Deliverable 1.1.2

Draft Curriculum Delivery

Coordinator: Maria Maleshkova
With contributions from: Maribel Acosta, Alexander Mikroyannidis
Quality Assessor: Elena Simperl

<table>
<thead>
<tr>
<th>Editor:</th>
<th>Maria Maleshkova, KIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliverable nature:</td>
<td>Other (O)</td>
</tr>
<tr>
<td>Dissemination level: (Confidentiality)</td>
<td>Public (PU)</td>
</tr>
<tr>
<td>Contractual delivery date:</td>
<td>30.04.2012</td>
</tr>
<tr>
<td>Actual delivery date:</td>
<td>30.04.2012</td>
</tr>
<tr>
<td>Version:</td>
<td>1.0</td>
</tr>
<tr>
<td>Total number of pages:</td>
<td>30</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Curriculum plan, training modules</td>
</tr>
</tbody>
</table>
Executive summary

Current developments on the Web are marked by the increasing importance and application of the Linked Data technologies, which are establishing themselves as an innovative way for publishing, interlinking and exploring datasets. Until now the main adopters have been research organisations, governmental institutions and a limited number of companies. However, with the increasing number of interested parties, committing to use Linked Data core principles, and through the plenitude of applications build on top of the already available data, it seems that the time is right for the wider industrial adoption of Linked Data. EUCLID aims to support this development by addressing the need for trained data practitioners who are able to apply a Linked Data approach as part of their data business solutions.

In particular, one of the main objectives of the project is to provide and extensive training curriculum, covering the main technologies, tools, use cases and skills that need to be acquired in order to complete both basic, as well as more complex, tasks related to dealing with Linked Data. This deliverable described the second updated version of the curriculum, including the adjusted content of the individual modules, the different learning methods and materials, the targeted skills level that is to be achieved, and the expected competencies.

The second part of the deliverable details the curriculum plan. This is the second version of the curriculum, which includes restructuring of the modules, rescaping of the content and refining the examples, tools and exercises. The updated curriculum is divided into six modules, covering three levels of topics – introductory, advanced and expertise. We continue to follow the approach of having modules build on each other, adding more specific knowledge such as visualisation approaches or building applications on top of Linked Data. The redefined six modules are – Introduction to Linked Data and Application Scenarios, Querying Linked Data, Providing Linked Data, Interaction with Linked Data, Creating Linked Data Applications, and Scaling-up. Some of the main changes here were the unification of the previous Visualizing Linked Data, and Analysing and Mining Linked Data into the Interaction with Linked Data module. The scope of some of the individual modules was also redefined, since the main aim is to support data practitioners in completing different tasks and not simply to teach certain technologies or tools. Furthermore, improvements were done based on feedback and comments, gathered after sharing and presenting the curriculum. However, the main driving factor was the practical orientation of the content, examples, and exercises in each module, and of the curriculum as a whole.

The third part of this deliverable introduces a number of related programs and initiatives, which aim to provide training on Linked Data and data in general. We aim to identify such existing courses and curricula, determine the main topics that they address and identify overlaps with the EUCLID curriculum. This will not only help us to make sure that we have not missed any relevant content but will also contribute toward the objectivity and completeness in terms of the covered topics. In addition, we can better align with further existing initiatives in the field, initiate collaborations and plan the organisation of the training events accordingly. This deliverable presents a few new curricula, which were not yet introduced in D1.1.1.

The final section of the deliverable includes a summary of related training events, courses and organisation. This information helps us not only to determine, which are currently the most popular topics but also helps us to identify the main players, which might also turn out to be prospective collaboration partners. Furthermore, we can track their activities and keep the here presented curriculum up-to-date and aligned with commonly offered training events and courses.
**Abstract (for dissemination)**

One of the main objectives of EUCLID is to provide and extensive training curriculum, covering the main technologies, tools, use cases and skills that need to be acquired in order to complete both basic, as well as more complex, tasks related to dealing with Linked Data. This deliverable described the second updated version of the curriculum, including the adjusted content of the individual modules, the different learning methods and materials, the targeted skills level that is to be achieved, and the expected competencies.

