
Deliverable 1.1.1

Curriculum Plan

Coordinator: Maria Maleshkova

With contributions from: Maribel Acosta, Elena Simperl

Quality Assessor: Barry Norton

Editor:	Maria Maleshkova, KIT
Deliverable nature:	Other (O)
Dissemination level: (Confidentiality)	Public (PU)
Contractual delivery date:	31.07.2012
Actual delivery date:	31.07.2012
Version:	1.0
Total number of pages:	23
Keywords:	Curriculum plan, training modules

Executive summary

Currently Linked Data technologies are becoming more and more important in the context of exposing and accessing structured data on the Web. Moreover, they are establishing themselves as a de facto standard for publishing and interlinking data sets for seamless exchange, integration, and reuse. This trend becomes even more evident through the growth in terms of the number of organizations committing to use Linked Data core principles but also through the plenitude of applications build on top of the already available data. In this context, EUCLID address the need for trained data practitioners who are able to apply a Linked Data approach as part of their data business solutions.

To this purpose, EUCLID provides a comprehensive educational curriculum, supported by multi-modal learning materials and highly visible eLearning distribution channels, tailored to the real needs of data practitioners. This deliverable describes the first version of the Linked Data training curriculum plan. The first section describes the purpose of the curriculum, its target groups, the different methods and materials that it encompasses, as well as the targeted level of knowledge that is to be achieved by participating in the trainings. We also describe what should be the prerequisite qualification for the targeted training groups.

The second part of the deliverable focuses on describing in grater detail the actual curriculum plan. This is the main part of the deliverable and it includes a first draft of a detailed outline about the topics that are to be covered and the way they are to be structured and delivered. The curriculum plan is divided into seven main modules, which encompass three levels of topics– introductory, advanced and expertise. The main modules built on each other adding more specific knowledge in particular application areas, such as Linked Data visualisation and analysis. The seven main modules are – Introduction to Linked Data and Application Scenarios, Querying Linked Data, Publishing, Interlinking and Cleansing Linked Data, Visualizing Linked Data, Analysing and Mining Linked Data, Creating Mashups and Linked Data Application, and Scaling-up. The curriculum plan in its current form captures the content and topic to be covered, however, the individual modules will undergo a number of further revisions. In particular, one distinguishing approach of EUCLUD for content delivery is the improvement of provided trainings based on gathered feedback, in order to produce really high-quality teaching materials. Therefore, the curriculum plan described here should be seen as means for capturing the relevant topics and structuring the training content and not as a final detailed plan for delivering courses for Linked Data practitioners.

The final section of the deliverable takes into consideration existing programs and initiatives for providing training materials related to Linked Data. Given the current popularity of this topic and the growing interest in the area, it is natural that there are already some teaching materials and courses. Our purpose it to identify these and to develop a “mapping” that determines the overlap and the coverage of the EUCLID curriculum. Based on this analysis we can objectively argue for the completeness of the provided curriculum, in terms of the covered topics but also taking into consideration the training goals and target audience. Furthermore, we can better align with further existing initiatives in the field, initiate collaborations and plan the organisation of the training events accordingly. This deliverable only identifies a list of relevant Linked Data training activities and organizers, giving only a few details. Therefore, it represents only a first draft that is to be extended with specific outlines and topics in the deliverables to come.

Document Information

IST Project Number	FP7 - 296229	Acronym	EUCLID
Full Title	Educational curriculum for the usage of Linked Data		
Project URL	http://www.euclid-project.eu/		
Document URL	http://www.euclid-project.eu/resources/deliverables/		
EU Project Officer	Werner Janusch		

Deliverable	Number	1.1.1	Title	Curriculum Plan
Work Package	Number	1	Title	Course Production and Delivery

Date of Delivery	Contractual	M3	Actual	M3
Status	version 1.0		final - <input checked="" type="checkbox"/>	
Nature	other			
Dissemination level	public			

Authors (Partner)	Maria Maleshkova (KIT), Maribel Acosta (KIT), Elena Simperl (STI-R)			
Responsible Author	Name	Maria Maleshkova	E-mail	maria.maleshkova@kit.edu
	Partner	KIT	Phone	+49 721 608 4 7363

Abstract (for dissemination)	<p>EUCLID provides a comprehensive educational curriculum, supported by multi-modal learning materials and highly visible eLearning distribution channels, tailored to the real needs of data practitioners. This deliverable describes the first version of the Linked Data training curriculum plan. The first section describes the purpose of the curriculum, its target groups, the different methods and materials that it encompasses, as well as the targeted level of knowledge that is to be achieved by participating in the trainings. The second section provides the actual curriculum plan and a detailed outline of the included seven modules. The final section takes into consideration other existing initiatives and curricula that cover Linked Data topics and proposes an alignment with the topics covered by EUCLID's training plan.</p>
Keywords	Curriculum plan, training modules

Version Log			
Issue Date	Rev. No.	Author	Change
09.07.2012	0.1	Maria Maleshkova (KIT)	Initial deliverable draft
20.07.2012	0.2	Maria Maleshkova (KIT)	Complete draft, without the curriculum mapping section
26.07.2012	0.3	Maria Maleshkova (KIT)	Revised and improved initial draft
29.07.2012	0.4	Maria Maleshkova (KIT)	Revised and improved draft
30.07.2012	1.0	Maria Maleshkova (KIT)	Final version

