on top of the integration framework to allow third party applications to consume and interact with the data that was exposed.

We believe that this proposed architecture could be a good starting point for the integration of heterogeneous data lying in isolated APIs, with data already published on the linked data cloud, although the correct transformation and integration of data from such an automated service is put to doubt. Although the educational linked data would increase in number, the quality of the produced data would still be ambiguous, since the semantic web is still in a premature stage and the appropriate technologies that would provide accurate transformations and mappings are yet to be implemented. Despite of these facts though, the main concept and the procedure for the implementation of an interoperable web of data, is well described and feasibly implementable to a satisfactory point.

[13] refers to the creation of a linked dataset that would contain information about ICT tools, such as wikis, chats or graphic editors, aiming at making educators task of retrieving educational tool relevant information much easier. The information concerning the various ICT tools would be described with the use of the Ontoolcole<sup>6</sup> vocabulary, while the dataset would be able to automatically import hundreds of ICT tool descriptions from third-party sources. On top of the dataset they aimed at creating two tools that would exploit the dataset and provide useful services to the educators. The first tool would be an educational ICT tool search system that would retrieve information that was published in the dataset, according to the queries the users made, while the second tool would provide an educational data publishing platform, which would enable the creation of educational data to enrich and evolve current description of ICT tools. Finally, they expected that third party contributors, such as educational institution and user communities, would also provide useful information to the existing data cloud, which would be to their benefit as more information would be interconnected to their dataset. The whole picture is shown in Figure 11.

---

<sup>6</sup> <http://www.gsic.uva.es/ontologies/ontoolcoleModel.owl>



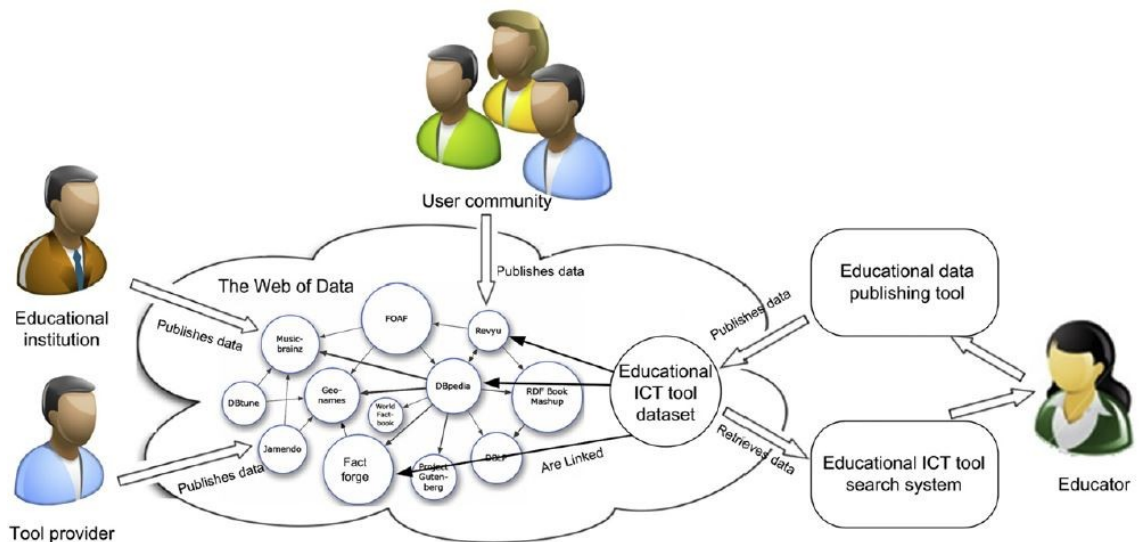


Figure 11: Main components and actors of the educational tool data infrastructure

The main problem they faced with this approach was the mapping that was required to be made with the existing vocabularies in order to align with them. This is graphically shown in Figure 12 where we can see some of the external datasets that would provide information to the educational tool dataset through mapping that would be made with their vocabularies,

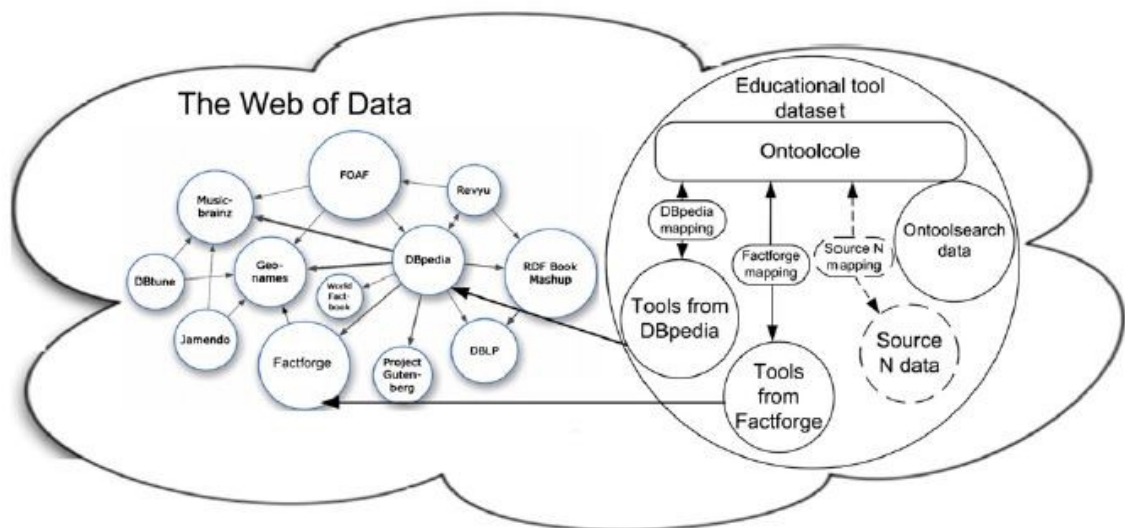


Figure 12

The final evaluation results showed that, when the data collected from the Web were related to the Ontoolcole ontology, educators were able to understand the produced dataset and found it useful for discovering ICT tools. Also the technological possibilities that would be able to support the educators learning tasks were addressed, when the questions they provided in order to evaluate the dataset returned a satisfying number of

potentially useful ICT tool descriptions. However, this is not enough to conclude that the information obtained from the Web of Data was adequate for educators to select a specific tool that would serve their needs, since more educational-specific information is required that cannot be currently retrieved from the Web of Data. Moreover, working with third-party external data had some more disadvantages than the ones mentioned above, since some of the tool descriptions that were found in them were incomplete or even contained errors. We can see that although the semantic web offers quite a few benefits such as collaborative publishing and maintenance of information over a commonly understood format, many questions and problems still remain unsolved since the appropriate technological infrastructure is quite immature. For example, the mapping that needed to be made between the ontologies, in this case, is a general semantic research topic that limits the capabilities of exploiting to the full range of benefits the semantic web has to offer.

The use of Semantic Web (SW) technologies in educational systems can help the accomplishment of AAAL: Anytime, Anywhere, Anybody Learning, where most of the World Wide Web resources are reusable learning objects supported by standard technologies and learning is facilitated by intelligent pedagogical agents. Motivated by this trend, [14] proposes a reference model for a Semantic Web-based educational system.

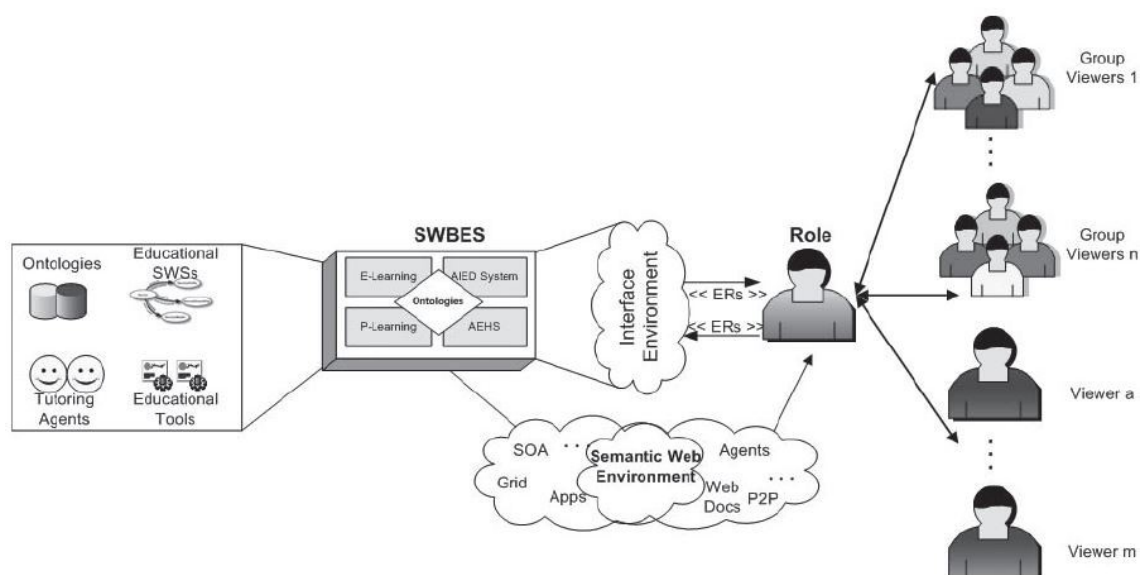


Figure 13: Reference Model of Semantic Web-based Educational systems

They state that the components that apart Semantic Web-based educational systems are the following, which can also be seen in Figure 13:

- Roles
- Interface environment
- Educational Resources
- Semantic Web-based Educational System
  - Ontologies, Pedagogical agents, Tools, Services
- Semantic Web Environment

A Semantic Web-based Educational system can be used by different users that are assigned different roles each time and the system should respond accordingly for each one of them. For example, a user with a teacher role is responsible for providing learning content and design learning activities, to monitor learners' interactions, to configure learners' strategies and to support students' evaluation, while a learner's main interest would be to interact with the system in order to receive personalized educational content in order to improve his knowledge and fulfill his learning goals. Every time the system should respond according to the goals of every user, providing him with the appropriate information. The interface environment represents the communication interface between a user with a specific role and the Semantic Web-based system and it is very important that it adjusts its interface for every different role, in order to facilitate different functionalities. The Semantic Web-based Educational system, being the heart of the system, is the one responsible for serving each different user with the education relevant resources he requested. This system involves ontologies, which are used to define the educational concepts, pedagogical agents, which help the learning process in various ways, tools, which can be used for user collaboration, simulation or authoring, and semantic web services, which can support a number of different educational activities, such as personalization of educational content and recommendations. Finally, the semantic web environment represents the interaction environment available to Semantic Web-based Educational systems and users in order for them to discover, browse, select, and invoke semantically annotated resources on the Web by employing semantic technologies and architectures. If such systems were to be implemented in the future, then users would be able to utilize appealing features, such as adaptive and personalized learning environments, enhanced instruction and learning capabilities,

information sharing, storage, retrieval and new ways of collaboration with peers. Inspired by the proposed reference model, an innovative e-learning solution named Intelligent Web Teacher [15] came from research projects. This system managed to model educational domains knowledge, users' competences and preferences by a Semantic Web approach, in order to create personalized and contextualized learning activities and to allow users to communicate, to cooperate and to dynamically create new content to deliver and information to share. We can therefore see that the semantic web is not so far away from becoming the new web, since innovative applications are able to be developed with the current infrastructure.

## **2.4 Similar Applications**

Since the Semantic Web started to emerge many initiatives were made by various organizations and communities towards making their data openly available on the LOD cloud. Furthermore, a large amount of useful applications have been developed, which proved the usefulness and importance of the Semantic Web. These applications operate by exploiting the datasets that are published on the cloud in order to provide useful and innovative services to the end users. In this subsection, we are going to present some of the most popular and relevant to our application initiatives that were made in this context. Most of them are university cases that decided to publish their data as Linked Open data, while in the last section we are also going to present the DBpedia case study, since it is currently the largest, most well-known and most interlinked dataset in the cloud.

### **2.4.1 OU Linked Open Data**

The OU (Open University) is the first UK University to expose and publish its organizational information in Linked Open Data format [16] and this was accomplished as part of the LUCERO project (Linking University Content for Education and Research Online), in the context of which the data extraction, transformation and maintenance were performed. This enabled the combination of various datasets, which were scattered among the different organizations of the OU, so that they could be accessible in an open way through a common web interface. By doing this, members of the public, students, researchers and organizations were able to easily search, extract and reuse OU's information.

The university decided to publish information concerning courses they offer, podcasts and academic publications. In order to achieve this they used quite a few already published ontologies, many of which are used in our application, since they are also supporters of the Semantic Web's trend, which is to reuse already available information. Some of the common ontologies that were also used in our application, are the following: Courseware, XCRI and MLO.

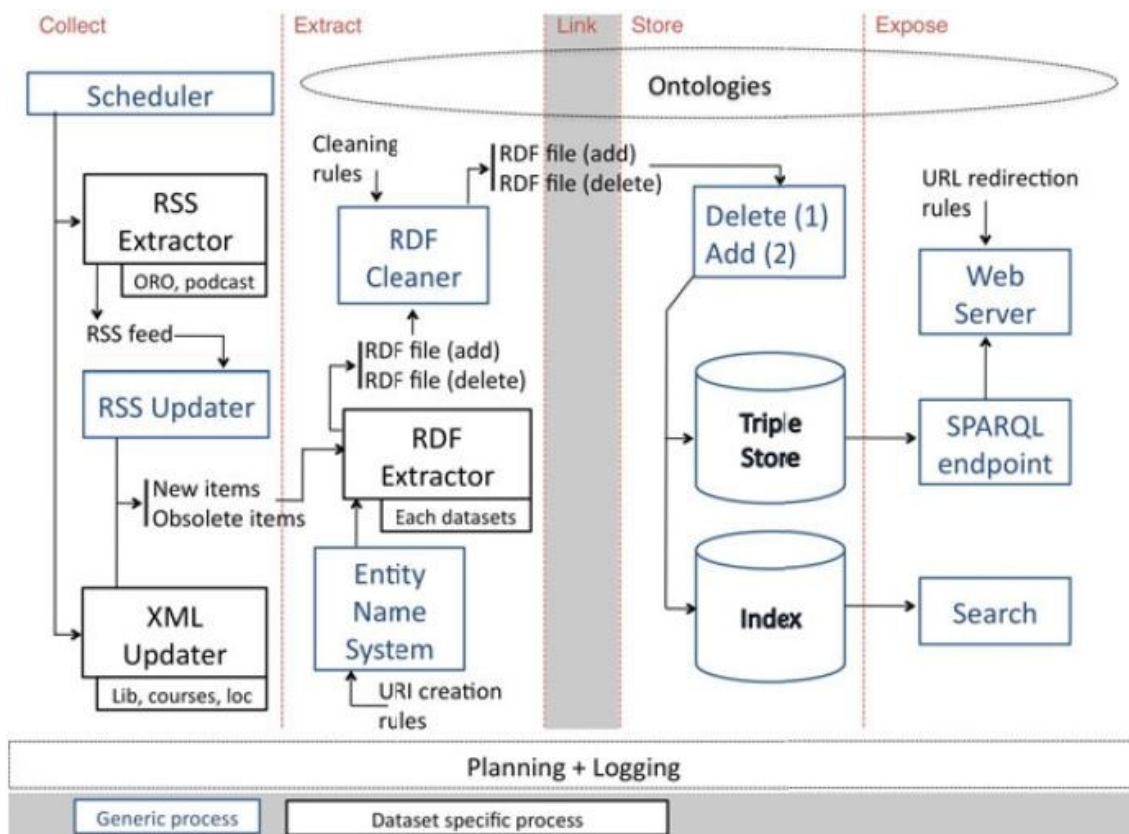


Figure 14: The OU system architecture

The architecture of their system is shown in Figure 14. We can see that it is divided into five different stages, which are the following:

1. Data collection
2. Data extraction
3. Data linkage
4. Data storage
5. Data exposure

During the collection phase a scheduler frequently retrieves newly created data that is scattered among different organizational data repositories and with different formats. After the data is gathered it is passed to the extraction phase which is responsible for

extracting the information that is held in a different format, cleaning it and finally converting it into RDF. After the RDF information is inter and intra linked with external datasets, it is stored in the triple store and then it is exposed to the web where it can be queried through the provided SPARQL endpoint.

A very interesting application developed on top of the exposed dataset is the OU Expert Search application. Its task is to identify people who have relevant expertise in a topic of interest, which addresses many common problems such as finding colleagues that are experts in a particular subject [21] or assembling a consortium with the necessary range of skills for a project proposal.

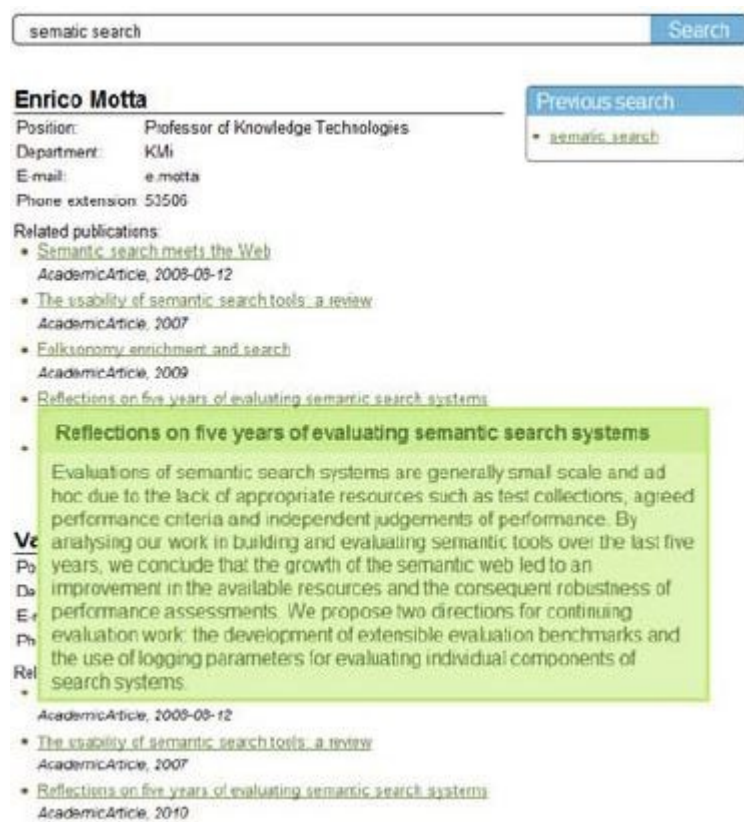


Figure 15: The OU Expert Search application

A screenshot of the system is shown in Figure 15 where we can see one of the returned results about an expert in a relevant to the search topic. We can see that for each expert the system provides his contact details, the top publications concerning the searched for topic and some information for each publication.