**Keywords**

Curriculum plan, training modules
Table of Contents

EXECUTIVE SUMMARY ............................................................................................................. 2
DOCUMENT INFORMATION ........................................................................................................ 3
TABLE OF CONTENTS ................................................................................................................. 4
LIST OF FIGURES AND TABLES ............................................................................................... 5
ABBREVIATIONS ...................................................................................................................... 6

1 INTRODUCTION ...................................................................................................................... 7
  1.1 TARGET AUDIENCE AND GOAL LEVEL OF KNOWLEDGE – REVISED ................................................. 7
  1.2 SKILLS AND KNOWLEDGE PREREQUISITES – REVISED .................................................................... 8
  1.3 DELIVERY METHODS AND MATERIALS – REVISED ........................................................................... 8
  \hspace{0.5cm} 1.3.1 Delivery Methods ...................................................................................................................... 8
  \hspace{0.5cm} 1.3.2 Training Materials ................................................................................................................... 9
  \hspace{0.5cm} 1.3.3 Revised Materials Production Process .................................................................................. 10

2 REVISED STRUCTURE OF THE CURRICULUM .......................................................................... 12
  2.1 INTRODUCTION AND APPLICATION SCENARIOS ........................................................................ 12
  \hspace{0.5cm} 2.1.1 Module Detailed Outline ........................................................................................................... 13
  \hspace{0.5cm} 2.1.2 Available Materials .................................................................................................................. 13
  2.2 QUERYING LINKED DATA .............................................................................................................. 15
  \hspace{0.5cm} 2.2.1 Module Detailed Outline ........................................................................................................... 15
  \hspace{0.5cm} 2.2.2 Available Materials .................................................................................................................. 15
  2.3 PROVIDING LINKED DATA ............................................................................................................ 16
  \hspace{0.5cm} 2.3.1 Module Detailed Outline ........................................................................................................... 16
  \hspace{0.5cm} 2.3.2 Available Materials .................................................................................................................. 17
  2.4 INTERACTION WITH LINKED DATA ................................................................................................. 17
  \hspace{0.5cm} 2.4.1 Module Detailed Outline ........................................................................................................... 18
  2.5 CREATING LINKED DATA APPLICATIONS ................................................................................... 19
  \hspace{0.5cm} 2.5.1 Module Detailed Outline ........................................................................................................... 19
  2.6 SCALING UP .................................................................................................................................. 19
  \hspace{0.5cm} 2.6.1 Module Detailed Outline ........................................................................................................... 19

3 ALIGNMENT WITH RELATED TRAINING ACTIVITIES ................................................................ 21
  3.1 PLANETDATA TRAINING CURRICULUM – REVISED ....................................................................... 21
  3.2 OPEN DATA INSTITUTE (ODI) ....................................................................................................... 22
  3.3 LEAN SEMANTIC WEB .................................................................................................................. 24
  3.4 FURTHER TRAINING INITIATIVES .................................................................................................. 26
  3.5 VIDEO LECTURES, SUMMER SCHOOLS, TUTORIALS AND RELATED EVENTS ................................... 28

4 CONCLUSION .......................................................................................................................... 30
List of Figures and Tables

Figure 1: Revise Materials Production Process ........................................................................................................... 10
Figure 2: Curriculum Structure ........................................................................................................................................... 12
Figure 3: EUCLID Curriculum Plan ................................................................................................................................. 14
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

Current developments in the context of sharing data on the Web are marked by the growing importance and use of Linked Data, which is becoming a de-facto standard for publishing and accessing structured data on the Web. This trend is supported by the increased number of governmental organizations, research institutes, but also companies, which deal with data exchange, manipulation and maintenance in their daily business activities. As a result, there is a 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.

This deliverable presents a revised version of the initially introduced curriculum, which includes restructuring of the modules, rescooping of the content and refining the examples, tools and exercises. The updated curriculum is divided into six modules, covering three levels of topics – introductory, advanced and expertise. We continue to follow the approach of having modules build on each other, adding more specific knowledge such as visualisation approaches or building applications on top of Linked Data. The redefined six modules are – Introduction to Linked Data and Application Scenarios, Querying Linked Data, Providing Linked Data, Creating Linked Data Applications, and Scaling-up. Some of the main changes here were the unification of the previous Visualizing Linked Data, and Analysing and Mining Linked Data into the Interaction with Linked Data module. The scope of some of the individual modules was also redefined, since the main aim is to support data practitioners in completing different tasks and not simply to teach certain technologies or tools. Furthermore, improvements were done based on feedback and comments, gathered after sharing and presenting the curriculum. However, the main driving factor was the practical orientation of the content, examples, and exercises in each module, and of the curriculum as a whole.