Table of Contents

EXECUTIVE SUMMARY	2
DOCUMENT INFORMATION	3
TABLE OF CONTENTS.....	5
LIST OF FIGURES AND TABLES.....	6
ABBREVIATIONS.....	7
1 INTRODUCTION	8
1.1 PURPOSE OF THE CURRICULUM PLAN	8
1.2 TARGET AUDIENCE AND GOAL LEVEL OF KNOWLEDGE	8
1.3 SKILLS AND KNOWLEDGE PREREQUISITES	9
1.4 DELIVERY METHODS AND MATERIALS	9
2 STRUCTURE OF THE CURRICULUM	12
2.1 INTRODUCTION AND APPLICATION SCENARIOS	12
2.1.1 <i>Motivating Scenario</i>	13
2.1.2 <i>Module Detailed Outline</i>	14
2.1.3 <i>Interactive Examples</i>	14
2.2 QUERYING LINKED DATA	14
2.3 PUBLISHING, INTERLINKING AND CLEANSING LINKED DATA.....	15
2.4 VISUALIZING LINKED DATA	16
2.5 ANALYSING AND MINING LINKED DATA	16
2.6 CREATING MASH - UPS AND LINKED DATA APPLICATIONS	17
2.7 SCALING - UP	18
3 RELATIONSHIP TO OTHER LINKED DATA TRAINING ACTIVITIES	19
3.1 PLANETDATA TRAINING CURRICULUM.....	19
3.2 VIDEO LECTURES AND SUMMER SCHOOLS.....	20
3.3 PROFESSIONAL TRAINING CURRICULA.....	21
4 CONCLUSION	22
REFERENCES.....	23

List of Figures and Tables

Figure 1: Materials Production Approach	10
Figure 2: Modules Revision Process	11
Figure 3: EUCLID Curriculum Plan	12

Abbreviations

DBMSs – Database Management System standard

DL – Distance Learning

DoW – Description of Work

EU – European Union

IT – Information Technology

KR – Knowledge Representation

LD - Linked Data

LOD – Lined Open Data

OWL – Ontology Web Language

OWL-S – OWL for Services/ OWL-based Web Service Ontology (formerly DAML-S)

RDF/S – Resource Description Framework / Schema

RDBMS - Relational Database Management System

SPARQL – SPARQL Protocol and RDF Query Language

SQL – Structured Query Language

WP – Work Package

WSMO – Web Service Modelling Ontology

XML - Extensible Markup Language

1 Introduction

Linked Data has achieved great success and popularity during the past few years, paving its way towards becoming a de-facto standard for publishing and accessing structured data on the Web. Therefore, it is not a surprise that this trend becomes evident in the context of government organizations but also companies, which deal with data exchange, manipulation and maintenance in their daily business activities. This trend naturally raises the interest in Linked Data technologies but also motivates the need for trained data practitioners, who can apply Linked Data solutions in different contexts and as part of various solutions. EUCLID aims to address precisely this need by providing an extensive training curriculum, backed up by a range of social and community-based activities, which aim to disseminate but also to gather feedback about the provided training materials. Therefore, the scope of EUCLID is twofold, first to provide a training plan and a set of training resources that will enable the development of skills and knowledge based on Linked Data technologies, and second, to expose this resources to the community through a range of social activities, such as engagement in online forums and Linked Data-centric events, and to gather opinions and comments, in order to further improve the content of the training courses.

The purpose of this deliverable, in particular, is to provide the first draft of the curriculum plan. Since the curriculum plan prescribes the structure and the content of the training materials that are to be generated by EUCLID, it represents a major milestone of the project. The curriculum plan given in this deliverable is actually more than a first drafted version. In an effort to provided high-quality training, suitable for the data practitioner's needs, we already presented the curriculum for revision to a number of experts and gathered feedback, which already lead to the restructuring and rearrangement of the modules, as well as to the adjustment of some of the content.

1.1 Purpose of the Curriculum Plan

In order to achieve the objective of the EUCLID project and to be able to provide training materials suitable for an audience with different level of background knowledge and aiming to apply Linked Data technologies in a variety of application areas, it is important to provide a curriculum plan that encompasses basic and fundamental training as well as covers more advanced topics, which require some prior experience in the field. Similarly, it is also beneficial to provide examples and use case scenarios that demonstrate the practical applicability of the learned concepts as well as motivate the technology and design decisions that are taken. Therefore, the curriculum plan describes the covered topics but, moreover, it also specifies the approach taken towards communicating the necessary knowledge about Linked Data technology and principles. It provides a set of modules that gradually build up the level of knowledge of the trainees and complements the theoretical teaching with a set of examples and hands-on tasks, which show how to directly apply the learned skills.

In summary, the curriculum plan takes a practice- and application-oriented approach towards communicating essential Linked Data knowledge that would help data practitioners to apply this new technology in the context of their daily tasks.

1.2 Target Audience and Goal Level of Knowledge

As already mentioned, the target audience of this curriculum plan is data practitioners and professionals, who already use or aim to adopt Linked Data as means for publishing and accessing structured data over the Web. The EUCLID trainee candidate is someone who aims to gain a broader and deeper understanding of how to deal with and manage data, in accordance with Linked Data principles, in the context of his/her professional activities.