Another interesting application is the Social Study application which combines the university's dataset information with information extracted from Facebook's Open Graph. In this context, the application extracts useful information about the user from his Facebook profile and combines it with the OU exposed information in order to

provide him with recommendations, about courses the university offers that may fit his interests. Figure 16 below, shows the top ranked courses that were returned from the system to the user as suggestions. Next to each result we can see the user's interest concepts, which were stated in his Facebook profile that made each of the shown courses to be returned as results.

## Social Study – Course Suggestions based

[Logout](#)

### Courses

[The technology of music](#) (techno, music, )

[Relational database systems](#) (techno, simple, )

[Challenging obesity](#) (eating, )

[Professional Graduate Certificate in Education: Secondary Music](#) (music, )

[The music dissertation](#) (music, )

[Voices and texts](#) (music, )

[Performances and repertoires](#) (music, )

Figure 16: The Social Study application

### 2.4.2 Linked Open Data University of Münster

As in the case of OU in UK, the University of Münster was the first university in Germany to publish its data as Linked Open Data, as part of the Linked Open Data University of Münster (LODUM) project [17]. The data that was included in their dataset contained scientific data and publications, as well as administrative data such as building databases and course schedules. They used the Bibliographic Ontology (BIBO) to describe bibliographic resources, such as books and articles, the Dublin Core (DC) ontology for metadata wherever it was required, the Friend of a Friend (FOAF) ontology for information concerning people and organizations, the W3C Basic Geo vocabulary for all things that were geo referenced and finally the Teaching Core Vocabulary (TEACH) was used to describe courses they offer. The architecture of their system is shown in Figure 17 and it is quite similar to the architecture that was used in the OU case. A scheduler periodically fetches data that is distributed among different systems and after it is converted into RDF through the custom triple factories it is pushed on the central triple store. The management tools support tasks such as vocabulary

management and link discovery between the initially disconnected datasets imported into the triple store, while the retrieval process exposes the stored data on a web interface in a human friendly way. The data can also be queried through the provided SPARQL endpoint. In order to demonstrate the use of the LODUM infrastructure they developed two interesting applications on top of the dataset, which are worth mentioning. The first one was a productivity map of the university's researchers. It rendered the university buildings in 3D, where the building height indicates the number of publications written by researchers working in the respective building. We can see a screenshot of the application's results in Figure 18. The second application, called The Westfälische Wilhelms-Universität Münster (WWU), was a mobile app that offered students, staff and visitors information about cafeteria menus, navigation instructions to university buildings, and an overview map of all university buildings on their mobile phone.

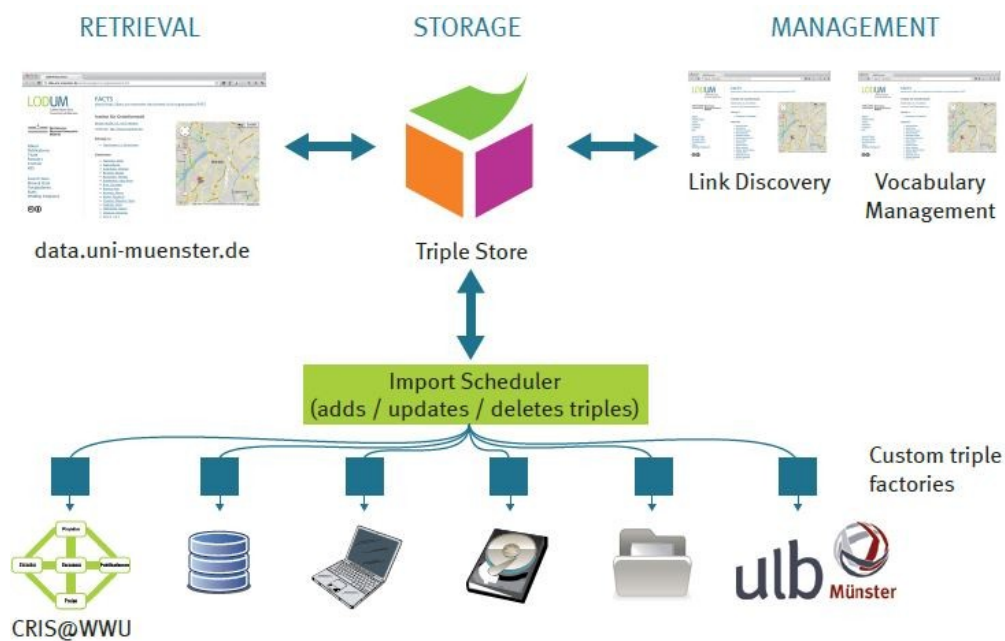


Figure 17: The LODUM workflow



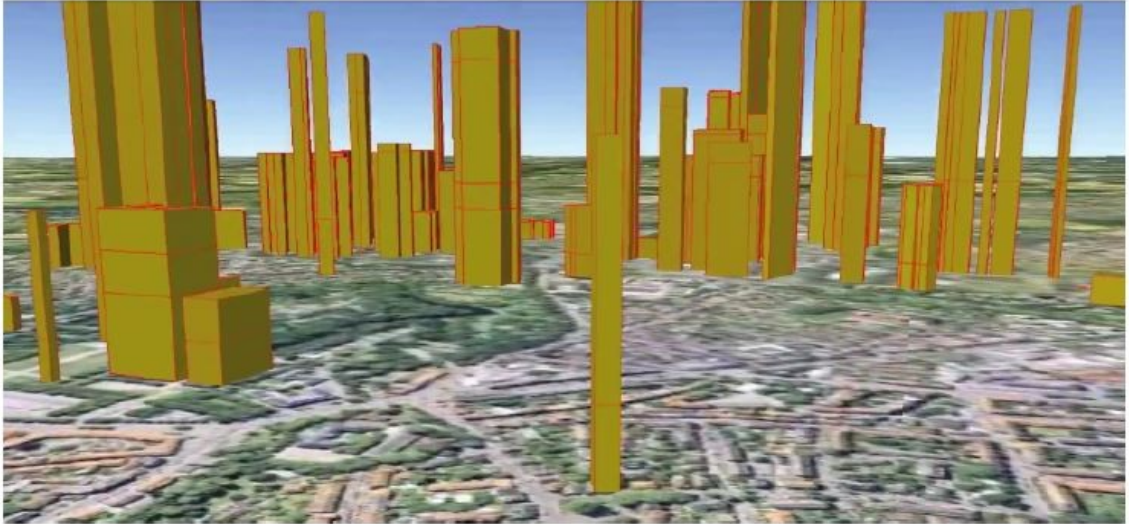


Figure 18: Productivity map of Münster's researchers

### 2.4.3 Tsinghua University Open Data

[18] presents a general process of building linked open university data, and uses Tsinghua University as a showcase. Tsinghua University exposed information about its campus buildings, its educational administration, such as course and exam schedules, its faculty members and basic information concerning the university. In Figure 19 we can see its current datasets and the connections between them.

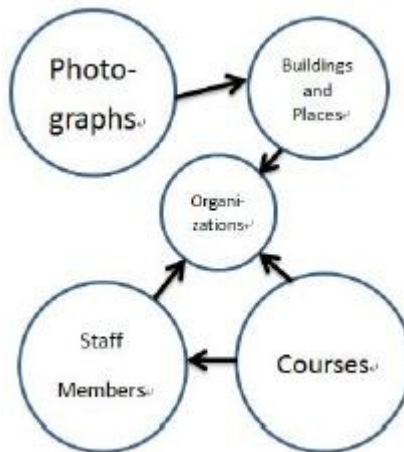


Figure 19: Tsinghua University datasets

It reused popular vocabularies relevant to the educational sector, which were the following: FOAF, AIIISO, Org and CourseWare. Most of their unstructured data was extracted from the university's website after the relevant webpages were crawled and in the process a lot of inconsistencies were noticed, which had to be corrected later on.

After the data from various sources was gathered it was assigned with the relevant URI and it was converted into RDF with the assistance of tools such as D2R Server and ConvertToRDF. Their data was then exposed as in the other cases through a web interface and a SPARQL endpoint. Finally, they developed two applications on top of the exposed data cloud. The first one, called CampusAssistant, exploited the Buildings and Places dataset and the Photographs dataset, in order to provide a searchable map for finding buildings and generally navigational information concerning Tsinghua University. The CourseFinder was an application which provided various ways for searching course relevant information, such as lecturers teaching a specific course.

#### 2.4.4 The Data.dcs Project

The University of Sheffield's Department of Computer Science is another great example of an initiative made by an academic institution to publish their legacy data into the LOD cloud [19], thereby assigning to their data a machine readable format. They decided to publish information concerning members of their department, their publications and research groups they belong to. The triplification of the data was quite a challenge since data concerning members and their research groups were embedded into HTML pages, while data concerning their publications were provided in RSS format. Thereby two different methods had to be developed in order to extract the appropriate information from each individual source.

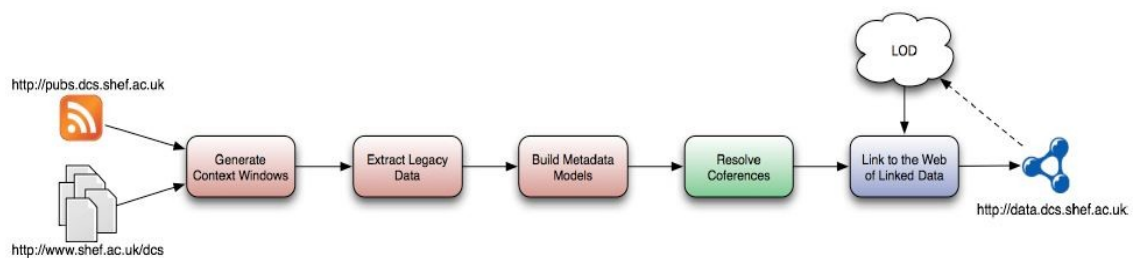


Figure 20: Three staged approach to convert legacy data to linked data

In Figure 20 we can the three staged approach the project followed in order to convert legacy data to linked data. The three phases which are pipelined are the following:

- Triplification
- Coreference Resolution
- Linking to the LOD cloud

The first phase included the triplification of legacy data, which was in an incompatible format as mentioned above. In order to extract the appropriate information, context windows were derived from the HTML pages and the RSS feeds, where each context window contained information about members or publications, and Hidden Markov Models were used to extract member or publication information from the windows. From this extracted information the triples were created using the FOAF ontology to describe members and the BIBTEX to describe members publications.

The second phase involved the resolution of co-references, since some instances that were created during the triplification phase might have been referring to the same entity e.g. equivalent people appearing in separate web pages. This was done by applying SPARQL rules in the dataset that discovered similar instances. Furthermore, in order to provide rich RDF descriptions, the information concerning each separate entity should be linked with each other. For example, a member of the department should be linked with the publications he has made and with the research group he belongs to. This is shown in more detail in Figure 21, where we can see that Mathew Rowe is a member of the “Organizations, Information and Knowledge Group” research group, and that he has published a book with title “Proceeding of Linked Data on the Web Workshop, WWW, Madrid, Spain”.

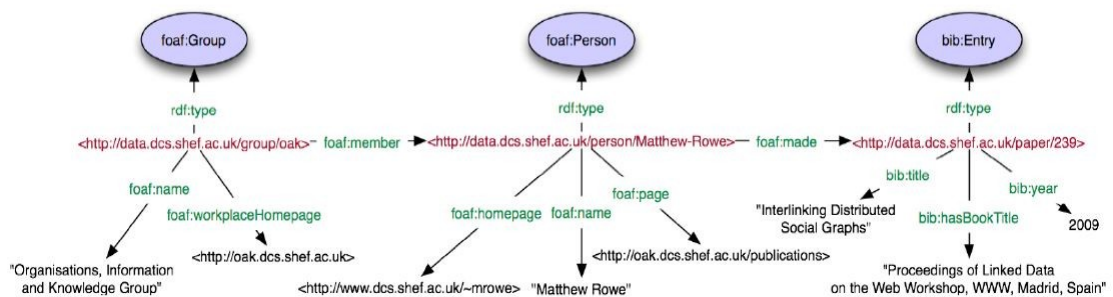


Figure 21: A snippet of the interlinked dataset

The third phase involved the interlinking with other available datasets in the LOD cloud. This was done by querying the LOD cloud for similar entities and related information resources, and linking this information with that in the produced dataset.

## 2.4.5 The LOD.CS.UNIPA Project

The LOD.CS.UNIPA project [20] refers to the initiative the University of Palermo took to publish their Computer Science’s department website data, in LOD format.

Currently they have transformed the available data in an LOD compatible format where each item has a URL that can be dereferenced and their next goal is to create interlinks with other LOD datasets. Their dataset currently contains information about users (teaching staff and administrative staff), courses (with their modules) and subjects, while in order to describe their data they re-used well known ontologies such as AISO, AISO Roles and FOAF, on top of which they made some additions to fit their exact needs.

The transformation of the website data to LOD compatible data had three phases, which were ontology building, data extraction and storing, and access interface creation, as shown in Figure 22.



Figure 22: Data transformation phases

In the ontology building phase specific properties and classes were extracted from the above mentioned ontologies, while additional adjustments were made in order to conform to the Italian University system. For example some entities such as the ECTS values and Erasmus eligibility were not included in these ontologies and it was necessary to develop a more extensive ontology.

In the data extraction and storing phase, a dump-based technology was used in order to store the data that was in the SQL database into dump files. After that, Triplify was used to extract the information gathered in the dump files and convert it into RDF triples. Finally, ARC2 was used as the triple store for the generated RDF data.

The access interface was a SPARQL endpoint that allowed querying the triple store, while also providing an unambiguous resource identifier allowing correct URI dereferencing. This allowed users to access the URI and use information contained in a single resource.

#### **2.4.6 The AcademIS Ontology**

The AcademIS ontology [21] is a relatively new ontology created in the context of representing academic activity and collaborations within higher educational institutions

(HEIs). It provides a tool that can be used by HEIs in order to assist them in processes such as evaluation of research and educational and collaboration activity among academics. More specifically, it provides an ontology that covers the teaching activities and connections of academics, and facilitates the research networking in educational institutions with emphasis on the formulation of collaborations among the academic staff. On top of the provided ontology, it also presents services and tools that provide visualization of the various teaching and research collaboration associations in order to support processes such as the quality evaluation of academic organizations. The ontology was developed as an extension of the Vivo Ontology since it already provided various research concepts and their associations. Among them are descriptions for publications in journals, conferences, or the publication of books, the collaborations between faculty members, the equipment of the academic institution and the related events that take place within an institution. The AcademIS Ontology introduces the following concepts (which can also be viewed in Figure 23), that were not provided by the Vivo Ontology, in order to describe concepts such as teaching collaborations, courses and their relationship information, like prerequisites, proposed and completed thesis topics, scholarships, internships etc.:

- Teaching collaborations
- Internships
- Scholarships
- Thesis

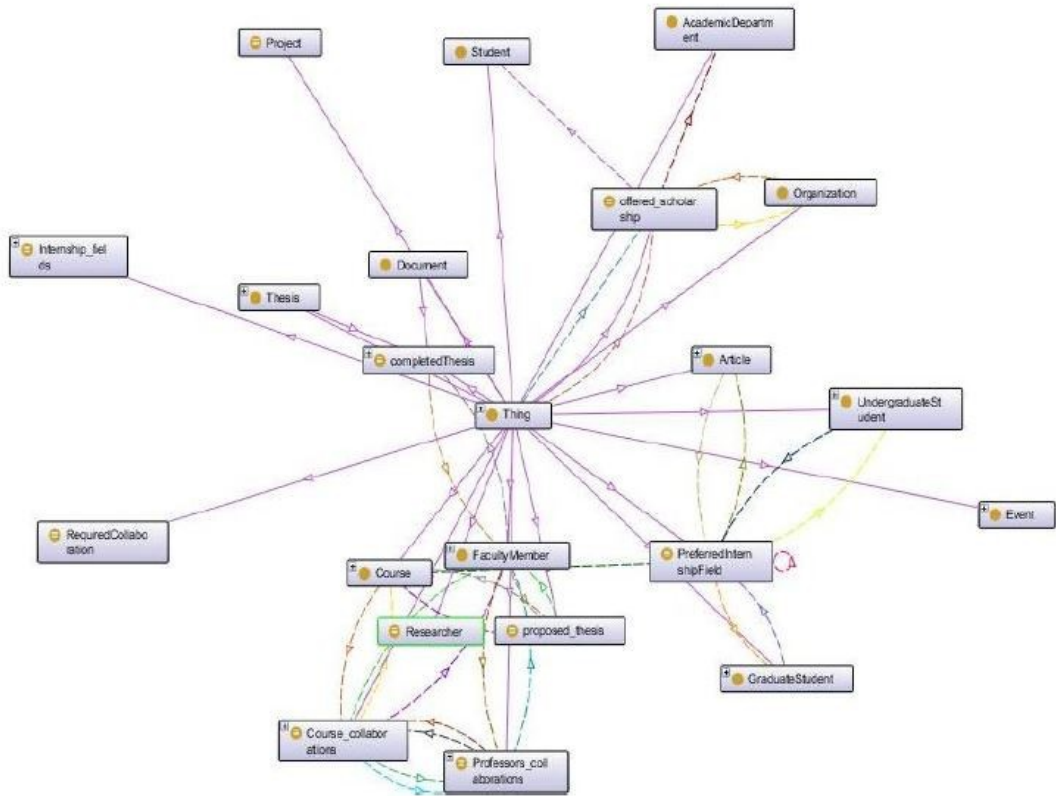


Figure 23: The newly introduced components of the AcademIS

The Teaching collaborations concept describes the collaborations between the professors in a specific academic department of the HEI. The Internships concept describes the various information concerning a student that wants to follow an internship within an organization, while the Scholarships concept provides information about all the scholarships provided by a specific department. Finally, Thesis expresses the necessary information about theses within an academic department.

In order to demonstrate the benefits of the AcademIS ontology the IREMA (Institutional REsearch MAnagement) tool was used on top of the ontology in order to provide visualization services that would assist the above mentioned needs. In Figure 24 we can see one of the visualization services provided by the IREMA tool, which demonstrates the overall research activity of all the scientific fields within a HEI.