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.

In the following sections we revisit and, if necessary, redefine the targeted audience, the goal level of knowledge, the skills and knowledge prerequisites, the delivery methods and the materials.

1.1 Target Audience and Goal Level of Knowledge – Revisited

As stated in the first version of the curriculum description, the main 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. This also motivates the practical orientation of the modules and the use of directly appreciable examples. However, throughout the first year of the project we discovered that the materials are use by a very broad audience, including researchers, students, professionals, managers, technology experts, ect. Overall, the curriculum and the trainings can be of benefit for anyone who aims to gain a broader and deeper understanding of how to deal with and manage data, in accordance with Linked Data principles.

However, it is important to point out that EUCLID’s training materials focus more on providing real business application examples and relating the learnt topics to those example. In particular, 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.

Regarding, the targeted skills and knowledge that are to be achieved, the curriculum continues to provide three main levels of expertise. 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. The benefit of having such a progressive structure of the modules is that instead of trying to bring all trainees to the same level, the curriculum can be adapted in order 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.

1.2 Skills and Knowledge Prerequisites – Revised

Based on the gained experience, some previous knowledge in IT development and engineering is very useful. This makes it easier to grasp the main concepts, the application functions and the presented approaches. We also found out that examples help a lot in compensating for lacking expertise. For the more advanced modules, some knowledge in the corresponding fields is very useful. Naturally, the precise pre-requirements depend on the candidate’s goals and on the training context. However, overall, so far we have not identified any specific boundaries that would hinder the use of the training materials and restrict it to only specific target audience.

1.3 Delivery Methods and Materials – Revised

In this section we share the experience that we have gained in terms of improving and revising the delivery methods and materials of EUCLID’s curriculum. In particular we reflect on the new process of generating training content and on the different channels used to disseminate the materials. The curriculum plan is based on a set of delivery methods and materials that complement each other and updates of the curriculum influence both the format and the ways of communicating the content.

1.3.1 Delivery Methods

Currently the curriculum foresees two main methods for delivering the learning content – via online channels and directly by a professional trainer. Given that 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. Furthermore, my making the materials available in a series of online channels, the achieved outreach is much greater and really anyone interested in the topic can benefit. For example, this is clearly visible from the viewer’s statistic of the webinar for module 3 on providing Linked Data, which demonstrate that the participants were spread all over Europe, included the US, Brazil and southern Asia.

Online communication channels, such as platforms for sharing slides, videos or complete training courses are very useful for supporting self-training and distance learning. Self-training requires individual engagement, where the candidate explores and learns the course material by himself/herself. There is usually no direct trainer support and onsite tutorial. The course material are accessible and acquired online, via the EUCLID’s website and 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 contrast, distance-learning provides more guidance to the students in terms of the learning plan but also the support in terms of interaction with trainers or gathering feedback. EUCLID does not plan to organise a distance-learning event, however, the materials are well suited to be used as a basis for such a course. In particular, the combination of the live webinars, guided tutorials, and the official course materials can easily support such a learning approach. As previously pointed out, EUCLID aims to shift discussions, answering of questions, and feedback collection from the restricted interaction of the trainer with the students to the openly available and growing Linked Data online community. As a result, the produced training materials in combination with the community interaction provide a pseudo distance-learning platform, available to a very wide audience.

EUCLID’s materials have also proven to be very useful as part of on-site trainings. The trainings were conducted both for professionals, as well as for people with more of a research background. Further details on the training events, the corresponding topics, audience and expected impact are available in Deliverable 2.3.2 “Interim real-world community engagement report”.

8
1.3.2  Training Materials

Both the content-generation process, as well as the format of the training materials themselves, have undergone some improvement and reshaping during the initial ten months of the project. In particular, currently EUCLID generates the following materials for each of the curriculum’s modules:

- **Description and detailed outline of the module**

  The description of the module as well as the detailed outline serve as a starting point for specifying the content, the technologies, tools and examples that are to be covered by the module. These are a short and concise way of introducing the covered topics and providing a short overview of the content. This type of content is reused on the website, for describing presentations on SlideShare\(^1\) and for collecting feedback on the envisioned structure of the materials before it is finalized. Currently, the descriptions of all modules are available and the first four modules also have detailed outlines.