The training participants can have no prior knowledge in the field but can also already have a given level of past experience with some of the introduced technologies. Therefore, an initial quick evaluation or short test would be useful in order to identify, which module and which level would be best to begin with. Furthermore, during the training the participants are expected to complete short assignments, work through hands-on exercises, and devote some time after the course hours to familiarize themselves with the introduced concepts.

Given the context and the target group of the EUCLID project, it is to be expected that the concrete training materials focus more on providing real business application examples and relating the learnt topics to those example. Therefore, the goal is to provide a curriculum plan that includes use cases and scenarios, which directly demonstrate the practical applicability of the learned concepts and technologies.

Considering the goal level of knowledge, the curriculum provides three possible levels of expertise that can be gained through the trainings. This first level is based on a set of modules that communicate the basic and fundamental skill that are required in order to begin adopting and applying Linked Data technologies. The second level deals with more advanced topics, also specializing in different areas such as visualization and query processing. The last level aims to provide expertise knowledge that is rather specific to an area of use and requires somewhat extensive prior knowledge. Therefore instead of aiming to bring all trainees to the same level, the curriculum plan is structured in such a way as to be adaptable to address the particular needs of the audience and serve both professionals who strive to become experts or want to simply get some basic knowledge in the area, alike.

1.3 Skills and Knowledge Prerequisites

All potential course participants require a basic level of computer literacy. A background in IT development and engineering is an advantage but not necessarily a requirement. Still, lack of basic background knowledge in IT might require further individual learning. Most important is that candidates have a keen interest in learning about the foundations of Linked Data, the Semantic Technology, Databases, and that they have an intention to produce, consume and manage data on the Web by applying the Linked Data principles. Naturally, the precise pre-requirements depend on the candidate's goals and on the training context. For example, a targeted level of expertise knowledge can be achieved, given a solid background in the particular field.

1.4 Delivery Methods and Materials

The curriculum plan is based on a set of delivery methods and materials that complement each other and are targeted towards addressing the different needs of the trained data professionals. This subsection outlines the possible delivery methods and the format of the modules, while each participant can choose the best combination of using and accessing the training materials.

Delivery methods:

Since the main target group of the curriculum plan is data professionals, naturally self-training and distance learning are the main means of communicating the courses content. These types of training methods are relatively flexible when it comes to geographical location and times-slot allocation and are, therefore, suitable for on-the-job but also parallel-to-the-job training.

- **Self-training**

Self-training is based on strong individual engagement and motivation, where the candidate explores and learns the course material by himself/herself. The course material can be acquired online, for example, via the EUCLID online portal. The materials could be used in printed or online form. It is up to the student to decide what level of expertise in the area of Linked Data technologies he/she wants to acquire. In this context there is usually no direct trainer support and onsite tutorial. It is only the official course material, which the candidates learn by themselves, with the support of existing community forums for possible question answering and discussions.

- **Distance learning**

The EUCLID project does not directly foresee the organization of distance learning events, however the combination of the offered live webinars, guided tutorials, the official course materials and access to recorded live webinars can easily serve as the basis for supporting such a learning approach. The only difference is that the project aims to shift the discussion and question-answering part to the growing Linked Data online community, thus not only benefiting directly from the knowledge and know-how of experts in the field but also encouraging new Linked Data practitioners to become involved in some of the existing online forums.

- **On-site training**

On-site training is a very effective and gives the opportunity for high-level interaction between the trainer and the trainee. This will be realised through the organization of a number of training events, possibly collocated with

Linked Data-centric conferences and events. This method is the best way to learn and get in contact with trainers for discussions and direct questions.

Training materials:

EUCLID aims to provide high-quality training materials. For this purpose it implements two main content-generation approaches. First, the individual curriculum modules are delivered both as written content as well as a webinar. As visualised in Figure 1, after the webinar is delivered, all training materials, including written documents, examples, presentation slides as well as the video recording of the webinar, will be used to create an eBook. Second, EUCLID implements an approach of revising the training content based on gathered feedback (see Figure 2). After the webinar is given, comments and suggestions will be gathered from the audience. Furthermore, each draft of the modules will be made available online and community feedback will actively be gathered. As a result, the final version of training materials will have undergone at least one round of improvements and revision.

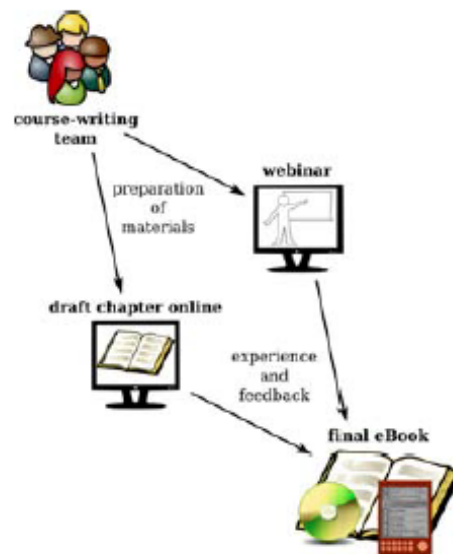


Figure 1: Materials Production Approach

The available forms of the training materials are the following:

- **Webinars**

A webinar is an online lecture with some interactive elements. This online training includes interactive presentation, lecture, workshop or seminar that is transmitted over the web. Participants can directly interact by giving, receiving and discussing material of the course in real time. The first webinar is already scheduled for the first week of August.