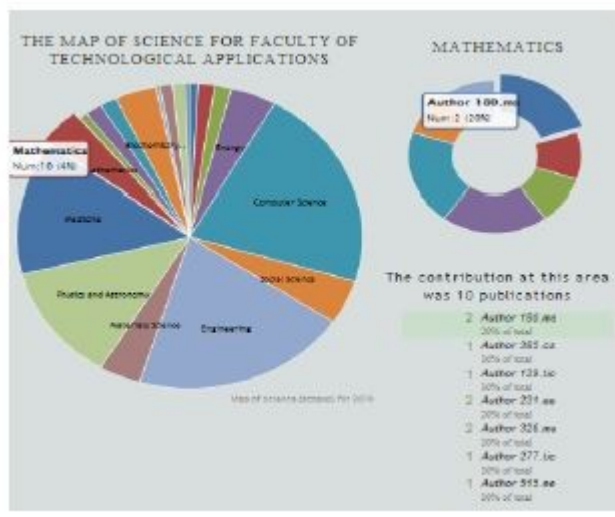


Figure 24: Map of science

The educational collaboration network service’s results are shown in Figure 25. This service examines the teaching collaborations of a specific professor in the courses in which he participates.

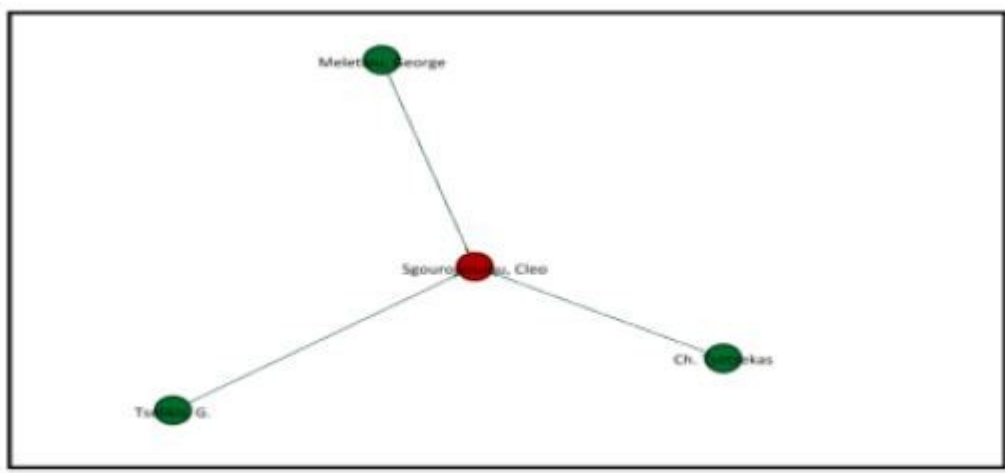


Figure 25: Collaboration networks of professors

In general, we can see that ontologies such as the one described above can provide many benefits to different stakeholders of a HEI, since many services can be developed on top of the provided concepts that can assist in the previously tedious processes that were required by each individual stakeholder.

### 2.4.7 DBpedia

DBpedia [22] is currently the largest dataset in the LOD cloud with millions of triples and interlinks making it one of the most important knowledge bases openly

available to the public. The dataset is so large that it is divided into smaller ones each describing a different domain. The currently available datasets and their number of triples are shown in Figure 26 . We can see that its largest datasets are Infoboxes, which are going to be discussed later on, with approximately 15.5 million triples and second comes the Articles dataset, with 7.6 million triples, which is the heart of Wikipedia. The main reason for DBpedia’s size and success is that it is a community effort allowing individual parts to participate in the construction of this magnificent knowledge base. DBpedia’s datasets are derived from Wikipedia and it allows to ask sophisticated queries against these dataset and to link other datasets on the Web to Wikipedia data.

<b>Dataset</b>	<b>Description</b>	<b>Triples</b>
<i>Articles</i>	Descriptions of all 1.95 million concepts within the English Wikipedia including titles, short abstracts, thumbnails and links to the corresponding articles.	7.6M
<i>Ext. Abstracts</i>	Additional, extended English abstracts.	2.1M
<i>Languages</i>	Additional titles, short abstracts and Wikipedia article links in German, French, Spanish, Italian, Portuguese, Polish, Swedish, Dutch, Japanese, Chinese, Russian, Finnish and Norwegian.	5.7M
<i>Lang. Abstracts</i>	Extended abstracts in 13 languages.	1.9M
<i>Infoboxes</i>	Data attributes for concepts that have been extracted from Wikipedia infoboxes.	15.5M
<i>External Links</i>	Links to external web pages about a concept.	1.6M
<i>Article Categories</i>	Links from concepts to categories using SKOS.	5.2M
<i>Categories</i>	Information which concept is a category and how categories are related.	1M
<i>Yago Types</i>	Dataset containing rdf:type Statements for all DBpedia instances using classification from YAGO [16].	1.9 M
<i>Persons</i>	Information about 80,000 persons (date and place of birth etc.) represented using the FOAF vocabulary.	0.5M
<i>Page Links</i>	Internal links between DBpedia instances derived from the internal pagelinks between Wikipedia articles.	62M
<i>RDF Links</i>	Links between DBpedia and Geonames, US Census, Musicbrainz, Project Gutenberg, the DBLP bibliography and the RDF Book Mashup.	180K

Figure 26: The DBpedia datasets

In Figure 27 we can see DBpedia’s main components and their interactions with each other. At the lowest layer are the Wikipedia dumps.



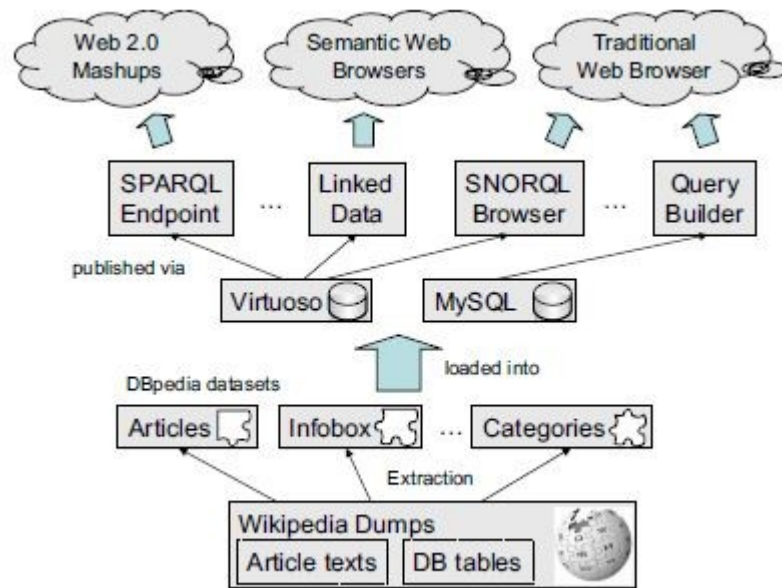


Figure 27: Main DBpedia components

These dumps include structured information that is periodically cached in relational database tables, as well as structured information which is extracted from Wikipedia’s article texts, which are firstly parsed by specific software. Based on these database dumps, they use two different methods for extracting the information and creating the datasets:

- They map the relationships that are already stored in relational database tables to RDF and
- They extract additional information directly from the article texts and infobox templates within the articles.

DBpedia’s infoboxes are data attributes for concepts that are described in Wikipedia articles and have a structured representation that allows them to be easily parsed and therefore important knowledge can be extracted. An example of a Wikipedia infobox is shown in Figure 28 where we can see how information for Busan metropolitan city is represented in a structured manner. We observe that specific attributes describing a city are assigned with relevant information described in the Wikipedia article.

```

{{infobox City Korea|
  full_name=Busan Metropolitan City|
  image=[[Image:Haeundaebeachbusan.jpg|
    250px|Haeundae Beach, Busan]]|
  rr=Busan Gwangyeoksi|
  mr=Pusan Kwangyŏksi|
  hangul=부산 광역시|
  hanja=釜山廣域市|
  short_name=Busan (Pusan; 부산; 釜山)|
  population=3,635,389 ...|
  area=763.46 km2|
  government=[[Metropolitan cities of
    South Korea|Metropolitan City]]|
  divisions=15 wards (Gu),
  <br>1 county (Gun)|
  region=[[Yeongnam]]|
  dialect=[[Gyeongsang Dialect|
    Gyeongsang]]|
  map=[[Image:Busan map.png|Map of
    South Korea highlighting the city]]|
}}

```

Busan Metropolitan City	
	
Korean name	
Revised Romanization	Busan Gwangyeoksi
McCune-Reischauer	Pusan Kwangyŏksi
Hangul	부산 광역시
Hanja	釜山廣域市
Short name	Busan (Pusan; 부산; 釜山)

Figure 28: Wikipedia infobox

These infoboxes allow extraction algorithms to detect and recognize them, using pattern matching techniques. After they are detected, they are then parsed and transformed to RDF triples. After the creation of the RDF triples they are stored in Virtuoso triple store and in MySQL databases and they are made available for external exploitation through SPARQL endpoints or appropriate browsers.

DBpedia's datasets are interlinked with numerous other datasets on the LOD cloud allowing the discovery and exploration of additional information. The importance of interlinking is shown by an application that was developed on top of the DBpedia dataset. The *Search DBpedia.org*, as it is called, allows users to explore the DBpedia dataset together with information from interlinked datasets such as the Geonames, the RDF Book Mashup and the DBLP bibliography datasets, by taking advantage of the relations in the data, enabling the narrowing of the returned search results in different domains of interest.

## 2.5 Conclusions

We have presented the most relevant to our dissertation research work that has been done. We firstly saw some of the work done concerning the semantic web in general and the main research directions that are to be followed in this sector. Then we explained

some of the Linked Data design issues that engineers may face when building their ontologies. We saw that there is a great need for development of tools that will assist with ontology reuse issues, such as ontology mapping and data fusion. In the third subchapter, we took a look at what is currently going on in the educational sector as far as linked data is concerned and we saw that a lot of work and effort is being put in order to develop a concrete semantic system that will collect the educational resources and provide a unique interface for accessing and exploiting them. Since there is currently a great interest in the Linked Open Data movement and many organizations take the initiative to publish their data on the cloud, so that it is made available to the public for exploitation, in the fourth subchapter, we presented DBpedia's community effort and some of the University's that have started this ongoing process. Of course there are also other examples such as the University of Bristol, the University of Edinburgh, the University of Oxford and the University of Manchester which decided to publish in LOD format their library catalogue records. We also saw many different approaches followed in order for each institution to convert their data, which was available in different incompatible formats, into RDF. This stage of the process was the most difficult one for all projects, since each data source required a different method to extract the unstructured data and convert it into a machine readable format. The inter and intra linking of the generated data was also not trivial, since the correct mappings had to be made between the generated triples, because the same data may have been present in different data sources. Finally, some applications developed by some of the institutions were presented in order to demonstrate how easily the data that is published can be combined and exploited for developing very useful and informative applications that can make the end-user's searching and browsing experience much easier. We strongly believe that the Semantic Web is still in an immature phase and that when the correct tools and approaches are developed the web will no longer be the same.

## **3 Problem Definition**

Sharing of resources and information across the web has been a tedious and of great importance task in the scientific community. Producing information and putting it on the web is just half way down the road. The rest is about making it available in a structured

format, so that it can easily and openly be accessible in order for it to be reused and shared among different communities and individuals. The preservation and accessibility of these resources is also of great importance, in order to ensure the transparency and reproducibility of them. Even these days knowledge reuse is a great problem, since the appropriate technologies are not yet quite mature in order to effectively enable the reuse and sharing of knowledge and currently technologies such as repositories, with proprietary interfaces and query mechanisms, are used. On top of that, as we mentioned in chapter 1, the technologies that are already used do not seem to agree on common terms in order to allow the interoperability of resources.

Since the last few years various communities promoted the exploitation and reuse of various domain resources, there has been a large movement towards the development of various technologies that would enable this vision to become a reality. This movement refers to the so called Linked Open Data Movement. At the same time, the Semantic Web started to emerge and its Linked Open Data principles started to prevail as the technology that would finally enable the web to become a hosting environment for interoperable information, which would be able to be reused, shared and understood not only by humans, but also by machines. Already a great amount of vocabularies have been developed by various organizations covering a wide range of domains, allowing the reusability of concepts and the continuous growth of the Linked Open Data cloud, with the addition of new interlinked datasets.

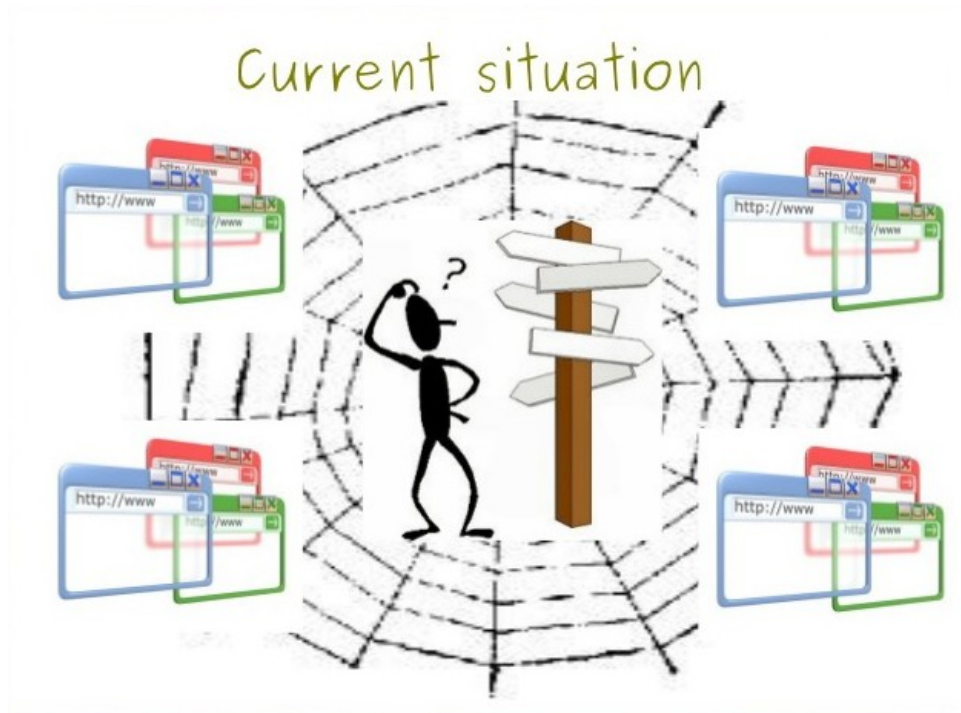


Figure 29: Before the Semantic Web

To be more specific, the main problem of the Web 1.0, as shown in Figure 29, is that when we want to search for something over the web we have to go through a large amount of documents, having to follow links that let us discover them, and most of the times the information included in a large percentage of them is irrelevant to the subject of our interest. It is a fact, that resources are generally duplicated and dispersed among different systems and databases, and the key concepts within these resources as well as their inter and intra connections are not explicitly shown to users. Moreover, the information lying in the documents has no conceptual schema and no inference can be applied in order to extract any further knowledge that would help our investigation. This makes the user's task very tedious and time consuming and for an amateur user there is a great chance, he will not find what he was looking for in the first place. The emerging Semantic Web technologies aim in solving these exact issues.

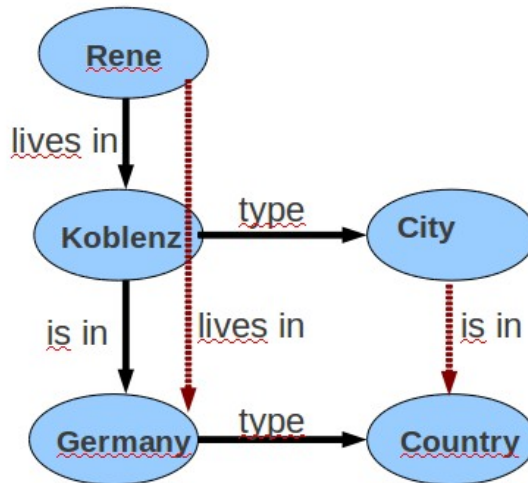


Figure 30: Linked Open Data in a nutshell

Figure 30 simply illustrates the form that the information will eventually have when it is published over the Linked Open Data cloud, using Semantic Web technologies. We can easily see the difference in the data representation model, over that of Web 1.0. It is now clear to the human and to the machine that is going to parse the data that Rene lives in Koblenz and in Germany and that Koblenz is in Germany. In the conceptual level, we also know that Koblenz is a city and Germany is a Country, while it is stated that a City is in a Country. This time the information is explicitly stated in a structured way and through its inter and intra connections additional information can easily be inferred, thereby making the end-users investigation experience much easier.

## 4 Implementation

In this chapter we are going to describe in detail the implementation of the application and the workflow that was followed in order to convert the information of interest from a non-interoperable format into RDF/RDFa. Firstly, we are going to present the information that we decided to publish as Linked Data and the ontologies that were used or extended in order to describe this information in a conceptual level. Also we will provide an overview of the instance level created on top of the ontology. Later on in this chapter, we describe the system infrastructure that was used and the workflow that was followed in order to build the application from the conceptual level

up to the data presentation layer. Furthermore, we present the various methods with which someone could access the generated dataset and we also present some examples of how someone could use the application and some screenshots of the application's user interfaces.

## 4.1 Data Categories and Schema

Below we present the information concerning the International Hellenic University that we decided to publish as Linked Data as well as the vocabularies that were used in order to describe them.

### 4.1.1 The Datasets

Currently the dataset mainly involves information concerning IHU's School of Science and Technology while from the other Schools only basic information is described. To be more specific, the following datasets were created which involved information from various aspects of the university:

- *People*: This dataset describes information concerning the university's faculty staff. It describes general information about a faculty member and the position he holds in the university ranking, such as Dean or Professor, and it links with the Organizations and Courses datasets to describe in which organization a member belongs to and what courses that member may be currently teaching.
- *Organizations*: The Organizations dataset describes general information about the International Hellenic University and the Departments (Schools) it consists of. It contains links to the Programs dataset in order to inform about the Programs the University and each separate School offers.
- *Courses*: This dataset contains description about each course the School of Science and Technology offers. It provides many details concerning each course, such as the course's learning outcomes, its assessment methods, its ECTS credits, the term it is offered in, its total hour's engagement etc. It contains links to the Organizations dataset to describe the University and School it is offered by and to the People dataset to connect with the faculty members the course is taught by.
- *Programs*: The Programs dataset describes information about the degrees the university offers. It describes information such as the level of degree (e.g.

Postgraduate), the duration, its cost, the qualification it offers (e.g. MSc, MBA etc.), when it starts, the study mode it is offered in etc. It contains links to the Courses dataset to describe what courses each Program offers.