- **Related Materials and further reading**

  These materials provide pointers to relevant presentations, videos, books and articles that can be used to deepen the acquired knowledge. We also include sources that were used as a basis for the content of some of the chapters, since the trainees should be aware of the standard literature in the field. Currently, the first four modules have a list of related materials.

- **Examples and exercises**

  The examples are included directly as part of the slides and the written chapter in order to directly demonstrate how the learned theory can be used in practice. In fact, some of the topics can be effectively conveyed only with the help of good examples, such as for instance “Querying Linked Data with SPARQL\(^2\)”. The exercises are usually at the end of each module and are part only of the complete multimedia versions of the content. In contrast, the examples occur in the slides, are discussed in the webinars and are naturally a part of the website and the textual materials. The exercises are an effective means for self-assessment and can be both simple questions and interactive tasks, requiring to apply some of the learned principles or to use introduced tools. For instance, the introductory chapter has multiple choice questions, while the module on querying Linked Data requires the running of SPARQL against an available SPARQL endpoint. Until now, we have completed the examples for the first three chapters and are working on the fourth one.

- **Presentation slides**

  Each module has a set of PowerPoint slides, which can be used for doing trainings. In fact, these are also the slides that we use to conduct the webinars. The slides cover the topic in detail and are available both on the website and on EUCLID’s SlideShare channel. The slide decks of the first three chapters are available online and we are in the process of completing the fourth one.

- **Webinars**

  The webinars are conducted based on the slides for each module. The webinars are broadcasted live from the Podium facility of the Open University\(^2\), after which a recording is made available both from the same facility and also through the EUCLID channel in Vimeo\(^3\). Furthermore, parts of the webinars are integrated directly in the versions of the module available as eBooks. We just recorded the third public webinar on publishing Linked Data, which was very well received.

- **Screencasts of tools**

  The EUCLID screencasts consist of short clips (2-3 minutes) that provide a quick overview and a walkthrough of a representative set of tools and platforms related with the EUCLID modules. The screencasts are made available in the EUCLID Vimeo channel and are also included in the eBook chapters in order to better explain the relevant

---

1. [http://www.slideshare.net/EUCLIDproject](http://www.slideshare.net/EUCLIDproject)
2. [http://stadium.open.ac.uk/podium](http://stadium.open.ac.uk/podium)
3. [https://vimeo.com/euclidproject](https://vimeo.com/euclidproject)

- **eBooks**

The EUCLID’s eBooks have a multimedia format, including detailed text description, embedded videos, images, exercises and they encompasse all the content for each module in a structured and interactive way. The eBook serves as the basis for self-learning, as well as for revisiting certain topics after a training is completed. In order to maximise the impact of the training materials on the community and bring them closer to as many people as possible, the eBook chapters are made available for a variety of platforms and formats including:

- Web browsers (HTML format)
- Apple iPad (iBook format)
- eReaders (ePUB format)
- Amazon Kindle devices (AZW3 format)

Until now, Module 1 has been released in all the above formats. Module 2 will be released shortly and Chapter 3 will be released in May of this year. 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.

### 1.3.3 Revised Materials Production Process

Based on the gained experience during the initial months of the project, we have revised and improved the materials production process (see Figure 1). In particular, we have implemented two main changes. First, the recording of the first webinar takes place before the production of the first version of the HTML chapter. This turned out to be necessary, since the slides and the other available materials are not enough to create a complete textual description. This also ensures that the content of the slides and the webinar are well aligned with the eBook descriptions. Second, we allow for enough time for quality assurance by domain experts. In this way we eliminate mistakes that might have occurred and make sure that we have covered all topics relevant for the module.