- **Online materials**

Each module is based on a set of online materials, including the slides to the webinars, examples, training exercises and self-assessment questions. Learning materials are structured according to learning outcomes, which are made explicit for the learner; embedded questions, strategically apportioned throughout the course text, support the validation of learnt outcomes.

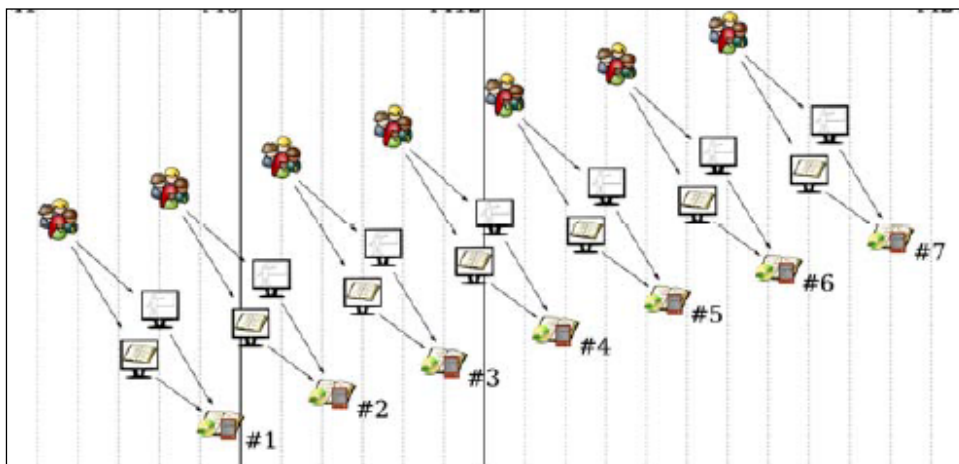


Figure 2: Modules Revision Process

- **eBooks**

An eBook is created for each curriculum module, based on the online materials and the webinar. It encompasses all the content to a particular topic, in a structured and interactive way. eBooks serve very well as the basis for self-learning as well as for revisiting certain topics after a training is completes as part of an EUCLID event, for example. After each webinar, the gathered feedback is used to restructure the module content for final delivery as an eBook series. Therefore, the eBooks represent the final outcome of the training materials revising process.

2 Structure of the Curriculum

This section described in greater detail the first version of the EUCLID curriculum plan. The outline is based on seven main modules that encompass three levels of topics – introductory, advanced and expertise. Figure 3 visualizes the individual models and their grouping into the three levels of expertise gained.

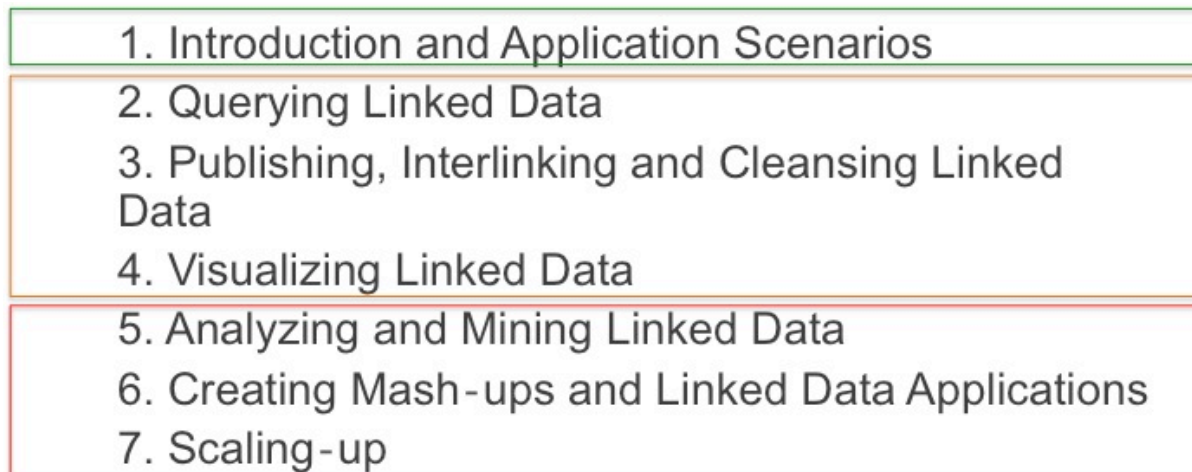


Figure 3: EUCLID Curriculum Plan

As it can be seen, the modules are defined in such a way as to gradually build up trainee’s knowledge. Furthermore, it enables course participants with previous knowledge or a specific area of interest to only briefly go over the introductory materials and directly dig into one of the more advanced modules. As already mentioned earlier, this proposed curriculum will provide a basis for the development of professional training courses, to ensure an organized guidance for their content, focusing on the necessary skills and technologies for Linked Data processing and use. Naturally, the proposed curriculum can be adjusted in order to better suit the trainees’ previous level of knowledge and intended goals of the training. Furthermore, individual courses can be designed by adjusting or mixing parts of the different modules in order to address only a more specific area of interest or to provide only a basic or advanced level of knowledge, for example, leaving out the expertise training.