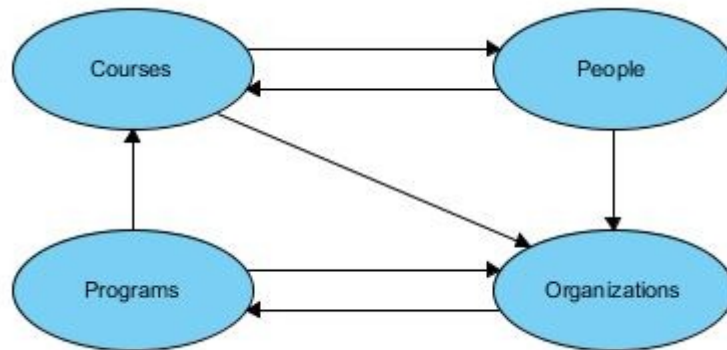


Figure 31: OpenIhu datasets and their interactions

In Figure 31 we can see the datasets that we described above and their interactions. We can see that the Programs dataset interacts with the Organizations and Courses dataset, the Courses dataset with the Organizations and People datasets, the People dataset with the Organizations and Courses datasets and finally the Organizations dataset with the Programs. Following this architecture we achieve having all data intra connected with each other so that machines, such as software agents, can traverse through the whole graph and discover further information that might be useful to the end user. Also, some bi-directional relationships between the datasets are present to make the navigation though the graph easy from any direction someone may start traversing the dataset.

#### 4.1.2 Vocabularies and Extensions

In this subchapter we are going to present the conceptual level of our application. More specifically, we will present in detail the schemas that were used in order to describe data concerning the International Hellenic University. The main idea was to reuse already published ontologies that are made available to the public through their Unique Resource Identifiers (URIs), since this is one of the main concepts of the Semantic Web. There is no point for creating new ontologies for concepts that are already described and accepted as they are by the community. Furthermore, the reuse of ontologies helps the creation of a global descriptor for commonly used concepts that will assist semantic agents and applications to understand their meaning and thereby provide useful services to the end user. By following this approach we imported in our



own ontology the ontologies that are going to be presented later on, in order to construct a unique ontology that would fit our exact needs. In cases where some of our concepts could not be described by the provided schemas, new properties and classes were created which were added either as extensions to the reused ontologies or as newly created concepts.

### ***The Reused Ontologies***

In order to conform to the Semantic web proposals we decided to create our ontology by reusing already published ontologies that are widely accepted and fit our exact needs. Below the list of reused ontologies is presented:

- *ReSIST Courseware Ontology*: The ReSIST Courseware Ontology <sup>7</sup>is used to represent the various educational courses and resources within the ReSIST project. Part of this ontology was used in our application in order to describe the courses that the IHU offers. We described information such as the objectives of a course, the number of ECTS credits it offers, the assessment methods of the course, the instructor(s) of the course etc. This ontology is used in other similar applications developed by educational institutions, such as the Open University of UK.
- *Metadata for Learning Opportunities (MLO)*: Metadata for Learning Opportunities (MLO)<sup>8</sup> is a European standardized model addressing metadata sufficient for advertising a learning opportunity. The MLO standard is also designed to facilitate semantic technologies and web architectures to support several mechanisms for exchange of the information and aggregation of information by third party service suppliers. It was used by our application in order to describe concepts concerning the programs and courses the University offers, the university's departments and the University itself. In the case of the university and its departments it was used to describe the programs the university and each separate department offers, while in the case of the department's description it was used to describe the learning outcomes of an offered course. The programs the University offers also used quite a few concepts offered by the MLO ontology. Among them were concepts concerning the cost of a program, its duration, its language of instruction, the qualification it offers etc.

---

<sup>7</sup> <http://courseware.rkbexplorer.com/ontologies/courseware>

<sup>8</sup> [http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/mlo\\_rdfs.xml](http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/mlo_rdfs.xml)

- *XCRI-CAP - XCRI Course Advertising Profile*: XCRI<sup>9</sup> is a UK-oriented project which aims establishing a specification to support the exchange of Course-Related Information. A key activity for XCRI is the development of an XML specification, the XCRI Course Advertising Profile (or XCRI-CAP for short), but it's also offered in a RDFs format which is the one we used in our application. Using this specification learning providers can publish their courses information in a standard, interoperable and semantically enhanced format, so that it can be collected easily by organizations with course search services. Opening up the offerings of learning providers creates new possibilities for value-added services and information channels for universities, colleges, and training providers. The XCRI ontology was used in order to describe the study mode (Part time or Full time) a university program is offered in, and the learning outcome of a course.
- *Lehigh University Benchmark (LUBM) Ontology*: The Lehigh University Benchmark<sup>10</sup> is developed to facilitate the evaluation of Semantic Web repositories in a standard and systematic way. The benchmark is intended to evaluate the performance of those repositories with respect to extensional queries over a large data set that commits to a single realistic ontology. It consists of a university domain ontology, customizable and repeatable synthetic data, a set of test queries, and several performance metrics. The LUMB ontology was used in order to describe concepts concerning people that work in the university, and concepts associated with the university and its departments. More specifically, it described the position and ranking a person that works in the university holds. For example it was used to describe if a person is an associate professor, a full professor, a lecturer, a dean etc. Furthermore, some of its properties were used to assign a person to the organization he works for, which in this case is the IHU, and to provide a person's email address. For the cases of the university and its departments the ontology's respective classes were used to create the relevant instances, while there were no actual properties used for these two cases since properties from other ontologies covered the conceptual spectrum.

---

<sup>9</sup> [http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/xcri\\_rdfs.xml](http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/xcri_rdfs.xml)

<sup>10</sup> <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl>

- *Portal Ontology*<sup>11</sup>: The Portal Ontology is a sub-ontology of *The AKT Reference Ontology*<sup>12</sup> which has been developed by the AKT partners to represent the knowledge used in the CS AKTive Portal testbed. It is the main ontology, describing people, projects, publications, geographical data, etc. It imports the AKT Reference Ontology<sup>13</sup> (Support Ontology) which is a top level ontology, providing basic definitions used by the portal ontology, including a simple representation of temporal objects, but in our case it was not used since it did not provide any fitting conceptual descriptions. The Portal Ontology was used in our application to represent that a Department (e.g. the School of Science and Technology) of the University is an organizational part of the University organization. Furthermore, it was used to describe properties of people that work in the university. Examples of such properties are the person's appellation, his gender and his research interests.
- Some properties from the *Dublin Core Ontology*<sup>14</sup> were also used, but since no instances of the ontology's classes were made it was not imported into the ontology, but the properties were used directly through the ontology's namespace. The title and description metadata properties were used in many of the ontology's classes in order to give a title and a general description to the specific instance of that class.

In general, we tried to make use of the above ontologies in order to describe concepts concerning the IHU, while in the process modifications and assertions were made to them in order to extract the best outcome. For example, in many cases a combination of the ontologies classes was made in order to describe a concept that could not be entirely described by a single ontology. Another important point worth mentioning is that the selection of the above ontologies was made not only by taking into account their recognition by the semantic community, but by how well they integrated together in order to provide a new ontology that would be well formed and consistent. Attempts were made to use other ontologies that may have been more popular, but due to inconsistencies they were discarded.

---

<sup>11</sup> <http://www.aktors.org/ontology/portal>

<sup>12</sup> <http://www.aktors.org/publications/ontology/>

<sup>13</sup> <http://www.aktors.org/ontology/support>

<sup>14</sup> <http://dublincore.org/documents/dces/>

## *A Deeper Insight*

At this point we are going to take a deeper look and explain the way the IHU Ontology was constructed by reusing the above mentioned ontologies and by applying modifications and extensions. We are going to present the classes and properties that were reused; the various combinations of classes that were applied and the extensions that were added in cases where the reused ontologies did not provide any descriptions. We should note that the final Ontology can be viewed/accessed/downloaded for further use/extension from the following URL: <http://lod.ihu.edu.gr/ontologies/v2/ihu-academic.owl>, while Figure 32 shows the HTML page<sup>15</sup> that gives a general description of the Academic Ontology.

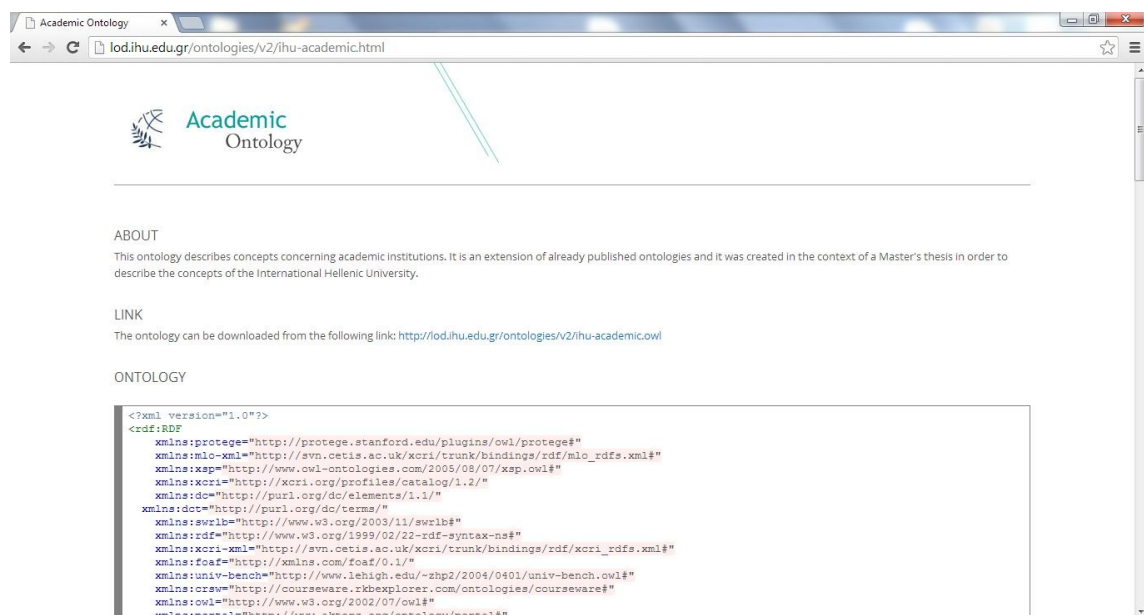


Figure 32: The Academic Ontology webpage

In Table 1 we can see the classes that were used from each of the ontologies presented above. The most used ontology concerning its classes was the LUMB ontology, since 14 classes were used to describe 14 different concepts. Some of the classes were not used directly though, but through inference rules that were applied to the concepts mainly through inheritance. From the ReSIST Courseware Ontology 2 classes were used, while from the MLO, XCRI-CAP and Portal ontologies 2, 3, 1 and 2 classes were used respectively. The ReIST Courseware Ontology was used in combination with the MLO and XCRI-CAP Ontologies to describe Courses and their Assessment methods. The MLO was used to describe the Programs the IHU provides, which was made through the use of the Learning Opportunity Instance and Learning

<sup>15</sup> <http://lod.ihu.edu.gr/ontologies/v2/ihu-academic.html>

Opportunity Specification concepts it provided. Also the Learning Opportunity provider concept was used in order to describe the International Hellenic University itself and its Departments. From the XCRI-CAP ontology the Resource concept was used in order to describe more extensively the Courses and the Programs the university offers, while it was also used to describe the Qualification a Program offers to its students. The LUMB ontology in combination with the Portal and ReSIST Courseware Ontologies was mainly used to describe the People who work at the IHU and the ranking each one of them holds (e.g. Professor, Associate Professor etc.). Furthermore, the Schedule class provided by the LUMB ontology was used to assign to a Program the various courses it offers, while the Department and University classes were used in combination with the MLO provided classes in order to describe the respective concepts. The Portal ontology was used in combination with the LUMB ontology to describe the Organizations, which in our case were the University and its Departments, and the People working them.

Table 1: Classes used from the reused Ontologies

	<b>ReSIST Courseware Ontology</b>	<b>MLO</b>	<b>XCRI-CAP</b>	<b>LUMB</b>	<b>Portal Ontology</b>
<b>Classes</b>	Course Assessment-Method	LearningOpportunityInstance LearningOpportunityProvider Learning Opportunity Specification	Resource	Department University Person Dean Employee Faculty Lecturer Professor Assistant Professor Associate Professor FullProfessor Schedule	Organization Person

				Organization	
--	--	--	--	--------------	--

In Table 2 we can see the datatype and object properties that were used from each Ontology. We can see that 8 properties were used from the ReIST Courseware Ontology, 9 properties from the MLO ontology, 2 properties from the XCRI-CAP Ontology, 2 properties from the LUMB ontology and 5 properties from the Portal Ontology.

Table 2: Properties used from the reused Ontologies

		ReSIST Courseware Ontology	MLO	XCRI-CAP	LUBM	Portal Ontology
<b>Properties</b>	<b>D a t a t y p e</b>	course-objectives	start	studyMode		full-name
		detailed-description	level languageOfInstruction	learningOutcome		
		is-taught-present	engagement duration			
		number-of-credits	cost			
		total-hours-engagement				
	<b>O b j e c t</b>	has-instructor	offeredAt		listedCourse	organization-part-of
		has-assessment-method	qualification		worksFor	has-research-interest
		taught-at	offers			has-appellation
						has-gender

--	--	--	--	--	--

In Table 3 we can see a summary of the classes and properties that were created and were used in conjunction with the ones provided by the reused ontologies in order to describe in more detail the concepts of the IHU. The Academic Ontology consists of 5 classes, 3 datatype properties and 5 object properties. The is-specialization-stream-of property is the inverse of the has-specialization-stream property, while the is-course-of-specialization-stream property is the inverse property of the specialization-stream-course property. Later on, we will see in the provided diagrams the relationships between Academic Ontology's classes and the classes from the reused ontologies, so that the properties of interest could be used in our ontology in order to describe certain concepts.

Table 3: Concepts and properties from the Academic Ontology

		<b>Academic Ontology</b>
<b>Classes</b>		Qualification
		Program
		AcademicCoordinator
		AcademicAssistant
		SpecializationStream
<b>Properties</b>	<b>Datatype</b>	course-term
		course-type
		faculty-type
	<b>Object</b>	is-instructor-of
		has-specialization-stream
		is-course-of-specialization-stream

		is-specialization- stream-of
		specialization- stream-course

In Table 4 we can see the domain and range of each property. In order to save space the following namespace prefixes are used for each ontology:

- ReSIST Courseware Ontology -> crsw
- MLO -> mlo
- XCRI-CAP -> xcri
- LUMB -> univ-bench
- Portal Ontology -> portal
- Academic Ontology -> ihu-owl
- Dublin Core Ontology -> dc

We can observe that most properties have a more generalized domain and range (the classes are in the higher level of hierarchy) which allowed us through inheritance and equivalence rules to make available some of these properties to our own classes through the various extensions we made.

Table 4: Domain and range of properties

	Property	Domain	Range
<b>Datatype</b>	crsw:course-objectives	crsw:Course	xsd:String
	crsw:detailed-description	crsw:Course	xsd:String
	crsw:is-taught-present	crsw:Course	xsd:String
	crsw:number-of-credits	crsw:Course	xsd:float
	crsw:total-hours-engagement	crsw:Course	xsd:float
	mlo:start	rdfs:Resource	xsd:String
	mlo:level	rdfs:Resource	xsd:String
	mlo:languageOfInstruction	rdfs:Resource	xsd:String
	mlo:engagement	rdfs:Resource	xsd:String



	<p>mlo:duration</p> <p>xcri:studyMode</p> <p>xcri:learningOutcome</p> <p>portal:full-name</p>	<p>rdfs:Resource</p> <p>xcri:presentation</p> <p>rdfs:Resource</p> <p>portal:Person</p>	<p>xsd:String</p> <p>xsd:String</p> <p>xsd:String</p> <p>xsd:String</p>
<b>Object</b>	<p>crsw:has-instructor</p> <p>crsw:has-assessment-method</p> <p>crsw:taught-at</p> <p>mlo:offeredAt</p> <p>mlo:qualification</p> <p>mlo:offers</p> <p>univ-bench:listedCourse</p> <p>univ-bench:worksFor</p> <p>portal:organization-part-of</p> <p>portal:has-research-interest</p> <p>portal:has-appellation</p> <p>portal:has-gender</p> <p>crsw:is-instructor-of</p> <p>ihu-owl:has-specialization-stream</p> <p>ihu-owl:is-course-of-specialization-stream</p> <p>ihu-owl:is-specialization-stream-of</p> <p>ihu-owl:specialization-stream-course</p> <p>ihu-owl:course-term</p>	<p>crsw:Course</p> <p>crsw:Course</p> <p>crsw:Course</p> <p>crsw:Course</p> <p>rdfs:Resource</p> <p>rdfs:Resource</p> <p>univ-bench:Schedule</p> <p>owl:Thing</p> <p>owl:Thing</p> <p>portal:Organization, portal:Person, portal:Organization-Unit</p> <p>portal:Person</p> <p>portal:Person</p> <p>portal:Person</p> <p>ihu-owl:Program</p> <p>crsw:Course</p> <p>ihu-owl:Specialization-Stream</p> <p>ihu-owl:Specialization-Stream</p> <p>ihu-owl:Specialization-Stream</p> <p>ihu-owl:Specialization-Stream</p> <p>crsw:Course</p>	<p>portal:Person</p> <p>crsw:Assessment-Method</p> <p>portal:Organization</p> <p>mlo:LearningOpportunityProvider</p> <p>rdfs:Resource</p> <p>mlo:LearningOpportunitySpecification</p> <p>univ-bench:Course</p> <p>owl:Thing</p> <p>portal:Organization</p> <p>portal:Research-Area</p> <p>portal:Appellation</p> <p>portal:Gender</p> <p>crsw:Course</p> <p>ihu-owl:SpecializationStream</p> <p>ihu-owl:SpecializationStream</p> <p>ihu-owl:Program</p> <p>crsw:Course</p> <p>xsd:String</p>

	ihu-owl:course-type	crsw:Course	xsd:String
	ihu-owl:faculty-type	univ-bench:Person	xsd:String
	mlo:cost	rdfs:Resource	xsd:String

In Table 5 we can see two things. Firstly, we can see the classes that were directly used in order to create the respective instances. This means that we only show the instance types (e.g. crws:Course) that were directly asserted by us and not the inferred instance types. Secondly, we see the properties that were used for each class instance. Although these properties are listed for each respective instance type it does not mean that they only have as their domain that certain class. Most of the properties are made available to the class through inheritance from higher level classes that either were added by us or were already present by the initial ontology that was imported. The exact domain of each property is shown in Table 4. Furthermore, we only show the properties that were used by us to describe each concept although more properties may have been available for use for each instance type. Generally, we can see that the instance type with the largest number of properties used is the crsw:Course with maximum 15 properties in use for each instance. We can also observe from the table that all the classes that are used to describe people and their rankings (Faculty, Lecturer, AssociateProfessor, AssistantProfessor, FullProfessor, AcademicAssistant, Dean, AcademicCoordinator) have almost the same properties. This is due to inheritance, since we can see that most properties have a different namespace than the class's namespace, which means that they were inherited by higher level classes (portal:Person, univ-bench:Person etc.). Later on in this chapter, we are going to show examples of instances created for each class to better understand the concepts shown in the tables.