**Figure 1: Revise Materials Production Process**

The steps of the revised materials production process are the following:

1. Defining the detailed outline of the chapter and defining the content to be covered
2. Collect related materials, announce the dates for the webinars and the release
3. Discuss and finalise the proposed content
4. Create finalised outline, exercises and examples
5. First version of the slides
6. Internal webinar
7. Second version of the slides, first version of the HTML chapter, internal quality assurance
8. Finalise exercises and examples
9. Finalise slides, finalise HTML chapter
10. Public webinar
11. Public release

EUCLID implements an approach of revising the training content based on gathered feedback. For example, after the initial webinar, comments and suggestions are gathered from the audience. Furthermore, each draft of the modules is made available online and community feedback is actively be gathered. As a result, the final version of the training materials has undergone at least two rounds of improvements and revision.
2 Revised Structure of the Curriculum

In this section we introduce the revised curriculum, including details on the encompassed six modules. In particular, we updated the overall structure of the curriculum as well as some of the individual topics within the chapters.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Introduction and Application Scenarios</td>
</tr>
<tr>
<td>2.</td>
<td>Querying Linked Data</td>
</tr>
<tr>
<td>3.</td>
<td>Providing Linked Data</td>
</tr>
<tr>
<td>4.</td>
<td>Interaction with Linked Data</td>
</tr>
<tr>
<td>5.</td>
<td>Creating Linked Data Applications</td>
</tr>
<tr>
<td>6.</td>
<td>Scaling up</td>
</tr>
</tbody>
</table>

Figure 2: Curriculum Structure

Figure 2 shows the new structure of EUCLID’s curriculum, covering one introductory module, three basic modules, and two advanced modules. The main difference is that the previous chapters on analysing Linked Data and on visualising and mining Linked Data are currently unified into module 4 on interaction with Linked Data. While progressing with the content of the first few modules, we realised that this was necessary in order to be able to cover focused and at the same time practically oriented technologies and tools. We did not want to go into much detail in the individual areas but at the same time wanted to find a balance between the necessary level of explanation and covering the actually required knowledge and skills. The so defined module on interaction with Linked Data covers the topics of browsing, searching, visualizing and analyzing data and delivers the main techniques and tools relevant for completing tasks in the area.

The following sections describe each of the modules in more detail, providing a summary of the content, detailed outline of the topic, and a list of the already available materials.

2.1 Introduction and Application Scenarios

This module introduces the main principles of Linked Data, the underlying technologies and background standards. It provides basic knowledge for how data can be published over the Web, how it can be queried, and what are the possible use cases and benefits. As an example, we use the development of a music portal (based on the MusicBrainz® dataset), which facilitates access to a wide range of information and multimedia resources relating to music. The module also includes some multiple-choice questions in the form of a quiz, screencasts of popular tools and embedded videos.

The materials of the first module underwent a number of iterations, especially because the most effective process of developing the materials needed to be developed. Furthermore, we were initially unaware of some of the interdependencies between the materials, such as for example that the first webinar needs to take place before the first written version of the module can be created. This is also probably the reason why the introductory module took quite some time to prepare and each of the following modules progressed quicker. In summary we discovered that alignment between the different types of materials is very important and some need to be developed in parallel (for example, the slides and the examples), while others need to be completed sequentially, since one is a prerequisite for the others. Furthermore, quality assurance by independent readers turned out to be extremely helpful, contributing towards very good overall results.

Following is a detail outline of the module as well as a summary of all the available materials.

4 http://musicbrainz.org
2.1.1 Module Detailed Outline

All the materials to the first module are available at [http://www.euclid-project.eu/#chapter1](http://www.euclid-project.eu/#chapter1). The following listing describes the covered topics in detail.

1.1 Introduction
1.2 Motivation of the Course
1.3 Background Technologies
   - Internet
   - Hypertext
   - WWW
   - Web 1.0 (static)
   - Web 2.0 (dynamic)
   - Social Web
   - Web 3.0 (semantic)
   - Ontologies
1.4 Background Standards
   - HTTP
   - URI
   - XML
   - RDF
   - RDFS
   - OWL (OWL 2 Full, OWL 2 DL, OWL 2 EL, OWL 2 QL, OWL 2 RL)
   - SPARQL
1.5 Linked Data
   - Linked Data Principles
   - Rating Published Datasets
   - Growth of Linked Data on the Web
1.6 Case Scenario: a Music Portal
1.7 Examples
   - Marbles
   - Sigma
   - DBpedia Mobile

2.1.2 Available Materials

The introductory module is the first module, for which all materials are already available. Figure 3 provides an overview of all modules and the current state of completed slides, webinars and eBook chapters. As it can be been currently we are starting with the development of module four, focusing on visualising, searching, browsing and analysing Linked Data. All the materials to module 1: Introduction and Application Scenarios are available at:

- Outline
  - [http://www.euclid-project.eu/resources/curriculum](http://www.euclid-project.eu/resources/curriculum)
- Slides
- Webinar and screencasts
  - Part I [https://vimeo.com/61612182](https://vimeo.com/61612182)
  - Part II [https://vimeo.com/61612378](https://vimeo.com/61612378)
EUCLID 296229

- O Sig.ma https://vimeo.com/57931687
- O Data.gov.uk https://vimeo.com/57931688
- O BBC Music https://vimeo.com/57931689
- O MusicBrainz https://vimeo.com/57935375

- Exercises
  - O Quiz http://www.euclid-project.eu/node/30/take
- eBook chapter
  - O HTML http://www.euclid-project.eu/modules/chapter1
  - O iBook http://www.euclid-project.eu/sites/default/files/resources/Chapter1.ibooks
  - O ePUB http://www.euclid-project.eu/sites/default/files/resources/Chapter1.epub
  - O Kindle http://www.euclid-project.eu/sites/default/files/resources/Chapter1.azw3

Figure 3: EUCLID Curriculum Plan
2.2 Querying Linked Data

Querying Linked Data is the first of the two advanced topic modules. This module looks in detail at SPARQL (SPARQL Protocol and RDF Query Language) and introduces approaches for querying and updating semantic data. It covers the SPARQL algebra, the SPARQL protocol, and provides examples for reasoning over Linked Data. The module uses examples from the music domain, which can be directly tried out and run over the MusicBrainz dataset. This includes gaining some familiarity with the RDFS and OWL languages, which allow developers to formulate generic and conceptual knowledge that can be exploited by automatic reasoning services in order to enhance the power of querying.

2.2.1 Module Detailed Outline

All the materials to the second module are available at http://www.euclid-project.eu/#chapter2. The following listing describes the covered topics in detail.

2.1 Introduction and Motivation Scenario

2.2 SPARQL Terminology

2.3 Querying and Updating Linked Data with SPARQL

- Introduction to SPARQL
- Querying Linked Data with SPARQL
  - Query forms: ASK, SELECT, DESCRIBE, CONSTRUCT
  - Query patterns: BGP, UNION, OPTIONAL, FILTER
  - Sequence modifiers: DISTINCT, REDUCED, ORDER BY, LIMIT, OFFSET
- Updating Linked Data with SPARQL 1.1
  - Data management: INSERT, DELETE; DELETE/INSERT
  - Graph management: LOAD, CLEAR, CREATE, DROP, COPY/MOVE/ADD
- SPARQL Protocol: query operation, update operation

2.4 Reasoning over Linked Data

- SPARQL 1.1 entailment regimes
- RDFS entailment regimes, lacks of consistency check, inference limitations
- OWL properties, property axioms, axioms, class constructions

2.2.2 Available Materials

For the second module, almost all materials are available. We are still in the process of preparing the final version of the eBook chapters. The currently completed materials to module 2: Querying Linked Data are available at:

- Outline
  - http://www.euclid-project.eu/resources/curriculum
- Slides
  - http://www.slideshare.net/EUCLIDproject/querying-linked-data
- Webinar and screencasts
  - Part I https://vimeo.com/61618438
2.3 Providing Linked Data

The work on this module is not as advanced as on the first two ones. We have just completed the recording of the public webinar. This module covers the whole spectrum of Linked Data production and exposure. After a grounding in the Linked Data principles and best practices, with special emphasis on the VoID vocabulary, we cover R2RML, operating on relational databases, Open Refine, operating on spreadsheets, and GATECloud, operating on natural language. Finally we describe the means to increase interlinkage between datasets, especially the use of tools like Silk.

We also paid special attention to providing examples with supporting tools and recorded a number of screencasts for this purpose.

2.3.1 Module Detailed Outline

All the materials to the third module are available at http://www.euclid-project.eu/#chapter3. The following listing describes the covered topics in detail.