Each module is composed of a set of sub-topics, which are relevant for Linked Data practitioner, i.e. professionals who will need to be able to use computer systems and software to produce, consume, and manage Linked Data. These sub-topics are ordered, as the modules, based on the required level of proficiency, i.e. we start with introductory sub-topics, within the overall module topic, and progress to more advanced topics, which need already higher proficiency in the topic, as gained through earlier sub-topics. As a result, the more proficient a candidate is in the respective topic, the ‘later’ they can begin within the presented curriculum.

The curriculum aims to be very practice and application-oriented. Therefore, all the modules use examples based on a common motivation scenario. The scenario exemplifies the benefits of applying the Linked Data principles and demonstrates individual use cases where applying the learnt technologies can be very beneficial. A summary of the motivation scenario is given as part of the first module and in Section 2.1.1. The following sections describe each of the curriculum modules in greater detail.

2.1 Introduction and Application Scenarios

The first module aims to provide a general overview of the main topics related to using Linked Data. It is only an introduction and some of the topics are only mentioned and then discussed in greater detail in one of the following modules. The main goal of this module is to describe the overall motivating scenario and to teach the fundamental

Linked Data principles, while briefly describing the context of the technologies and possible application solutions. Module summary:

- 4 Units
- Goal: Introduction to the foundations and fundamental technologies of Linked Data.
- Outcome: Overview of the main Linked Data principles, concepts, scenarios and use cases.

It is important to point out that the first module is described in greater detail than the other modules because it will be used to deliver the first webinar and to produce the pilot eBook. Therefore, it is further along in terms of preparing the outline, the content materials, the used the examples and the interactive elements, such as a SPARQL endpoint for test query execution. The following modules currently only consist of a detailed outline of the covered topics and a set of gathered relevant materials and use cases.

2.1.1 Motivating Scenario

As a motivating scenario for EUCLID we consider the provision of a music-based portal, and the challenges and benefits of using Linked Data in creating such a portal. In fact a number services and mash-ups exist, which already drawn benefit from existing Linked Open Data and Semantics-based technologies; we point out as examples [1], [2] and [3]. Two major artifacts that these, and other, Linked Data-based approaches in the music domain share in common are the Music Ontology [4] and the MusicBrainz dataset [5]. These two artifacts will be used as examples throughout the EUCLID training material.

In order to provide a useful portal, the developer in this scenario would like to bring together a number of disparate components of data-oriented content:

- 1) Musical content - content exists in the third-party commercial setting (links into download and streaming providers, e.g. Amazon/iTunes and Spotify/Last.fm), the license-free setting (e.g. the Live Music Archive 'etree'), and the grey market setting (e.g. YouTube);
- 2) Music and artist metadata - while the MusicBrainz dataset is the primary source (it's primary artifact being a relational database, but EUCLID material will demonstrate how this is transformed into RDF via D2R), it is weak on biographical and genre information, for which alternative sources will be discussed (including Freebase and DBpedia);
- 3) Review content - reviews exist that are already Linked Data-oriented (e.g. BBC Music Reviews), that are semi-structured but unlinked (e.g. Pitchfork) and that are largely unstructured (the Web in general);
- 4) Visual content - photographic depictions, album covers and videos exist on the Web but are loosely coupled in terms of semantic interlinking.

The portal developer will use common identifiers to bring together this disparate content, and furthermore to offer interesting mash-ups using the inter-linkage to further data from the Linking Open Data Cloud, e.g. geographical and biographical exploration, and the possibility to provide engaging visualisations over this.

The developer will also seek to improve the quality of the semantic interlinking of the content they aggregate and contribute back to the Linking Open Data Cloud. In particular they will improve the linking of artists and works to visual content and to reviews, in the latter case crawling review content and publishing external annotations. They will also seek to improve classification within the metadata, encoding genre information - at least with respect to the emphasis of their portal - along the way demonstrate the use of the Google Refine technology.

Finally, prototypical examples from the portal will demonstrate the use of RDFa annotation of human-readable content and demonstrate the link to emerging Web technologies that inherit from semantics, such as Google

RichSnippets, Facebook OpenGraph and schema.org annotation.

2.1.2 Module Detailed Outline

1. Motivation scenario (as described in Section 2.1.1)
 - a. Music-based portal, exemplifying some of the common challenges of dealing with structured data on the Web and demonstrating the need for data integration, data exploration, and data visualisation
2. Linked Data foundations (introduction to the underlying technologies - HTTP, URIs, XML, RDF, SPARQL)
 - a. Evolution of the Web
 - b. Web technology basics (HTTP, URIs)
 - c. Describing and exchanging data (XML)
 - d. Semantics on the Web
 - i. RDF, RDFs, Ontologies, OWL, Vocabularies
 - ii. Music-based portal example and the use of semantics/used vocabularies
 - e. Querying semantic data (SPARQL)
 - i. Music-based portal example with a simple query
3. Introduction to Linked Data
 - a. Principles of Linked Data
 - b. The Web of Data
 - i. Types of data, popular data domains, web of data growth
 - c. Exploring the Web of Data (applications, analysis and visualisation techniques)
4. Introduction of Linked Data use case scenarios

2.1.3 Interactive Examples

Each of the modules will contain a set of examples, self-assessment questions and further interactive materials. The first module only presents an overview of some of the advanced topics but it already includes some interactive examples so that the trainees can get a first impression of some of the courses that follow. This is especially true for the interactive SPARQL query examples, which serve as a teaser for the following module (see Section 2.2) that explores SPARQL use and processing in much more detail.