Table 5: Properties used for each class instance

<b>Classes</b>	<b>Properties</b>				
crsw:Course	crsw:course-objectives	crsw:detailed-description	crsw:is-taught-present	crsw:number-of-credits	crsw:total-hours-engagement
	crsw:has-instructor	crsw:has-assessment-method	crsw:taught-at	mlo:offeredAt	xcrl:learningOutcome
	ihu-owl:course-	ihu-owl:course-	ihu-owl:is-course-of-	dc:title	

	term	type	specialization-stream		
ihu-owl:Program	ihu-owl:has-specialization-stream	mlo:qualification	univ-bench:listedCourse	mlo:cost	mlo:duration
	mlo:languageOfInstruction	mlo:engagement	mlo:start	mlo:level	xcri:studyMode
	dc:title				
ihu-owl:SpecializationStream	ihu-owl:is-specialization-stream-of	ihu-owl:specialization-stream-course	dc:title		
ihu-owl:Qualification	dc:title	owl:sameAs			
univ-bench:Department	portal:organization-part-of	mlo:offers	dc:title	dc:description	
univ-bench:University	mlo:offers	dc:title	dc:description	owl:sameAs	
univ-bench:Faculty	ihu-owl:faculty-type	portal:full-name	univ-bench:worksFor	portal:has-research-interest	portal:has-appellation
	portal:has-gender	crsw:is-instructor-of			
univ-bench:Lecturer	ihu-owl:faculty-type	portal:full-name	univ-bench:worksFor	portal:has-research-interest	portal:has-appellation
	portal:has-gender	crsw:is-instructor-of			
univ-bench:AssociateProfessor	ihu-owl:faculty-type	portal:full-name	univ-bench:worksFor	portal:has-research-interest	portal:has-appellation
	portal:has-gender	crsw:is-instructor-of			
univ-bench:AssistantProfessor	ihu-owl:faculty-type	portal:full-name	univ-bench:worksFor	portal:has-research-interest	portal:has-appellation
	portal:has-gender	crsw:is-instructor-of			
univ-bench:FullProfessor	ihu-owl:faculty-type	portal:full-name	univ-bench:worksFor	portal:has-research-interest	portal:has-appellation
	portal:has-gender	crsw:is-instructor-of			
ihu-	ihu-	portal:full-name	univ-bench:work	portal:has-research-	portal:has-appellation

owl:AcademicAssistant	owl:faculty-type		sFor	interest	
	portal:has-gender	crsw:is-instructor-of			
univ-bench:Dean	ihu-owl:faculty-type	portal:full-name	univ-bench:work sFor	portal:has-research-interest	portal:has-appellation
	portal:has-gender	crsw:is-instructor-of			
ihu-owl:AcademicCoordinator	ihu-owl:faculty-type	portal:full-name	univ-bench:work sFor	portal:has-research-interest	portal:has-appellation
	portal:has-gender				
crsw:Assesment-Method	owl:sameAs				

In Figure 33 we can see the complete ontology that was used in order to describe the concepts of the International Hellenic University. The arrows that are shown represent the inheritance relationships between the classes. This is the complete ontology including the various extensions and modification in the relationships made by us. For example, we can see that the classes we created (with the ihu-owl namespace prefix) are integrated into the class hierarchy so that the concepts of interest could be described in more detail. In the below diagrams we are going to show the extensions that were made by our part mainly by using the `rdfs:subClassOf` and `owl:equivalentClass` properties in order to create a well-integrated ontology. By applying these modifications we managed to provide to each class of interest the properties that are shown in Table 5.

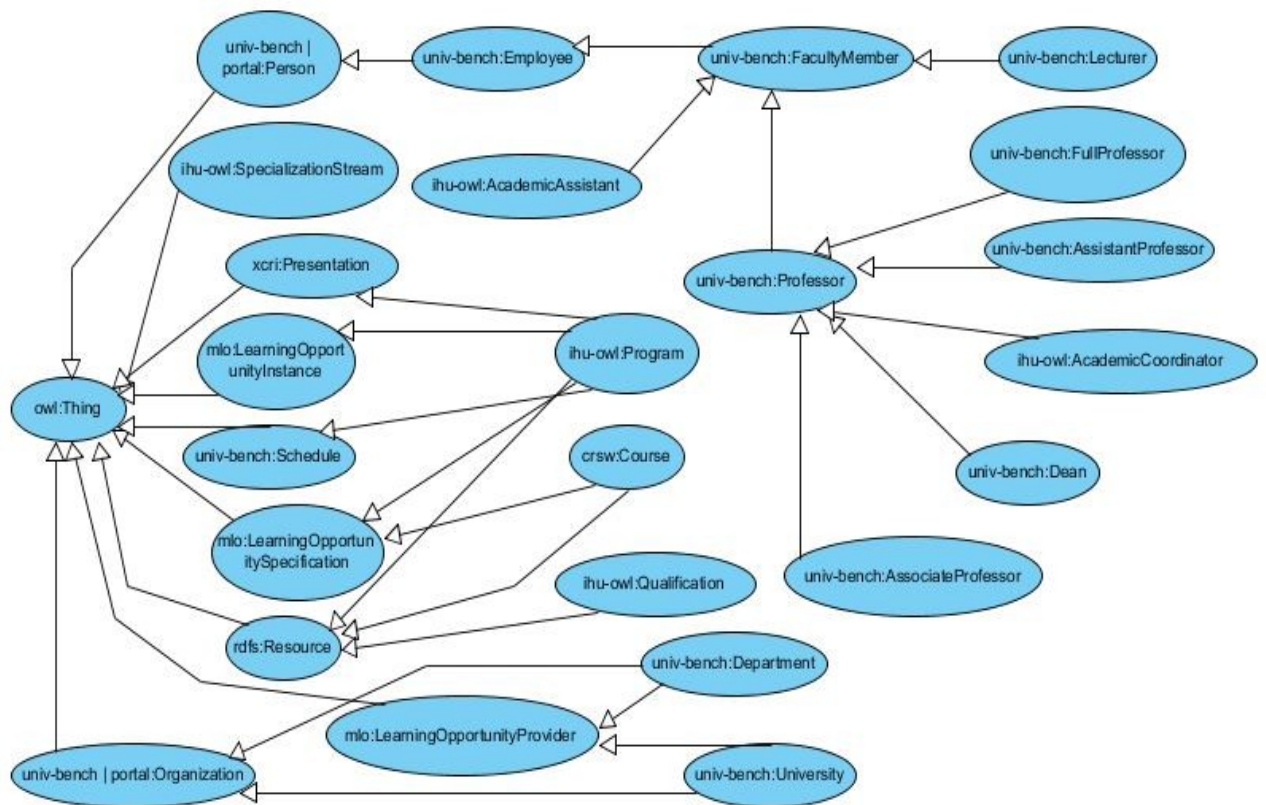


Figure 33: Inheritance relationships between classes

In Figure 35 we can see two of the extensions that we made to the LUMB ontology. Since it did not provide classes that described the Academic Coordinator and Academic Assistant roles in the university ranking structure, we decided to create them and integrate them into the ontology by adding them as subclasses of the univ-bench:Professor and univ-bench:Faculty classes respectively, which were already provided. This way the ihu-owl:AcademicCoordinator and ihu-owl:AcademicAssistant classes inherit all of the properties provided by the classes that are higher in the hierarchy and also relevant to the classes properties can be added. Although in our case no extra properties were added directly to the classes the ihu-owl:faculty-type property that was created and was assigned in a higher level class was inherited to these classes also. The code that led to the creation of these relationships is shown in Figure 34.

```
<owl:Class rdf:about="AcademicAssistant">
  <rdfs:subClassOf rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Faculty"/>
</owl:Class>
<owl:Class rdf:about="AcademicCoordinator">
  <rdfs:subClassOf rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Professor"/>
</owl:Class>
```

Figure 34

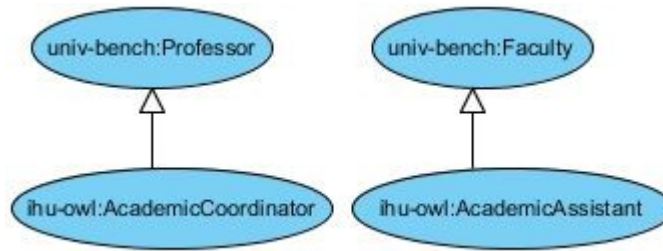


Figure 35

In Figure 36 we can see a different kind of relationship created between two similar classes of two different ontologies, while in Figure 37 we can see the code that generated this relationship. The owl:equivalentClass property was used between the portal:Organization and univ-bench:Organization classes in order to show that these two classes represented in two different ontologies are actually referring to the same thing, which is an Organization. This way the instances created from that level and below in the class hierarchy can use the properties provided by both classes, allowing us to give a more detailed and informative description of the instance.

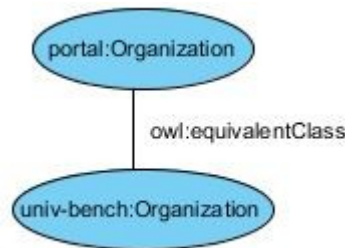


Figure 36

```
<owl:Class rdf:about="http://www.aktors.org/ontology/portal#Organization">
  <owl:equivalentClass rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Organization"/>
</owl:Class>
```

Figure 37

Figure 38 shows the ihu-owl:SpecializationStream class that we created in order to describe the specialization streams the MSc in Energy Systems provides. We can observe that it is not actually connected directly to the ontology by some kind of inheritance or equivalence relationship, but it is connected by setting it as a domain or range class for some properties that we created. More specifically, it was set as the range for the ihu-owl:has-specialization-stream property that we created in order to show the specialization streams a Program offers. Also the inverse of this property - the ihu-owl:is-specialization-stream-of - was created and the ihu-owl:SpecializationStream

class was set as the domain of that property in order to show the Program to which the specializations stream belongs to.



Figure 38

In Figure 39 we can see that the ihu-owl:Qualification class that we created in order to show the qualification a Program of the university offers, was set as a subclass of the rdfs:Resource class. The rdfs:Resource class is used mainly by the MLO and the XCRI-CAP ontologies and it is used by being set as the domain or range of most of its provided properties. So the reason why many of the classes we created were set as subclasses of the rdfs:Resource class was that we wanted them to inherit the relevant properties provided by these ontologies, always keeping in mind though the logical connection between the concepts.

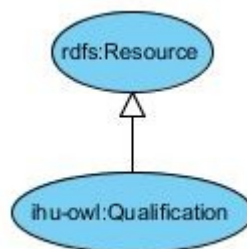


Figure 39

In Figure 40 we see the way we integrated the various concepts of four ontologies in order to provide an appropriate description for the Course concept. We decided to set the two relevant Course classes provided by the LUMB and ReSIST Courseware Ontologies as equivalent for the same reason as we did for the Organization classes from two other ontologies. We also decided to set the Course class as a subclass of the mlo:LearningOpportunitySpecification and rdfs:Resource classes since we wanted to use some of the properties that used these two classes as their domain or range. Although this could also had been done by setting them as equivalent, we thought that it was more appropriate to use the rdfs:subclassOf property since a course is a more specific concept than the concepts described by its superclasses. These modifications allowed us to use the MLO and XCRI-CAP properties shown in Table 5, to further describe the Course concept. On top of that, we also created two extra properties – the

ihu-owl:course-term and ihu-owl:course-type – in order to assign to each course the term it belongs to and its type (Required, Elective or Mandatory).

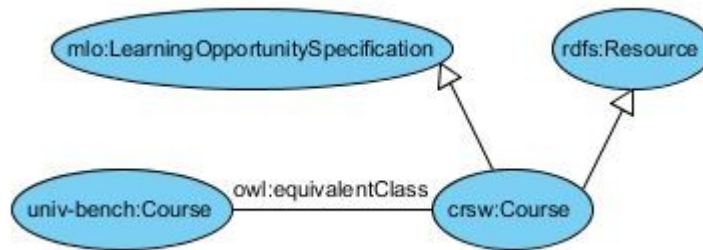


Figure 40

Figure 41 shows the equivalence that was set between the two Person classes from the LUMB and Portal ontologies, while Figure 42 shows that the univ-bench:University class was set as a subclass of the mlo:LearningOpportunityProvider class. The later was done so that we could use the mlo:offers property provided by the MLO ontology in every University instance that was created. This way the user or machine that would traverse the ontology could see through the University instance the various Programs it offers. In the case of the Person concept, except from the properties that were provided by the two ontologies (Portal and LUMB) we also needed a property which showed if a Person is a Faculty, a Visiting Faculty or an Adjunct Faculty member. Since this kind of property did not exist we created the ihu-owl:faculty-type property and we gave it as domain the Person class. This way all subclass instances of the Person class (e.g. univ-bench:AssociateProfessor, uni-bench: AssistantProfessor etc.) would inherit this property and use it in order to describe what kind of faculty type each person represents.

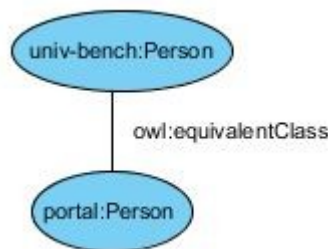


Figure 41



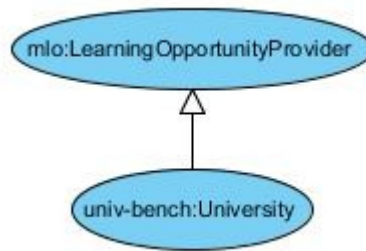


Figure 42

Figure 43 shows the multiple inheritance rules that were set for the Program concept. Since there was no actual class provided by any of the used ontologies, we created the Program class and by integrating it with concepts already provided by them we managed to describe the concept by mainly reusing the provided properties. Since we could not describe everything concerning this concept through the provided properties we created an extra property – the ihu-owl:has-specialization-stream property – and we set the ihu-owl:Program class as its domain. The rest of the properties that were inherited by the classes shown in the diagram can be viewed with each ontology’s prefix in Table 5.

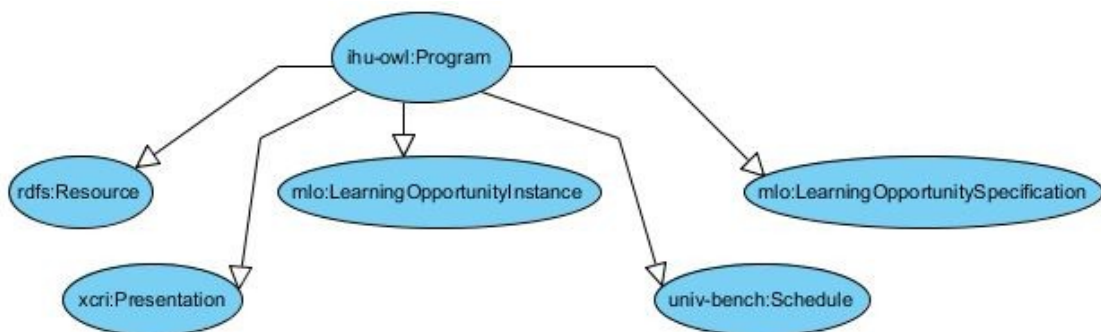


Figure 43

Finally, Figure 44 shows that we combined the two concepts from the LUMB and MLO ontologies in order to describe the Departments of the university. This way we could once again use the mlo:offers property provided by the MLO ontology in order to show the Programs each Department offers. Furthermore, the equivalence that was set in a higher level between the two Organization classes (department is by default a subclass of the Organization class as shown in Figure 33) of the Portal and LUMB ontologies led to the inheritance of the portal:organization-part-of property which tells that each Department is an organizational part of the IHU.

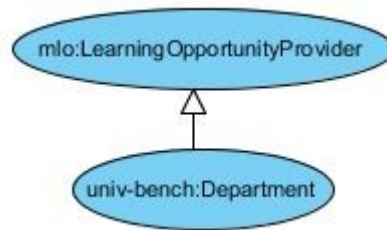


Figure 44

### 4.1.3 The Academic Ontology’s Instance Level

In this subchapter we are going to provide a presentation of the Academic Ontology’s instance level. We are going to present an indicative instance of most of the used classes including the properties that were used in order to provide a detailed description of each concept. More specifically we are going to present an instance for each of the following classes:

- Course
- Program
- University
- Department
- Dean
- AssociateProfessor
- SpecializationStream

We are not going to present the rest of the Person subclasses for which we created instances for, since they are almost identical to the AssociateProfessor class having as their only difference the `rdf:type` property. Also we are not going to present the `crsw:Assessment-Method` and `ihu-owl:Qualification` classes since they do not have any interesting properties except from the `rdf:type`, `dc:title` and `owl:sameAs` properties, with the last one connecting the instances with the respective instance of the dbpedia dataset. To avoid confusion we should also state that classes-properties and instances have different namespace prefixes which are shown in Table 6.

Table 6: URIs of instances and classes-properties

	Prefix	URI
<b>Instances</b>	ihu	<a href="http://lod.ihu.edu.gr:8080/academic/id/">http://lod.ihu.edu.gr:8080/academic/id/</a>
<b>Classes-Properties</b>	ihu-owl	<a href="http://lod.ihu.edu.gr:8080/academic/def/">http://lod.ihu.edu.gr:8080/academic/def/</a>

In Table 7 we can view the ihu:KnowledgeManagement instance which is one of the courses that are offered by the International Hellenic University. We can see that the properties that are used come from different ontologies and they are used in order to describe in detail this concept. There are properties describing the term the course is offered in, the type of course, the course's learning outcome, its cost etc. The properties that are shown in the table that have values without prefixes assigned to them are datatype properties while the rest are object properties. For example the crsw:has-instructor property is an object property and it is a reference to the ihu:Nick\_Bassiliades instance which is of type univ-bench:AssociateProfessor and ihu:AcademicCoordinator. This way each course is connected with the Persons it is taught by.

Table 7: An indicative Course instance

Individual	Property	Value
ihu:Knowledge_management	ihu-owl:course-type	Elective
	ihu-owl:course-term	First
	mlo:offeredAt	ihu: School_of_Science_and_Technology
	mlo:offeredAt	ihu: International_Hellenic_University
	rdf:type	crsw:Course
	xcrl:learningOutcome	Comprehend web Knowledge...
	xcrl:learningOutcome	Acquire essentialAcquire...
	xcrl:learningOutcome	Experiment with creating their...
	crsw:detailed-description	Basic concepts of Knowledge and...
	crsw:course-objectives	This module examines basic...
	crsw:has-instructor	ihu:Nick_Bassiliades
	crsw:has-assessment-method	Exam
	crsw:has-assessment-method	Coursework
	crsw:total-hours-engagement	30.0
	dc:title	Knowledge management
	crsw:number-of-credits	6
	crsw:taught-at	Ihu: International_Hellenic_University

In Table 8 we can observe the ihu:Energy\_Systems instance which is an instance of type Program. We can see that each Program is connected with the university it is offered by

and the Department it belongs to. Also since this particular program had specialization streams and since no relevant properties were offered by any ontology, the ihu-owl:SpecializationStream property was created in order to connect these two concepts.