3.1 Introduction and Motivation

3.2 Linked Data Lifecycle
   - Linked Data Principles
   - Tasks for Providing Linked Data

3.3 Creating Linked Data
   - Data extraction, giving names (URIs), selecting vocabularies

3.4 Interlinking Linked Data
   - Link discovery
   - Manual interlinking, automatic interlinking
   - Interlinking with SKOS

3.5 Publishing Linked Data
   - Describing dataset with metadata (VoID)
   - Making the dataset accessible (dereferencing HTTP URIs, RDF dump, SPARQL endpoint, RDFa)
   - Making the dataset searchable (search engine support)
   - Exposing the dataset in repositories (creating new ones - CKAN, using the Data Hub, the Linking Open Data Cloud)
3.6 Linked Data Publishing Checklist

3.7 Tools for Providing Linked Data
- OpenRefine: Extracting data from spreadsheets
- R2RML: Extracting data from RDBMS
- GATECLOUD: Extracting data from text
- CALAIS: Extracting data from text
- Silk: Interlinking data sets

2.3.2 Available Materials

For the third module, almost all materials are available. We are currently in the process of performing quality assurance of the written chapter and the final eBooks are supposed to be released in May. The currently completed materials to module 3: Providing Linked Data are available at:

- Outline
  - http://www.euclid-project.eu/resources/curriculum
- Slides
  - Module slides http://www.slideshare.net/EUCLIDproject/providing-linked-data
  - Presentation on Big Linked Data http://www.slideshare.net/EUCLIDproject/big-linked-data
  - Presentation on Creating Data Science Curriculum for Professionals http://www.slideshare.net/EUCLIDproject/data-science-curriculum-v-32
  - Presentation on Mapping Relational Databases to Linked Data http://www.slideshare.net/EUCLIDproject/r2-rml-londonsemweb201304
- Webinar and screencasts
  - Part I https://vimeo.com/64709409
  - Part II https://vimeo.com/64709408
  - OpenRefine (https://vimeo.com/62430786)

2.4 Interaction with Linked Data

We are currently focusing on preparing most of the materials of the fourth module. This module focuses on providing means for exploring Linked Data. In particular, it gives an overview of current visualization tools and techniques, looking at semantic browsers and applications for presenting the data to the end user. We also describe existing search options, including faceted search, concept-based search and hybrid search, based on a mix of using semantic information and text processing. Finally, we conclude with approaches for Linked Data analysis, describing how available data can be synthesized and processed in order to draw conclusions. The module includes a number of practical examples with available tools as well as an extensive demo based on analyzing, visualizing and searching data from the music domain. The fourth module is very practice oriented and all of the though technologies are backed up by examples and tools.
2.4.1 Module Detailed Outline

The current set of available materials can be found at http://www.euclid-project.eu/#chapter4. So far we have completed the module description and the detailed outline. We are currently in process of preparing the first set of slides and organising the rehearsal webinar. The following listing describes the covered topics in detail.

4.1 Introduction and Motivations

4.2 Linked Data Visualisation

- Visualisation Techniques
  - Visualizing the Linked Data Cloud
  - Requirement for Visualisation Tools
  - Visualizing Different Data Dimensions

- Existing Linked Data Visualisations
  - Simple bar and pie charts, histograms, line and scatterplots
  - Node-link tree and graph visualisations, in both 2D and 3D
  - Matrices, parallel co-ordinates
  - Timeline and topology plots, map and landscape views
  - Space-filling visualisations such as tree maps, rose diagrams, icicle, bubble and sunburst plots
  - Iconography, including star and glyph plots
  - Text-based

- Linked Data Browsers
  - sig.ma, sindice, OpenLink RDF Browser, Marbles, Disco - Disco Hyperdata Browser, Piggy Bank, part of SIMILE, Zitgist DataViewer, iLOD, URI Burner

- Browsers with Visualisation Options
  - Tabulator, IsaViz, OpenLink Data Explorer, RDF Gravity, RelFinder, DBpedia Mobile, LESS http://less.aksw.org
  - Further: SIMILE Exhibit, Haystack, FoaF Explorer, Humboldt, LENA, Noadster, mSpace, Revyv, RKBExplorer, Semanlink

- Visualisation toolkits
  - Information Workbench Linked Open Data, Graves

- SPARQL Visualisation

4.3 Linked Data Search

- Exploiting Linked Data Properties for Search
- Faceted Search and Browsing
  - Browsing in Taxonomies

- Concept-based Search and Hybrid Search
  - KIM, sig.ma, PoolParty

- Semantic Search Engines

- Searching for Semantic Data
  - sindice, Watson, Swoogle, Semantic Web Search Engine (SWSE)
4.4 Methods for Linked Data Analysis
- Features of Linked Data analysis
- Tools for Linked Data analysis

2.5 Creating Linked Data Applications

This module represents the first set of advanced topics, which are part of EUCLID’s curriculum. The fifth module gives details on technologies and approaches towards exploiting Linked Data by building Linked Data applications. In particular, it gives an overview of popular existing applications and introduces the main technologies that support implementation and development. Furthermore, it illustrates how data exposed through common Web APIs can be integrated with Linked Data in order to create mashups.