2.2 Querying Linked Data

Querying Linked Data is the first of the three advanced topic modules. These modules focus on data management and processing tasks that are crucial for using Linked Data in a business environment.

Some of the main tasks that need to be completed by the data practitioner are the retrieval and modification of existing datasets. This includes getting familiar with the corresponding query language, in this case SPARQL, learning how to create queries, dealing with data stores and exploring some further Linked Data query options. The main objective of the module is to give the trainees the tools and knowledge in order to be able to explore and update Linked Data.

Module summary:

- 3 Units
- Goal: Introduction to SPARQL query language, RDF stores and SPARQL engines.
- Outcome: Installing and working with SPARQL stores, ability to define SPARQL queries.

Module outline:

1. Querying Linked Data with SPARQL
 - a. SPARQL query methods
 - b. Query examples based on the use case scenario
2. Query processing
 - a. SPARQL-based querying
 - b. URI look up
 - c. Federated query processing
 - d. SPARQL-endpoints and RDF stores
3. Reasoning over Linked Data
 - a. Reasoning for data integration

This module is very practice-oriented; therefore, it will include a lot of examples, enabling the training participants to compose their own queries and do test query execution on a setup dataset.

2.3 Publishing, Interlinking and Cleansing Linked Data

Making existing structured data available on the Web, in accordance with the Linked Data principles is one of the main tasks that need to be performed by professionals in this field. The previous module teaches all the necessary skill for querying already published data. This module, however, focuses on how to make the data available in the first place. In particular, this encompasses the completion of three main tasks – publishing, interlinking and cleansing, which can be performed in this particular order but can also take place independently, depending on the goal that the data practitioner wants to achieve.

Module summary:

- 3 Units
- Goal: Introduction to techniques for extracting, transforming, and loading data in the form of Linked Data, data cleansing, data interlinking base on instance and vocabulary mapping.
- Outcome: Ability to publish, interlink and cleanse Web data sets.

Module outline:

1. Publishing Linked Data
 - a. Linked Data publishing guidelines
 - i. Data transformation
 - b. Linked Data publishing patterns
 - c. Further Linked Data publishing approaches

- d. Linked Data publishing checklist
2. Linked Data interlinking
 - a. Interlinking process
 - b. Multi-source interlinking
 - c. Instance and vocabulary mapping
 - d. Schema mappings and transformations
3. Linked Data cleansing
 - a. Data cleansing
 - b. Testing and debugging Linked Data

2.4 Visualizing Linked Data

This section describes the module outline to Visualizing Linked Data. In particular, the focus is on introducing tools for browsing and exploring Linked Data in a graphical way. Alongside with querying existing dataset and publishing structured data in accordance with Linked Data principles, visualizing data is also a task that is commonly performed by data practitioners and is especially suitable for certain use cases. In addition to following the graphs of a single dataset, this also includes creating charts or summaries over integrated data, which make the synergy effect of combining different pieces of information directly visible. The focus of the module is on introducing common techniques but especially on describing some of the commonly used tools in the area.

Module summary:

- 2 Units
- Goal: Introduction to visualization components and toolkits for Linked Data.
- Outcome: Ability to browse through Linked Data datasets, skills to generate charts based on integrated data.

Following is a module outline, which will be extended with specific examples that can be tested in different tools, in order to directly explore the practical applicability of the learnt techniques.

Module outline:

1. Linked Data browsers
 - a. Linked Data browsing solutions
2. Linked Data visualisation
 - a. Visualisation techniques
 - b. Visualisation toolkits

2.5 Analysing and Mining Linked Data

The final three modules cover a list of more advanced topics that are relevant for data practitioners who want to acquire expertise knowledge and skills in specific Linked Data application areas. In contrast to the previous modules, the set of expertise topics require familiarity and previous experience with the principles and technologies from the introductory course.

This module focuses on presenting approaches and tools for analysing and mining Linked Data. Data mining is a prominent and well-established research and application field that uses a diversity of methods in order to draw conclusions and make statements about existing datasets. Similarly, in the context of Linked Data, analysis and

mining can be applied in order to determine co-relations, derive trends, identify topics of interest, draw conclusions, and many more. To this end, this curriculum module touches on some of the common techniques used for LD analysis.

Module summary:

- Summary: 2 Units
- Goal: Introduction to statistical methods for data analysis and clustering.
- Outcome: Installing and working with statistical data analysis tools.

Module outline:

1. Methods for Linked Data analysis
 - a. Statistical data analysis for Linked Data
 - b. Data clustering for Linked Data
 - c. Linked Data analysis tools as platforms
2. Mining for Linked Data

Data analysis and mining encompasses a broad area of methods and tools. Therefore, this module will focus on exploring the most commonly used approaches in the context of Linked Data and provide examples by demonstrating some of the available applications.

2.6 Creating Mash-ups and Linked Data Applications

Linked Data can be exploited very effectively through the creation of mashups and applications built on top of it. Therefore, this module describes existing common approaches towards designing and implementing applications that consume Linked Data. In addition, this part of the curriculum also introduces techniques for creating mashups based on combining Web APIs and Linked Data. The module is very practically oriented and the trainees can benefit from some previous experience in the area of software development and engineering. Still, this is not a hard requirement, since the module also aims to convey the theoretical knowledge on how to approach the development of applications, while the particular implementation can be realised with the help of supporting tools.