Table 8: An indicative Program instance

Individual	Property	Value
ihu:Energy_Systems	dc:title	M.Sc in Energy Systems
	mlo:cost	4000
	mlo:duration	14-26 months
	mlo:offeredAt	ihu: School_of_Science_and_Technology
	mlo:offeredAt	ihu: International_Hellenic_University
	rdf:type	ihu:Program
	mlo:level	Postgraduate
	mlo:languageOfInstruction	English
	mlo:qualification	ihu:M.Sc
	mlo:start	October
	xcri:study-mode	Part time
	xcri:study-mode	Full time
	ihu-owl:specialization-stream	ihu:Renewable_Energy
	ihu-owl:specialization-stream	ihu:Energy_Systems_Management

Table 9 shows the IHU instance and the information it carries. It contains a general description of the University and it is connected with the Programs it offers so that the user or machine can easily navigate from instance to instance in order to find the information he is looking for easily. Also the owl:sameAs property connects our dataset with an external dataset, the dataset of dbpedia. By using this property we state that the IHU instance has a similar instance that is stated in the dbpedia dataset. This is an important concept of the Semantic Web since one of the Linked Data principles refers to the interlinking of datasets in order to provide a global connected graph of individual datasets where similar concepts are explicitly stated to assist the cloud's exploration.

Table 9: The IHU instance

Individual	Property	Value
ihu:International_Hellenic_University	dc:title	International Hellenic University
	rdf:type	univ-bench:University
	dc:description	The International Hellenic University (IHU) was established in...
	mlo:offers	ihu:Transnational_and_European_Commercial_Law_and_Alternative_Dispute_Resolution
	mlo:offers	ihu:Law_and_Economy
	mlo:offers	ihu:Sustainable_Development

mlo:offers	ihu:Banking_and_Finance
mlo:offers	ihu:Executive_MBA
mlo:offers	ihu:Management
mlo:offers	ihu:Black_Sea_Cultural_Studies
mlo:offers	ihu:Strategic_Product_Design
mlo:offers	ihu:ICT_Systems
mlo:offers	ihu:Energy_Systems
owl:sameAs	dbpedia:International_Hellenic_University

Table 10 shows the ihu:School\_of\_Science\_and\_Technology instance which is a Department instance. The instances of this type provide a general description of the Department, while they are also connected with the Programs that belong to the specific Department.

Table 10: An indicative Department instance

Individual	Property	Value
ihu:School_of_Science_and_Technology	dc:title	International Hellenic University
	rdf:type	univ-bench:University
	dc:description	Our strategic mission is to develop teaching programmes...
	mlo:offers	ihu:Transnational_and_European_Commercial_Law_and_Alternative_Dispute_Resolution
	mlo:offers	ihu:Law_and_Economy
	mlo:offers	ihu:Sustainable_Development
	mlo:offers	ihu:Banking_and_Finance
	mlo:offers	ihu:Executive_MBA
	mlo:offers	ihu:Management
	mlo:offers	ihu:Black_Sea_Cultural_Studies
	mlo:offers	ihu:Strategic_Product_Design
	mlo:offers	ihu:ICT_Systems
	mlo:offers	ihu:Energy_Systems

Table 11 and Table 12 describe two instances referring to people working at the IHU. The ihu:Ioannis\_Vlahavas instance is a Dean instance and it provides information about the appellation, gender and research interests of the person. It also states the University the Dean works for by using the univ-bench:worksFor property which is actually used for all Person instances. Table 12 describes the AssociateProfessor instance which has the same properties as the Dean instances (and with the rest of Person instances), with the difference that the former provides a connection between the Associate professor and the Courses he teaches. This is done by using the crsw:is-instructor-of property which has as its range the Course class.

Table 11: An indicative Dean instance

Individual	Property	Value
ihu:Ioannis_Vlahavas	portal:full-name	Professor Ioannis Vlahavas
	rdf:type	univ-bench:Dean
	portal:has-apellation	portal:Prof
	portal:has-gender	portal:Male-Gender
	portal:has-research-interest	portal:Artificial-Intelligence-Research-Area
	univ-bench:worksFor	ihu:International_Hellenic_University
	ihu-owl:faculty-type	IHU Faculty

Table 12: An indicative AssociateProfessor instance

Individual	Property	Value
ihu:Nick_Bassiliades	portal:full-name	Dr. Nikolaos Bassiliades
	rdf:type	univ-bench:AssociateProfessor
	rdf:type	ihu:AcademicCoordinator
	portal:has-apellation	portal:Dr
	portal:has-gender	portal:Male-Gender
	portal:has-research-interest	portal:Artificial-Intelligence-Research-Area
	portal:has-research-interest	portal:Ontologies
	portal:has-research-interest	portal:Knowledge-Management
	portal:has-research-interest	portal:Semantic-Web-Area
	univ-bench:worksFor	ihu:International_Hellenic_University
	ihu-owl:faculty-type	IHU Faculty
	crsw:is-instructor-of	ihu:Knowledge_management

Table 13 shows the ihu:Energy\_Systems\_Management instance which is an instance of the ihu-owl:SpecializationStream class. This class was created by us in order to connect the Programs with the specialization streams they offer and each course with the specialization stream it belongs to. For example in Table 13 we can see that the instance is connected with the Program it belongs to through the ihu-owl:is-specialization-stream-of property while it is also connected with the courses belonging to this particular specialization stream with the ihu-owl:specialization-stream-course property. These properties were also created by us for the reasons mentioned above.

Table 13: An indicative SpecializationStream instance

Individual	Property	Value
ihu:Energy_Systems_Manag	dc:title	Energy Systems Management

ement	rdf:type	ihu:SpecializationStream
	ihu-owl:is-specialization-stream-of	ihu:Energy_Systems
	ihu-owl:specialization-stream-course	ihu:Emission_Allowance_Markets
	ihu-owl:specialization-stream-course	ihu:Fossil_Fuels_and_Renewable_Energy_Economics
	ihu-owl:specialization-stream-course	ihu: Energy_Markets_Trading_and_Risk_Management
	ihu-owl:specialization-stream-course	ihu: Environmental_Law_and_Policy_for_Natural_Resources_and_Energy
	ihu-owl:specialization-stream-course	ihu: Energy_Efficiency_and_Savings
	ihu-owl:specialization-stream-course	ihu:Energy_Transportation

Finally, Table 14 shows some statistics concerning the dataset at its current state.

Table 14: Dataset Statistics

<b>Triples</b>	1421
<b>Classes</b>	17
<b>Entities</b>	120
<b>Distinct Subjects</b>	120
<b>Properties</b>	40
<b>Distinct Objects</b>	563

### *Utilizing RDFa snippets*

In this part we are going to provide a workflow of how someone could exploit our generated RDF data in order to enrich IHU's webpages with semantically rich metadata. The transformation of RDF data to RDFa snippets is quite easy, since there are quite a few tools that can do that automatically. Each snippet should be separated though in order to selectively be embedded into the appropriate webpages that they are associated with. In Figure 45 we can see a generated RDFa snippet, which describes the Knowledge Management course offered by the university. This snippet is ready to be embedded into a webpage in order to serve more descriptive information about its content, to search engines and various other software agents. Since the IHU website uses the Joomla CMS to deliver its content, we are going to give a brief guide of how this snippet could be embedded into the website through the CMS. After the RDFa snippet is obtained, the articles the CMS offers could be utilized. More specifically, this snippet could be embedded into the respective article that presents the Knowledge Management course. This is shown in Figure 46, where we tried to load a Custom

HTML module, containing the RDFa snippet, into the article. As we can observe from the article preview and from the actual generated page containing the article (Figure 47), the RDFa snippet is not visible to the end user but it is actually served along with the HTML content. Since the snippet is contained into the Custom HTML module the CMS offers, it could also be embedded in whatever part of the website through this module. In order for the page to be valid, the new tags that describe the content in RDFa format need to have a valid vocabulary. All we have to do to achieve this is replace the page's DOCTYPE with the one shown in Figure 48. In order for this change to take affect over the whole Joomla website the new DOCTYPE needs to be placed in the template that is used. After these changes are made the served webpages that contain this article (or Custom HTML module) will be semantically enriched with the RDFa snippet. In order to confirm that the served XHTML+RDFa content was valid we went the test webpage through a validation check whose results can be viewed in Figure 49.



```

<div xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns="http://www.w3.org/1999/xhtml"
xmlns:crsw="http://courseware.rkbexplorer.com/ontologies/courseware#"
xmlns:xcri-xml="http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/xcri_rdfs.xml#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:portal="http://www.aktors.org/ontology/portal#"
xmlns:swrl="http://www.w3.org/2003/11/swrl#"
xmlns:mlo="http://purl.org/net/mlo/"
xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"
xmlns:base="http://lod.ihu.edu.gr:8080/academic/id/"
xmlns:mlo-xml="http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/mlo_rdfs.xml#"
xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:xcri="http://xcri.org/profiles/catalog/1.2/"
xmlns:ihu="http://lod.ihu.edu.gr:8080/academic/def/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:univ-bench="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#"
xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
class="rdf2rdfa">
<div class="description" about="http://lod.ihu.edu.gr:8080/academic/id/"
typeof="owl:Ontology">
<div rel="owl:imports"
resource="http://lod.ihu.edu.gr/ontologies/v2/ihu-academic.owl"/>
</div>
<div class="description"
about="http://lod.ihu.edu.gr:8080/academic/id/Knowledge_management"
typeof="crsw:Course">
<div property="crsw:course-objectives"
content="This module examines basic concepts of Knowledge and Knowledge Manag
datatype="xsd:string"/>
<div property="ihu:course-term" content="First" datatype="xsd:string"/>
<div property="ihu:course-type" content="Elective" datatype="xsd:string"/>
<div property="crsw:detailed-description"
content="Basic concepts of Knowledge and Knowledge Management. Architecture and
datatype="xsd:string"/>
<div rel="crsw:has-assessment-method"
resource="http://lod.ihu.edu.gr:8080/academic/id/Coursework"/>
<div rel="crsw:has-assessment-method"
resource="http://lod.ihu.edu.gr:8080/academic/id/Exam"/>
<div rel="crsw:has-instructor"
resource="http://lod.ihu.edu.gr:8080/academic/id/Nick_Bassiliades"/>
<div property="xcri:learningOutcome"
content="Experiment with creating their Knowledge Management systems through a
datatype="xsd:string"/>
<div property="xcri:learningOutcome"
content="Comprehend web Knowledge Management languages and technologies, includ
datatype="xsd:string"/>
<div property="xcri:learningOutcome"
content="Acquire essentialAcquire essential skills on Knowledge Management Syst
datatype="xsd:string"/>
<div property="crsw:number-of-credits" content="6.0" datatype="xsd:float"/>
<div rel="mlo:offeredAt"
resource="http://lod.ihu.edu.gr:8080/academic/id/International_Hellenic_Univers
<div rel="mlo:offeredAt"
resource="http://lod.ihu.edu.gr:8080/academic/id/School_of_Science_and Technolo
<div rel="crsw:taught-at"
resource="http://lod.ihu.edu.gr:8080/academic/id/International_Hellenic_Univers
<div property="dc:title" content="Knowledge management" datatype="xsd:string"/>
<div property="crsw:total-hours-engagement" content="30.0" datatype="xsd:float"/>
</div>
</div>

```

Figure 45: RDFa instance

The screenshot shows a web editor interface for an article titled "Knowledge Management". On the left, there is a sidebar with a "Title" field containing "Knowledge Management" and an "Alias" field containing "knowledge-management". Below this is a "Section" field and a code editor showing HTML/RDFa markup. The main content area displays the rendered article, including the title "Knowledge Management", the instructor "Dr Nikolaos Bassiliades", and a table for "Teaching Hours and Credit Allocation" showing "30 Hours, 6 Credits" and "Module Assessment: Exam & Coursework". Below the table is an "Aims" section with a paragraph of text describing the module's focus on knowledge concepts, systems, and tools.