Module summary:

- 4 Units
- Goal: Overview of Web APIs and integrating Web APIs and Linked Data. Convey basic skills for creating mashups and building applications on top of Linked Data.
- Outcome: Familiarity with techniques for using Web APIs and Linked Data in order to create mashups and applications.

Mashups and applications, built on top of Linked Data, have been gaining in popularity and represent an important way of exploiting the data through the manipulation and integration of the underlying datasets but also through the provisioning of user interfaces that effectively address the end user's needs in a particular use case. This module introduces an overall scenario that is used to motivate and give examples throughout the individual sub-topics. It provides an introduction to Web APIs and describes common approaches for creating applications and mashups.

Module outline:

1. Motivation scenario (data integration based on mashups)
2. Linked Data applications
 - a. Architecture of Linked Data applications

- b. Deploying Linked Data applications
 - c. Linked Data application roles (publisher, consumer, third-party)
3. Introduction Web APIs
 - a. Underlying technology basics
 - b. Describing Web APIs
 - c. Web APIs search and access
 - d. Web API authentication
 - e. Web API integration with Linked Data
 4. Creating data mashups
 - a. Procedural coding
 - b. Challenges and shortcomings of creating mashups
 - c. Platforms for creating mashups
 - d. Mashups and application directories
 - e. Mashups based on Linked APIs

2.7 Scaling-up

The final module of the curriculum is devoted to dealing with large amounts of data and using and managing these in an effective way. It includes topics on scaling-up, on top of some of the learnt approaches throughout the curriculum, in order to be able to deal with numbers and sizes of Linked Data dataset that challenge the computational limits. In particular, it introduces approaches for Linked Data cloud development and big data processing. As already mentioned, this is a module with expertise topic level, which requires some previous knowledge in the field and might not be of immediate interest to all data practitioners to-be.

Module summary:

- 2 Units
- Goal: Introduction to large-scale data processing on computing clusters, cloud deployment.
- Outcome: Installing and working with distributed data processing systems.

The module is structured around two main topics – Linked Data cloud deployment and big data processing. Each of these topics is introduced and motivated with the help of examples.

Module outline:

1. Linked Data cloud deployment and processing
 - a. Linked Data and cloud deployment
2. Big data processing
 - a. Computing clusters for Linked Data
 - b. Processing systems

3 Relationship to Other Linked Data Training Activities

Given the growing popularity of Linked Data and the increasing interest in the area, it is natural that there are already some teaching materials and courses on related topics. Our purpose is to identify these and to develop a “mapping” that determines the overlap and the coverage that existing training initiatives and courses have with the EUCLID curriculum. Based on this analysis we can objectively argue for the completeness of the provided curriculum, in terms of the covered topics, but also taking into consideration the training goals and target audience. Furthermore, we can better align with existing initiatives in the field, initiate collaborations and plan the organisation of the training events accordingly.

Since the main objective of this deliverable is to describe the curriculum plan, in this section we only identify a list of relevant Linked Data training activities and organizers, giving only a few details. Therefore, this section represents only a first draft that is to be extended with specific outlines and topics in the deliverables to come.

3.1 PlanetData Training Curriculum

One of the major initiatives in the field that focus on big data is the PlanetData project, which aims to establish a community of researchers that supports organizations in exposing their data in new and useful ways. The project aims to enable maintainable large-scale access to structured data that has been exposed by organisations. To support the future production and consumption of large scales of data, there should be training and education programs, addressing the interested organisations in the industry and the academic community, on large-scale data management and the underlying approaches and technologies. Therefore, PlanetData provides an extensive curriculum for making training possible for both the academic and industrial communities.

The curriculum covers four main topics – Semantic Technology, which includes the Semantic Web, Database Technology, and Linked Data and Data Streams. While the last topic is obviously the most relevant one for taking into consideration for EUCLID’s curriculum, some of the remaining topics, such as Semantic Web are partially relevant as well. The complete Planet Data curriculum is available at [6], while the covered Linked Data topics are listed below.

1. Introduction
 - a. The Data Deluge
 - b. The Rationale for Linked Data
 - c. Intended Audience
 - d. Introducing *Big Lynx Productions*
1. Principles of Linked Data
 - a. The Principles in a Nutshell
 - b. Naming Things with URIs
 - c. Making URIs Deferencable
 - d. Providing Useful RDF Information
 - e. Including Links to other Things
3. The Web of Data
 - a. Bootstrapping the Web of Data
 - b. Topology of the Web of Data
4. Linked Data Design Considerations
 - a. Using URIs as Names for Things
 - b. Describing Things with RDF

- c. Publishing Data about Data
- d. Choosing and Using Vocabularies
- e. Making Links with RDF
- 5. Recipes for Publishing Linked Data
 - a. Linked Data Publishing Patterns
 - b. The Recipes
 - c. Additional Approaches to Publishing Linked Data
 - d. Testing and Debugging Linked Data
 - e. Linked Data Publishing Checklist
- 6. Consuming Linked Data
 - a. Deployed Linked Data Applications
 - b. Architecture of Linked Data Applications
 - c. Effort Distribution between Publishers, Consumers and Third

As it can be seen at a first glance, the EUCLID's curriculum covers all of the topics suggested by PlanetData, while some of the sections are covered in further depth. As work on the curriculum and the individual modules progresses, we will aim to develop a more precise mapping of the covered topics.

3.2 Video Lectures and Summer Schools

There are a number of videolecture organisations and summer schools that cover Linked Data topics as part of their presentation and training curricula. The list of events and presentations that are especially devoted to Linked Data is rather short, however, we were able to identify a number of courses that include this topic.

One main source of video-based training materials is available through videolectures¹. Videolectures has currently about 40 presentations and lectures on Linked Data, which can be analysed in order to determine the covered topics. One benefit of this source of training materials is that the presentations cover very introductory topics, as well as very advanced ones, and sometimes even provide discussions on controversial topics in the field. Therefore, it can be observed how the content of the lectures is adjusted to the target audience and which topics are relevant for the particular target groups.

The need for training on Linked Data principles and technologies has also been recognized by a number of summer schools for graduate and undergraduate students. Some of the training initiatives for young researchers that we identified are:

- ESWC Summer School²
- SUMMER SCHOOL ON ONTOLOGY ENGINEERING AND THE SEMANTIC WEB (SSSW)³
- LOD2 - Creating Knowledge out of Interlinked Data summer school⁴

Each of these summer schools includes Linked Data as part of their training topics and it would be helpful to analyse the level of detail of the presentations as well as the expected prerequisite skills and aimed for level of gained expertise at the end of the course.

¹ <http://videolectures.net/>

² <http://summerschool2012.eswc-conferences.org/>

³ <http://sssw.org/2012/>

⁴ <http://summerschool2012.eswc-conferences.org/>

3.3 Professional Training Curricula

Naturally, the importance of training in Linked Data technologies for professionals, who deal with publishing and managing structured data on the Web, has also been recognised. Some of the companies that provide trainings in this area are:

- Semantic Web Company⁵
- Semsphere⁶
- Ontotext⁷

We have examined in more detail the Semsphere training curriculum that offers three main levels of training, including specialist level, professional level and academy trainer level. The individual topics that are covered in each of the trainings can be seen below.

1. Specialist Level
 - a. Introduction to Semantic Technologies
 - b. Linked Data
 - c. Ontologies
2. Professional Level
 - a. Advanced Languages: OWL and RIF
 - b. Semantic Annotation
 - c. Dynamics, APIs and Services
3. Academy Trainer Level
 - a. Advanced Ontologies
 - b. Advanced Reasoning
 - c. Semantic Application Development
 - d. Training and Certification Methodology

It is important to point out that Linked Data is taught as part of the specialist level, which covers the fundamentals of semantic technologies and serves as the basis for the following training levels. Therefore, in the context of this curriculum, Linked Data is considered as one of the essential topics that have to be covered before the trainees can proceed onto more advanced or specialised subjects.

Furthermore, in addition to Linked Data training initiatives, video lectures, summer schools, and professional training courses we also plan to take into consideration further materials, including books (such as [7]) and online forums with trainings and hands-on examples. A detailed exploration and analysis of existing training curricula will help us to improve the presentation and structuring of the topics, described in this deliverable, while the use of presentation media and examples will aid in determining the most suitable way of communicating the curriculum content.

⁵ <http://www.semantic-web.at/>

⁶ <https://www.semsphere.com/>

⁷ <http://www.ontotext.com/>

4 Conclusion

With the growing importance and use of Linked Data principles and technologies, there is also an increased demand for trained data practitioners who are able to develop Linked Data-based solutions. The EUCLID project addresses precisely this need by providing an extensive training curriculum that communicates the fundamental background knowledge but also introduces some advanced and expert-level topics.

This deliverable presents the first draft of the EUCLID curriculum plan. It is based on seven main modules, which aim to gradually build up the trainee's knowledge in the field. The main covered topics include – Introduction to Linked Data and Application Scenarios, Querying Linked Data, Publishing, Interlinking and Cleansing Linked Data, Visualizing Linked Data, Analysing and Mining Linked Data, Creating Mashups and Linked Data Application, and Scaling-up. The modules are grouped into three main levels of topics – introductory, advanced and expertise. The curriculum aims to be practice and hands-on oriented, therefore examples, self-assessment questions and demo application are an important part of the presented materials.

EUCLID follows an approach for content delivery and revising, based on improving the created training materials by gathering comments and feedback, in order to produce really high-quality teaching courses. Therefore, the curriculum plan in its current form, as well as the content of the individual modules will undergo a number of further revisions. However, this deliverable captures the main topics and content that are to be covered by EUCLID.

References

- [1] <http://seevl.net/>
- [2] <http://www.bbc.co.uk/music/reviews/>
- [3] <http://etree.linkedmusic.org/about/>
- [4] <http://musicontology.com/>
- [5] <http://musicbrainz.org/>
- [6] Planet Data training curriculum [http://www.planet-data.eu/sites/default/files/pr-material/deliverables/D6.1 Training curriculum.pdf](http://www.planet-data.eu/sites/default/files/pr-material/deliverables/D6.1_Training_curriculum.pdf)
- [7] Tom Heath and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.