Figure 46

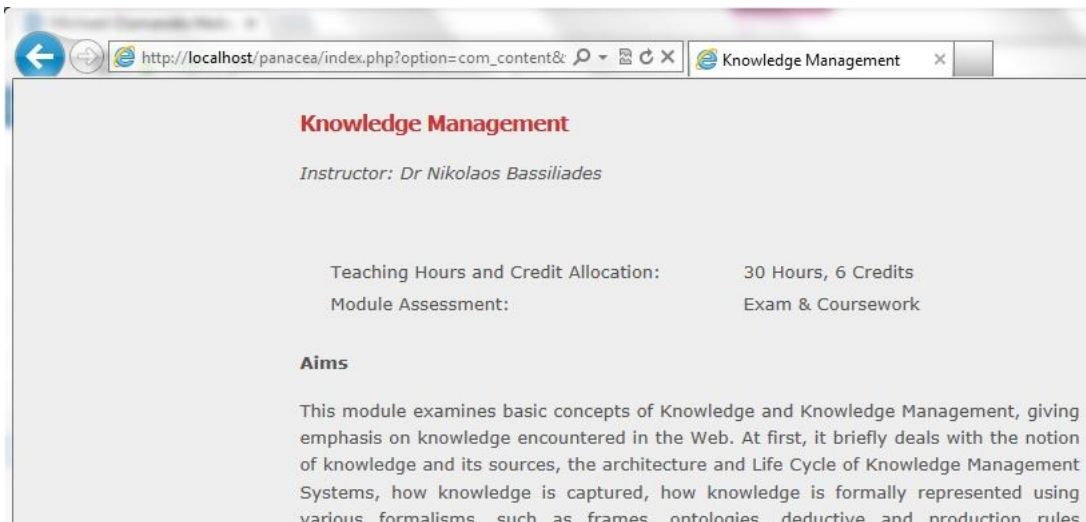


Figure 47

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
```

Figure 48

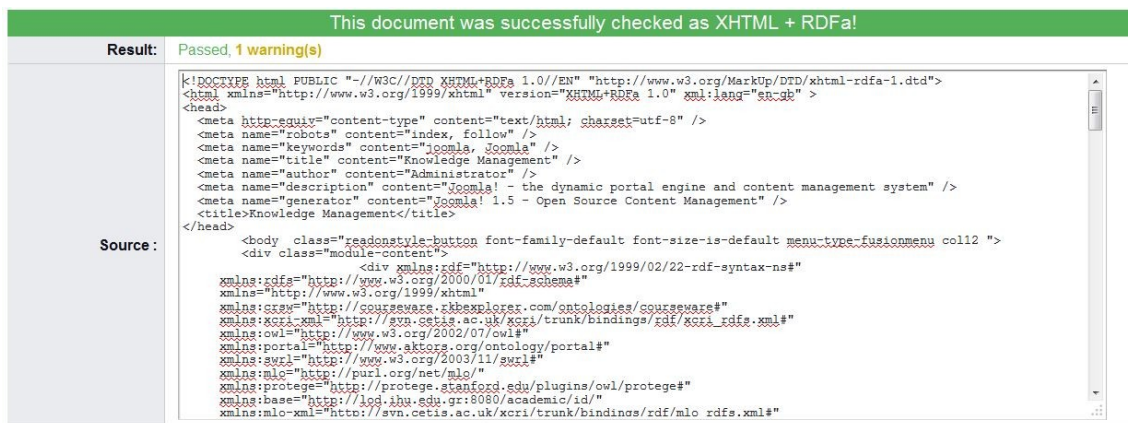


Figure 49

## 4.2 System Infrastructure and User interfaces

In this subchapter we are going to give an overview of the application infrastructure and we are also going to provide some examples of how someone could access the exposed data through the provided interfaces. We are firstly going to present the workflow that was followed in order to create the Academic Ontology and the instances on top of it, while later on some screenshots and examples are going to be presented in order to clarify the various ways an end user can easily access and exploit the dataset and the ontology.

### 4.2.1 The Workflow for Generating Linked Data

The workflow that was followed in order to create our Linked Data application had two phases which are the following:

1. The first phase concerned the creation of the conceptual layer on top of which later on we would create the instances that described the information concerning the IHU.
2. The second phase involved the creation of the instance layer and its exposure to the LOD cloud and to the Web, in each case with its respective format.

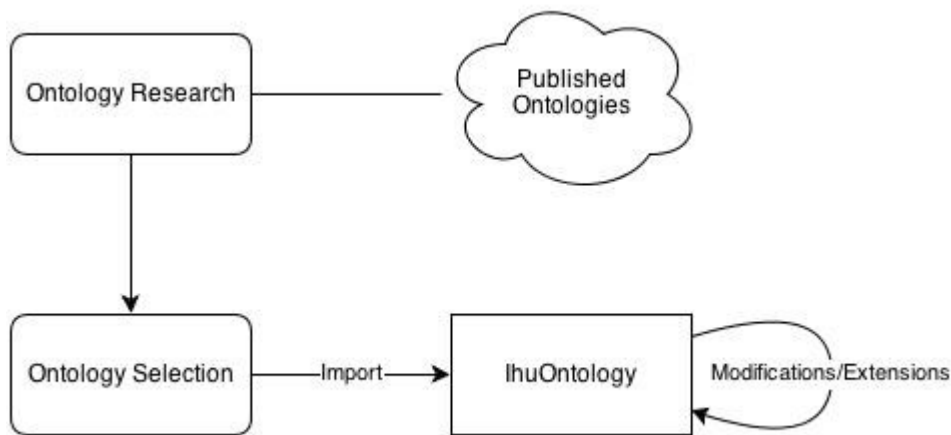


Figure 50: The Ontology creation process

In Figure 50 we can see the process that was followed in order to create the Academic Ontology, which describes the various concepts of the IHU. The first step of this phase involved an extensive research concerning the available ontologies that were already published in the cloud and were available for reuse. Since there was not a single ontology available that was able to describe all of the concepts that we were interested in, a combination of various ontologies had to be made in order to describe different domains of information. This was not a trivial task though, because although quite a few ontologies were available their integration was not always possible, since different combinations of ontologies led to inconsistencies. After the appropriate ontologies were selected they had to be imported into the Academic Ontology in order to create a single and concrete ontology that would describe the whole domain of the IHU. In this step we encountered quite a few problems relevant to ontology reuse and their integration [7] [8] which were discussed in the literature review section. Except of the inconsistencies that

were mentioned above, another problem was the domain coverage of each ontology and their actual integration. For example quite a few descriptions for concepts that we wanted to use were missing and this led to the creation of new descriptions which were integrated into the final ontology. Also the ontology integration was not done automatically by the tool as they were imported, but modifications had to be made in order to map similar concepts provided from different ontologies. After all these problems were resolved the Academic Ontology was ready for use and the second phase of the workflow took place.

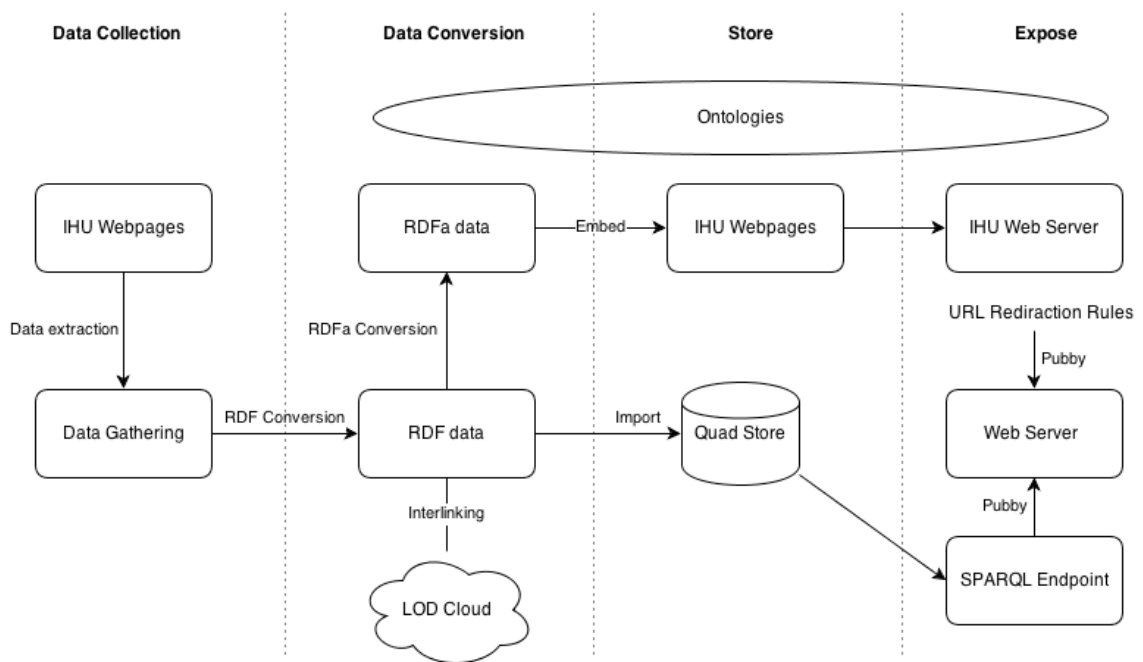


Figure 51: The workflow for generating Linked Data

The second phase of the workflow, which is the creation of the instance layer, can be seen in Figure 51 and is divided into four sub phases which are the following:

- Data collection
- Data conversion
- Data storage
- Data exposure

The data collection sub phase involved the process of extracting and gathering the information we wanted from the webpages of the IHU. Since the data that was present in the webpages was not in an appropriate format for publication it had to be gathered and passed to the second phase for further processing. Figure 52 and Figure 53 are screenshots from two of the pages that were used during this phase in order to gather

information. Figure 52 shows the way a course is displayed in the webpage, while Figure 53 shows among others the various schools/departments of the IHU, their programs and their specialization streams.

After the data was gathered from the relevant webpages it had to be converted into an appropriate format for publication. The first conversion that was made was into RDF which is the standard format for Linked Data publication. This conversion was made by creating RDF instances according to the conceptual layer that was created in phase one.

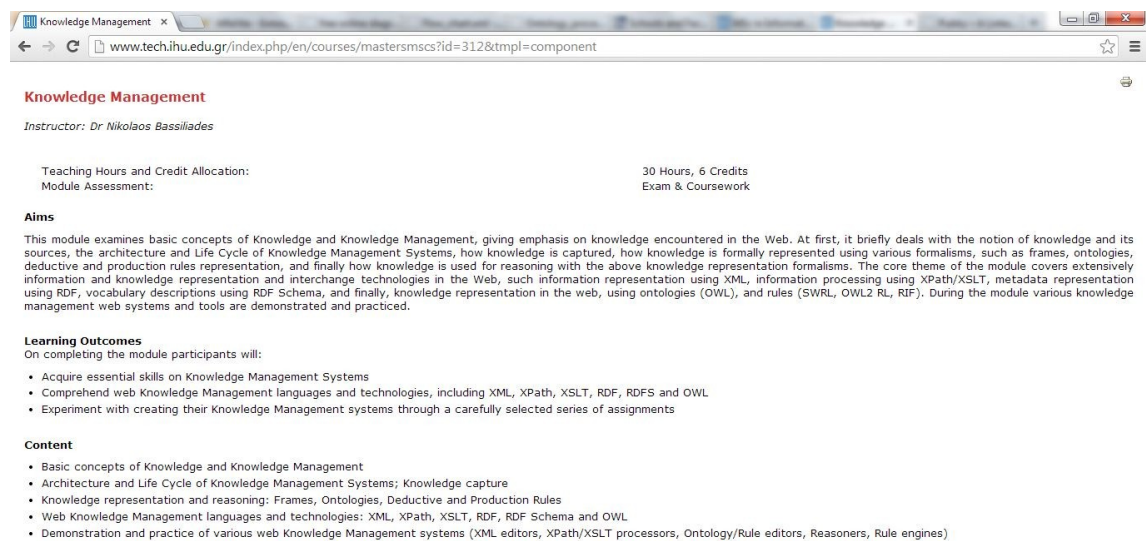


Figure 52: The IHU webpage for the Knowledge Management course

## Schools and Faculty

### Schools Faculty

The International Hellenic University comprises three Schools, each offering specialist programmes in cutting edge disciplines taught by leading instructors: the School of Economics & Business Administration, the School of Humanities and the School of Science & Technology. You are welcome to browse below and explore the different programmes on offer within each School:

- School of Economics and Business Administration
- School of Humanities
- ▾ School of Science and Technology Click to collapse

The **School of Science and Technology** offers the following postgraduate programmes:

- MSc in Information & Communication Technology Systems - Next intake October 2013
- MSc in Energy Systems - Next intake October 2013

Specialisation Streams:

- Energy Management Stream
- Renewable Energy Stream

Figure 53: The IHU's schools/departments and their offered programs

This way the data was not only created in an appropriate interoperable format but it was also interconnected so that it could easily be traversed and understood by crawlers and semantic agents, since the conceptual layer provided the semantics that are needed. Some of the instances were also interconnected with similar instances that were already published in other datasets in the LOD cloud. After the RDF data was ready, its conversion into the RDFa format was quite simple since there are quite a few tools available that can automatically do this task.

After the data was converted into RDF and RDFa they were ready to be stored. The storage of each data format was made in two different ways, since each format has a different purpose. The RDF data was stored into Virtuoso Quad Store<sup>16</sup> as triples, while the RDFa data snippets were embedded as semantically rich metadata into IHU's test HTML pages. As mentioned in the previous chapter only a test case of RDFa conversion was made in order to provide a methodology with which someone could embed the snippets into IHU's webpages. So the actual deployment of the RDFa snippets into IHU's web server was not actually implemented, but it is depicted in the figure to show the general methodology that should be followed.

The final sub phase involved the exposure of the dataset that was created, so that it could be exploited by users and machines. The Academic Ontology is also openly available through its URI, so that it can be reused and/or extended by any interested party. The three components shown in Figure 51 represent the three different ways the exposed data can be accessed and each of them is going to be described in the next chapter.

#### **4.2.2 Accessing IHU's Linked Data**

After the dataset was created we needed to provide a way so that it could be accessed by both users and machines. For users who do not have any particular technical knowledge a more user friendly interface had to be provided, while for machines the data had to be available in a raw RDF format so that it could easily be parsed and understood. Thereby we provided three different ways of access in the exposed data which are the following:

- SPARQL endpoint
- Web browser

---

<sup>16</sup> <http://virtuoso.openlinksw.com/>

- HTML snippets of RDFa (not currently available on a live server)

The first way the data can be explored is through the SPARQL endpoint Virtuoso Server provides for its triple store. This way a user or an application can make sophisticated queries to the SPARQL endpoint, which will return the relevant results from the quad store. Although technical knowledge is required by the user in order to access the dataset this way, it is the most appropriate if he requires exploring the dataset in depth and getting targeted results. Currently the dataset and the ontology can be queried through the following public SPARQL endpoint: <http://lod.ihu.edu.gr:8890/sparql>. Figure 54 and Figure 55 show two examples of how someone could run a query against the SPARQL endpoint in order to extract the data that is stored in Virtuoso quadstore. The first query shown in Figure 54 tries to retrieve all mandatory courses offered by the MSc in ICT Systems program. We can see that four courses were successfully returned as a result.

The screenshot shows a web interface for querying a SPARQL endpoint. At the top, the 'Default Graph IRI' is set to `http://lod.ihu.edu.gr:8890/ihu-academic-linked`. Below this is a 'Query' input area containing the following SPARQL query:

```
PREFIX univ-bench: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
PREFIX portal: <http://www.aktors.org/ontology/portal#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ihu-owl: <http://lod.ihu.edu.gr:8080/academic/def/>
PREFIX ihu: <http://lod.ihu.edu.gr:8080/academic/id/>

select ?title where
{
  ihu:ICT_Systems univ-bench:listedCourse ?s.
  ?s ihu-owl:course-type "Mandatory"^^xsd:string.
  ?s dc:title ?title.
}
```

Below the query input are four buttons: 'Execute', 'Save', 'Load', and 'Clear'. The 'Execute' button is highlighted. Below the buttons is a table displaying the results of the query:

title
"ICT Management and E-Business Strategy"^^<http://www.w3.org/2001/XMLSchema#string>
"Computer Networks"^^<http://www.w3.org/2001/XMLSchema#string>
"Web Information Systems"^^<http://www.w3.org/2001/XMLSchema#string>
"Decision support systems"^^<http://www.w3.org/2001/XMLSchema#string>

Figure 54: Querying the SPARQL endpoint

The second query shown in Figure 55 tries to retrieve all associate professors and the courses each one of them teaches. We can see that eight associate professors and nine courses are returned as a result, since one of them teaches two courses. These queries just scratch the surface of what can actually be done, since sophisticated queries can

return the exact information someone is searching for. For example, someone could search for courses that belong to programs that have a certain tuition fee threshold, or even search for courses that have a specific learning outcome and objective by providing some keywords relevant to their domain of interest.

The screenshot shows a SPARQL query interface. At the top, there are several PREFIX definitions:
 

```

    PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
    PREFIX univ-bench: <http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#>
    PREFIX cgm: <http://courseware.rkbexplorer.com/ontologies/courseware#>
    PREFIX portal: <http://www.aktors.org/ontology/portal#>
    PREFIX dc: <http://purl.org/dc/elements/1.1/>
    
```

 Below the PREFIX definitions is a SPARQL query:
 

```

    select ?name ?title where
    {
      ?p rdf:type univ-bench:AssociateProfessor.
      ?p portal:full-name ?name.
      ?p cgm:is-instructor-of ?c.
      ?c dc:title ?title.
    }
    
```

 At the bottom of the interface are buttons for 'Execute', 'Save', 'Load', and 'Clear'. Below the buttons is a table with two columns: 'name' and 'title'. The table contains ten rows of results, each showing a name and a corresponding title with their respective URIs.

name	title
"Dr. Efichios S. Sartzetakis" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"Introduction to Energy Economics and Quantitative Methods" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>
"Dr. Issac Meir" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"Energy Conversion Technologies" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>
"Dr. Issac Meir" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"Urban Energy Systems" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>
"Dr. Athanasia Pouloudi" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"ICT Management and E-Business Strategy" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>
"Dr. Maria Papadopoulou" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"Computer Networks" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>
"Dr. Nikolaos Bassiliades" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"Knowledge management" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>
"Dr. Athena Vakali" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"Web Information Systems" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>
"Dr. Michela Meo" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"Green ICT" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>
"Dr. Ioannis Stamelos" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>	"Advanced Information Systems: Design, Development and Maintenance" <sup>MA</sup> <http://www.w3.org/2001/XMLSchema#string>

Figure 55: Querying the SPARQL endpoint

The second and more user friendly way the RDF data can be accessed is through a web browser. Currently the data can be browsed and explored from the following URL: <http://lod.ihu.edu.gr:8080/academic>.

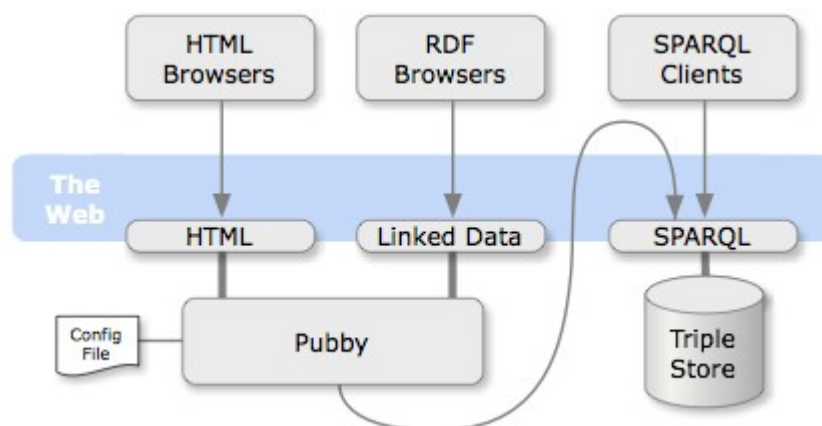


Figure 56: Pubby overview



This exposure method was implemented using Pubby<sup>17</sup>. Pubby is a Linked Data Frontend for SPARQL Endpoints and it can be used to add Linked Data interfaces to SPARQL endpoints. Figure 56 shows the way Pubby works in order to serve Linked Data sitting behind triples stores, as user-friendly HTML pages. An important task Pubby handles is that it provides dereferenceable URIs by rewriting URIs found in the SPARQL-exposed dataset into the Pubby server's namespace. Pubby also handles various details of the HTTP interaction, such as the 303 redirects required by Web Architecture, and content negotiation between HTML, RDF/XML and Turtle descriptions of the same resource. In a nutshell, what Pubby does is that it maps HTTP requests made by the client (web browser) to the corresponding URI of the resource that is stored in the triple store, and it returns the results to the client (by dereferencing the URIs) after having queried the SPARQL endpoint. This way a user can explore the dataset through a browser in a more user-friendly way than having to write sophisticated queries against a SPARQL endpoint. Figure 57 and Figure 58 show two of the dataset instances as they are returned by Pubby. Figure 57 shows the IHU instance and the triples it is associated with, while Figure 58 shows the Knowledge Management instance and its associated triples. We can see that Pubby returns a user friendly HTML page which contains the information that is associated with the specific instance that is queried. Furthermore, the user can easily navigate through the dataset by following the links that are generated by Pubby. Each of these links if clicked will again instruct Pubby to lookup in the SPARQL endpoint for the URI that link represents and return the results associated with it in an HTML page, after first having dereferenced the URI so that the results can correctly be served to the client (web browser).

---

<sup>17</sup> <http://wifo5-03.informatik.uni-mannheim.de/pubby/>

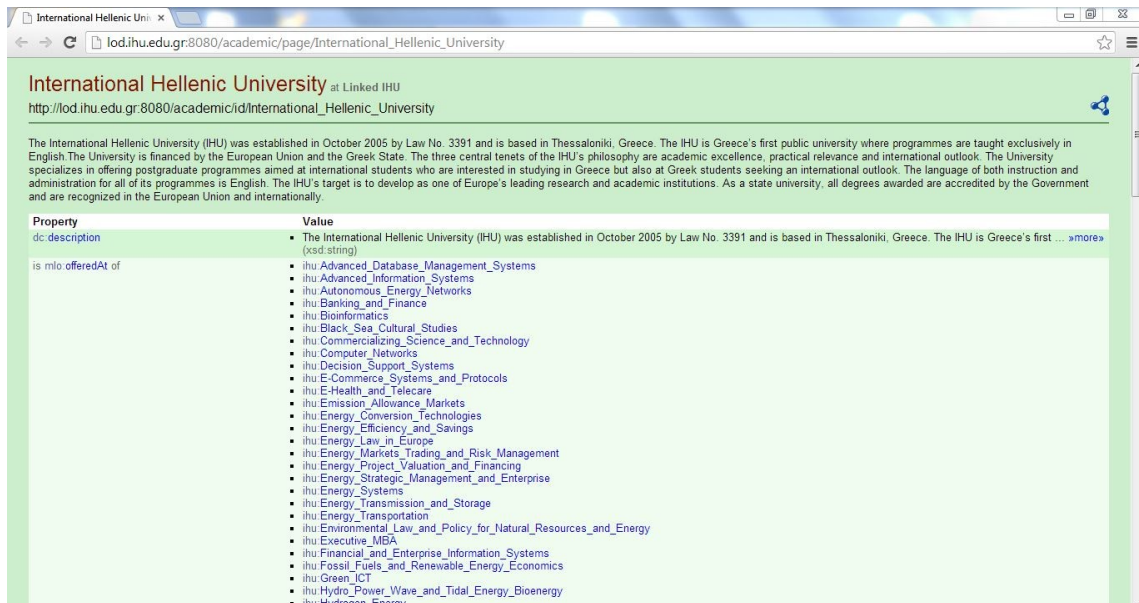


Figure 57: The IHU instance as generated by Pubby

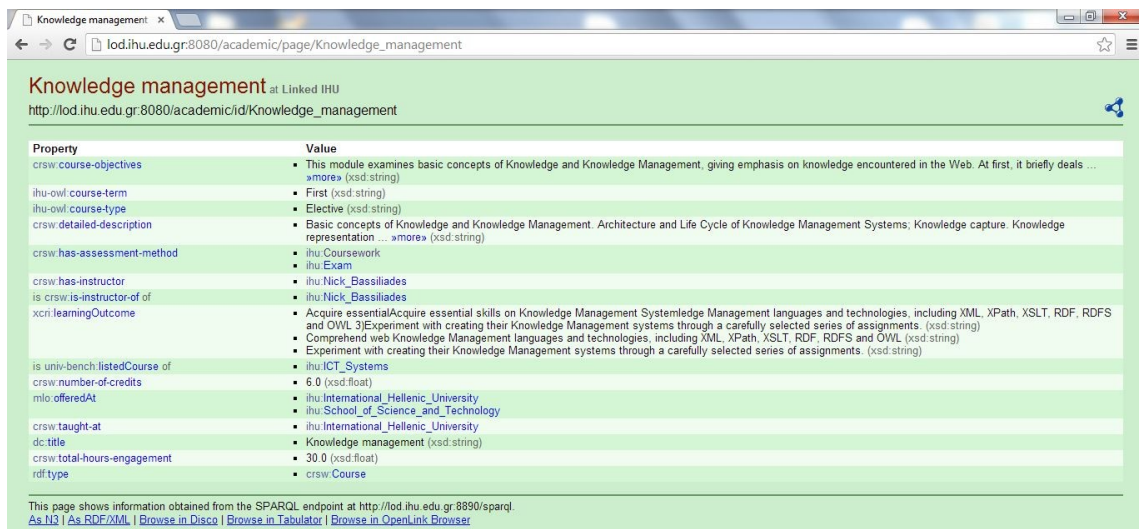


Figure 58: A course instance as generated by Pubby

The final way the data could be accessed is through the RDFa snippets that would be embedded into IHU's webpages, although this access method is mainly provided so that the data that is already provided by the IHU's webpages could be semantically enhanced, in order for them to be more easily understood and indexed by search engines and semantic agents.

In the next subchapter we are going to provide some examples of how the dataset could be accessed through its various interfaces, while also some sophisticated queries are going to be run against the SPARQL endpoint in order to demonstrate how

applications could utilize this functionality in order to consume information from the dataset.

## 5 Conclusions

As we previously mentioned, Linked Data and the Semantic Web in general, is an emerging field, which will hopefully lead to a more “understandable” web for both humans and machines.

The aim of this dissertation was to publish IHU’s data in a more interoperable and semantically enhanced format, so that it can be accessed and explored for further reuse. Since Linked Data is the appropriate format to do this, we created the Academic Ontology which was used as the conceptual layer on top of which IHU’s data was described. Furthermore, we offered two different kinds of methods for accessing the dataset that we created, with the first one being a SPARQL endpoint and the second through a web browser. This way access to the dataset is not only given to machines and/or users with technical knowledge, but also to users that are willing to explore the dataset in an easier and more user friendly manner. Finally, a brief guide was presented for how someone could transform the dataset into RDFa snippets and embed them into IHU’s webpages through its CMS.

The contribution of this effort is quite significant, since only a few academic institutions worldwide have taken the initiative to publish part of their data as Linked Data. The University’s data is now part of the Linked Open Data cloud and it is openly available for anyone to explore and consume. The bases have also been set for the development of interesting applications, which would be able to consume the information provided by the dataset and output useful services for the IHU and not only. The dataset that was created is also interlinked with dbpedia – the largest and most popular dataset currently available - which allows further exploration of the described data into external sources. Another important contribution is the creation of a new ontology – the Academic Ontology – which is also openly available for further use/reuse, by any third parties interested in describing concepts that concern academic institutions. From a more political point of view, this initiative gives to the university a

more open and extroversive personality, since it voluntarily makes its data openly available to the public to exploit and reuse.

## 6 Future Work

In this section, we are going to provide some information concerning the future work that can be done in order to ameliorate the state of the published data and to extend their longevity.

The instances that are currently available as Linked Data were created manually meaning that in the near future they could be out of date. This means that most of the processes that were done manually up to this point, need to be automated by various software components that would serve the following tasks. Firstly, a scheduler is needed which would periodically scan the sources from which the data are extracted and check if there are any differences with the data that is currently stored in the triple store. If the scheduler observes any differences then a specific extractor for each separate source would be needed in order to automatically gather the data and convert it from the incompatible format it is currently in (e.g. HTML, RSS etc.), to RDF. After the data was cleaned and converted into RDF, another automated process would fulfill the task of automatic interlinking with external datasets. There are currently quite a few tools that do this task automatically, such as the Silk<sup>18</sup> framework. When the interlinking process would complete then the new RDF data would need to automatically be added in the triple store, while data that is out of date would need to be deleted. If these processes were present they would enable the dataset to be completely autonomous and continuously up to date. They would also enable the growth of the dataset, since new data would be noticed and added in the dataset. In order for new concepts to be introduced in the dataset though, the extension/modification of the Academic Ontology would be required. This would not be a problem though since the ontology is currently openly available to the public for reuse and extension.

Concerning the RDFa snippets and the brief guide that was presented, an additional automated process could be developed which would implement the steps that were described in the guide. This process could be embedded into the functionality of the

<sup>18</sup> <http://wifo5-03.informatik.uni-mannheim.de/bizer/silk/>

above mentioned tasks. To be more specific, it could be embedded in the stage were the scheduler would have extracted and converted the data into RDF. At this point the “RDFa process” would need to convert the RDF data into RDFa snippets and embed each one of them into a specific article (or module) describing information associated with that particular snippet.

# Bibliography

- [1] Varitimou Anastasia, “Publishing Linked Data from existing governmental datasets”, International Hellenic University, 2012
- [2] C. Bizer, R. Cyganiak, and T. Heath, “How to Publish Linked Data on the Web,” 2007; <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial>.
- [3] T. Berners-Lee. Linked Data. <http://www.w3.org/DesignIssues/LinkedData.html>, 2006
- [4] B. Adida, M. Birbeck, S. McCarron, and S. Pemberton. RDFa in XHTML: Syntax and Processing. W3C Working Draft 18 October 2007, W3C Semantic Web Deployment Working Group, 2007
- [5] Linked Data - the story so far, Bizer, Christian, Heath, Tom and Berners-Lee, Tim (2009) Linked Data - the story so far. International Journal on Semantic Web and Information Systems, 5, (3), 1-22
- [6] W. Halb, Y. Raimond, and M. Hausenblas, "Building Linked Data for Both Humans and Machines," Proc. World Wide Web 2008 Workshop: Linked Data on the Web (LDOW 08), 2008; <http://events.linkedata.org/ldow2008/>
- [7] HS Pinto, JP Martins, Reusing ontologies, AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes, 2000
- [8] Elena Paslaru Bontas , Malgorzata Mochol , Robert Tolksdorf , "Case studies on ontology reuse", In Proceedings of the 5th International Conference on Knowledge Management IKNOW05, 2005
- [9] d'Aquin, Mathieu; Adamou, Alessandro and Dietze, Stefan (2013). Assessing the educational linked data landscape. In: ACM Web Science 2013, 02-04 May 2013, Paris, France
- [10] Stefan Dietze, Salvador Sanchez-Alonso, Hannes Ebner, Hong Qing Yu, Daniela Giordano, Ivana Marenzi, Bernardo Pereira Nunes, (2013) "Interlinking educational

resources and the web of data: A survey of challenges and approaches", Program: electronic library and information systems, Vol. 47 Iss: 1, pp.60 – 91

[11] Yu, Hong Qing; Dietze, Stefan; Li, Ning; Pedrinaci, Carlos; Taibi, Davide; Dovrolis, Nikolas ; Stefanut, Teodor ; Kaldoudi, Eleni and Domingue, John (2011). A linked data-driven & service-oriented architecture for sharing educational resources. In: 1st International Workshop on eLearning Approaches for Linked Data Age (Linked Learning 2011), 8th Extended Semantic Web Conference (ESWC2011), 29 May 2011, Heraklion, Greece

[12] Dietze, Stefan; Yu, Hong Qing; Giordano, Daniela ; Kaldoudi, Eleni ; Dovrolis, Nikolas and Taibi, Davide (2012). Linked education: interlinking educational resources and the web of data. In: The 27th ACM Symposium On Applied Computing (SAC-2012), Special Track on Semantic Web and Applications, 25 - 29 March 2012, Trento, Italy

[13] Ruiz-Calleja, A., Vega-Gorgojo, G., Asensio-Pérez, J.I., Bote-Lorenzo, M.L., Gómez-Sánchez, E., Alario-Hoyos, C.: A Linked Data approach for the discovery of educational ICT tools in the Web of Data. ;Computers & Education(2012)952-962

[14] Ig Ibert Bittencourt, Seiji Isotani, Evandro Costa, Riichiro Mizoguchi: Research directions on Semantic Web and education, Scientia Interdisciplinary Studies in Computer Science 19(1): 60-67, January/June 2008 © 2008 by Unisinos

[15] Capuano, N. Miranda, S. and Orciuoli, O.: IWT: A Semantic Web-based Educational System. IV Workshop of the Working Group on AI & E-Learning, AI\*IA: 11-16, 2009

[16] F. Zablith, M. Fernandez, and M. Rowe. The OU linked open data: production and consumption. In Proceedings of the 1st International Workshop on eLearning Approaches for the Linked Data Age (Linked Learning) at ESWC, Heraklion, Greece, 2011

[17] Kessler, C., Kauppinen, T.: Linked Open Data University of Muenster| Infrastructure and Applications. In: Demos the Extended Semantic Web Conference 2012 (ESWC2012). Heraklion, Crete, Greece (May 2012)

[18] Yuanchao Ma, Bin Xu, Yin Bai, Zonghui Li: Building Linked Open University Data: Tsinghua University Open Data as a Showcase, The Semantic Web, Lecture Notes in Computer Science Volume 7185, 2012, pp 385-393

- [19] Rowe, Matthew, and Fabio Ciravegna. "Data. dcs: Converting Legacy Data into Linked Data." LDOW 628 (2010)
- [20] Taibi, Davide, Giovanni Panasci, and Biagio Lenzitti. "LOD. CS. UNIPA Project: an experience of LOD at the University of Palermo."
- [21] Evangelia Triperina, Cleo Sgouropoulou, and Anastasios Tsolakidis. 2013. AcademIS: an ontology for representing academic activity and collaborations within HEIs. In Proceedings of the 17th Panhellenic Conference on Informatics (PCI '13). ACM, New York, NY, USA, 264-271
- [22] Auer, Sören, et al. "Dbpedia: A nucleus for a web of open data." The semantic web. Springer Berlin Heidelberg, 2007. 722-735

# Appendix

## The Academic Ontology

```
<?xml version="1.0"?>
```

```
<rdf:RDF
```

```
  xmlns:protege="http://protege.stanford.edu/plugins/owl/protege#"

```

```
  xmlns:mlo-xml="http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/mlo_rdfs.xml#"

```

```
  xmlns:xsp="http://www.owl-ontologies.com/2005/08/07/xsp.owl#"

```

```
  xmlns:xcri="http://xcri.org/profiles/catalog/1.2/"

```

```
  xmlns:dc="http://purl.org/dc/elements/1.1/"

```

```
    xmlns:dct="http://purl.org/dc/terms/"

```

```
  xmlns:swrlb="http://www.w3.org/2003/11/swrlb#"

```

```
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"

```

```
  xmlns:xcri-xml="http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/xcri_rdfs.xml#"

```

```
  xmlns:foaf="http://xmlns.com/foaf/0.1/"

```

```
  xmlns:univ-bench="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#"

```

```
  xmlns:crsw="http://courseware.rkbexplorer.com/ontologies/courseware#"

```

```
  xmlns:owl="http://www.w3.org/2002/07/owl#"

```

```
  xmlns:portal="http://www.aktors.org/ontology/portal#"

```

```
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"

```

```
  xmlns:swrl="http://www.w3.org/2003/11/swrl#"

```



```

xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:mlo="http://purl.org/net/mlo/"
xml:base="http://lod.ihu.edu.gr:8080/academic/def/">
<owl:Ontology rdf:about="">
  <rdfs:label>Ihu Academic Ontology</rdfs:label>
  <dc:title xml:lang="en">Ihu Academic Ontology</dc:title>
  <dc:description xml:lang="en">This ontology describes concepts concerning an academic
institution.</dc:description>
  <dc:creator>Michael Hickson</dc:creator>
  <dc:created>2013-11-07</dc:created>
  <dc:source rdf:resource="http://lod.ihu.edu.gr/ontologies/v2/ihu-academic.owl"/>
  <rdfs:seeAlso rdf:resource="http://lod.ihu.edu.gr/ontologies/v2/ihu-academic.html"/>
  <owl:versionInfo>2.0</owl:versionInfo>
  <owl:imports rdf:resource="http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/mlo_rdfs.xml"/>
  <owl:imports rdf:resource="http://svn.cetis.ac.uk/xcri/trunk/bindings/rdf/xcri_rdfs.xml"/>
  <owl:imports rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl"/>
  <owl:imports rdf:resource="http://courseware.rkbexplorer.com/ontologies/courseware"/>
</owl:Ontology>
<rdfs:Class rdf:about="Qualification">
  <rdfs:subClassOf>
    <rdfs:Class rdf:about="http://www.w3.org/2000/01/rdf-schema#Resource"/>
  </rdfs:subClassOf>
</rdfs:Class>
<rdfs:Class rdf:about="http://xmlns.com/foaf/spec/Agent"/>
<owl:Class rdf:about="http://www.aktors.org/ontology/portal#Organization">
  <owl:equivalentClass rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-
bench.owl#Organization"/>
</owl:Class>
<owl:Class rdf:about="AcademicAssistant">
  <rdfs:subClassOf rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Faculty"/>
</owl:Class>
<owl:Class rdf:about="Program">
  <rdfs:subClassOf rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Schedule"/>
  <rdfs:subClassOf rdf:resource="http://xcri.org/profiles/catalog/1.2/presentation"/>
  <rdfs:subClassOf rdf:resource="http://purl.org/net/mlo/LearningOpportunityInstance"/>

```

```

<rdfs:subClassOf rdf:resource="http://purl.org/net/mlo/LearningOpportunitySpecification"/>
<rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
</owl:Class>
<owl:Class rdf:about="AcademicCoordinator">
  <rdfs:subClassOf rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Professor"/>
</owl:Class>
<owl:Class rdf:about="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#University">
  <rdfs:subClassOf rdf:resource="http://purl.org/net/mlo/LearningOpportunityProvider"/>
</owl:Class>
<owl:Class rdf:about="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Person">
  <owl:equivalentClass rdf:resource="http://www.aktors.org/ontology/portal#Person"/>
</owl:Class>
<owl:Class rdf:about="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Department">
  <rdfs:subClassOf rdf:resource="http://purl.org/net/mlo/LearningOpportunityProvider"/>
</owl:Class>
<owl:Class rdf:about="http://courseware.rkbexplorer.com/ontologies/courseware#Course">
  <owl:equivalentClass rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-
bench.owl#Course"/>
  <rdfs:subClassOf rdf:resource="http://purl.org/net/mlo/LearningOpportunitySpecification"/>
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
</owl:Class>
<owl:Class rdf:about="SpecializationStream"/>
<owl:ObjectProperty rdf:about="is-course-of-specialization-stream">
  <rdfs:range rdf:resource="SpecializationStream"/>
  <owl:inverseOf>
    <owl:ObjectProperty rdf:about="specialization-stream-course"/>
  </owl:inverseOf>
  <rdfs:domain rdf:resource="http://courseware.rkbexplorer.com/ontologies/courseware#Course"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="has-specialization-stream">
  <rdfs:range rdf:resource="SpecializationStream"/>
  <owl:inverseOf>
    <owl:ObjectProperty rdf:about="is-specialization-stream-of"/>
  </owl:inverseOf>
  <rdfs:domain rdf:resource="Program"/>

```

```

</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="http://courseware.rkbexplorer.com/ontologies/courseware#is-
instructor-of">
  <rdfs:domain rdf:resource="http://www.aktors.org/ontology/portal#Person"/>
  <rdfs:range rdf:resource="http://courseware.rkbexplorer.com/ontologies/courseware#Course"/>
  <owl:inverseOf>
    <rdf:Description rdf:about="http://courseware.rkbexplorer.com/ontologies/courseware#has-
instructor">
      <owl:inverseOf rdf:resource="http://courseware.rkbexplorer.com/ontologies/courseware#is-
instructor-of"/>
    </rdf:Description>
  </owl:inverseOf>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="is-specialization-stream-of">
  <rdfs:domain rdf:resource="SpecializationStream"/>
  <rdfs:range rdf:resource="Program"/>
  <owl:inverseOf rdf:resource="has-specialization-stream"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="specialization-stream-course">
  <rdfs:range rdf:resource="http://courseware.rkbexplorer.com/ontologies/courseware#Course"/>
  <rdfs:domain rdf:resource="SpecializationStream"/>
  <owl:inverseOf rdf:resource="is-course-of-specialization-stream"/>
</owl:ObjectProperty>
<rdf:Property rdf:about="http://purl.org/dc/elements/1.1/relation"/>
<rdf:Property rdf:about="http://xcri.org/profiles/catalog/1.2/learningOutcome">
  <rdfs:domain rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
</rdf:Property>
<rdf:Property rdf:about="http://purl.org/dc/elements/1.1/language"/>
<rdf:Property rdf:about="http://purl.org/dc/elements/1.1/date"/>
<owl:DatatypeProperty rdf:about="course-type">
  <rdfs:domain rdf:resource="http://courseware.rkbexplorer.com/ontologies/courseware#Course"/>
  <rdfs:range>
    <owl:DataRange>
      <owl:oneOf rdf:parseType="Resource">
        <rdf:rest rdf:parseType="Resource">

```

```

    <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Mandatory</rdf:first>
    <rdf:rest rdf:parseType="Resource">
      <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#nil"/>
      <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
      >Elective</rdf:first>
    </rdf:rest>
  </rdf:rest>
  <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
  >Required</rdf:first>
</owl:oneOf>
</owl:DataRange>
</rdfs:range>
</owl:DatatypeProperty>
<owl:DatatypeProperty rdf:about="faculty-type">
  <rdfs:domain rdf:resource="http://www.lehigh.edu/~zhp2/2004/0401/univ-bench.owl#Person"/>
  <rdfs:range>
    <owl:DataRange>
      <owl:oneOf rdf:parseType="Resource">
        <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
        >Faculty</rdf:first>
        <rdf:rest rdf:parseType="Resource">
          <rdf:rest rdf:parseType="Resource">
            <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >Adjunct Faculty</rdf:first>
            <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#nil"/>
          </rdf:rest>
          <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
          >Visiting Faculty</rdf:first>
        </rdf:rest>
      </owl:oneOf>
    </owl:DataRange>
  </rdfs:range>
</owl:DatatypeProperty>

```

```
<owl:DatatypeProperty rdf:about="course-term">
  <rdfs:domain rdf:resource="http://courseware.rkbexplorer.com/ontologies/courseware#Course"/>
  <rdfs:range>
    <owl:DataRange>
      <owl:oneOf rdf:parseType="Resource">
        <rdf:rest rdf:parseType="Resource">
          <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >Second</rdf:first>
          <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#nil"/>
        </rdf:rest>
        <rdf:first rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
            >First</rdf:first>
      </owl:oneOf>
    </owl:DataRange>
  </rdfs:range>
</owl:DatatypeProperty>
</rdf:RDF>
```