Similarly to the previous module, this module is very practice oriented and we are planning on building on the examples and demo system developed for Interacting with Linked Data, in order to demonstrate how applications can be designed and implemented on top of the available data.

2.5.1 Module Detailed Outline

Currently we are in the process of refining the outline of the module and specifying the covered topics, tools and technologies. The following listing describes the envisioned content in detail.

5.1 Using Linked Data with Web APIs
- Introduction Web APIs
  - Underlying technology basics
- Common Web APIs
- Graph Store protocol
- Federated query processing
- Combining Linked Data with Web API data

5.2 Linked Data Applications
- Motivation scenario (data integration based on mashups)
- Existing Linked Data applications
- Application directories

2.6 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. This module is centred around two main topics – Linked Data cloud deployment and big data processing. In particular, it addresses the main issues of Linked Data and scalability. The module gives details on approaches and technologies for clustering, distributing, sharing, and caching data. Furthermore, it addresses the means for publishing data through cloud deployment and the relationship between Big Data and Linked Data, exploring how some of the solutions can be transferred in the context of Linked Data.

2.6.1 Module Detailed Outline

Currently we have identified the following topics to be covered by the module.

6.1 Linked Data Clustering
- Distribution, Sharing
- Caching

6.2 Linked Data Cloud Deployment

6.3 Big Data Technologies Applied to Linked Data

Naturally, the final modules are not so detailed as the initial ones but as we progress and create more learning materials we will also reach a further level of refinement. We plan to complete the materials to all the modules six months before the end of the project, allowing for enough time for sharing and announcing the achieved results.
3  Alignment with Related Training Activities

In this section we continue to analyse and compare existing training activities and curricula in the area of Linked Data. Deliverable 1.1.1 provided an initial overview of some of the related activities, here we extend these and add some further ones.

With the growing use and application of the Linked Data principles and related technologies, it is only natural that there are already some teaching materials and courses on the topic. In order to be able to provide a curriculum that is up to date but that is also competitive in terms of the offered training, we identify related training activities and describe how they align with the courses offered by the project. 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.

3.1  PlanetData Training Curriculum – Revised

Work on the PlanetData training curriculum has progressed and, therefore, we shortly revisit the offered topics and the availed materials. The PlanetData project aims to establish a community of researchers that supports organizations in exposing their data in new and useful ways. In this context, in order 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.

The curriculum has not be revised since it was last published and continues to cover four main topics – Semantic Technology, Database Technology, Linked Data and Data Streams. What is also relevant is the fact that the training materials are available as part of a series of video lectures http://videolectures.net/planetdata_training_curriculum/

From the topics covered by the PlanetData curriculum, the Linked Data topic is the most relevant one for the here described curriculum. However, some of the remaining topics can be used to determine partial coverage of such as Semantic Technologies and module 1, which introduces all the main underlying principles. The complete Planet Data curriculum is available at5, while the covered Linked Data topics are listed below. Unfortunately the Linked Data section is not available in the form of videos as part of the videolectures.

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 Defererencable
    d.  Providing Useful RDF Information
    e.  Including Links to other Things

2.  The Web of Data
    a.  Bootstrapping the Web of Data

b. Topology of the Web of Data

3. 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

4. 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

5. Consuming Linked Data
   a. Deployed Linked Data Applications
   b. Architecture of Linked Data Applications
   c. Effort Distribution between Publishers, Consumers and Third

Since we were currently preparing the module on providing Linked Data, some on the topics from section 4. Recipes for Publishing Linked Data were taken into consideration. In particular, we created our own summary (checklist) of the steps covered in the process of making data available in the form of Linked Data. Section 1 through 3 were partially covered by the introductory module, where we describe the main relevant technologies, the underlying principles and the vision of the Web of Data.

Finally, section 5 on consuming Linked Data is somewhat similar to module 5, which described how to build applications on top of Linked Data. The main difference here is that we take a very practical approach by giving specific examples, set of tools and solutions that can directly be used. We do not focus so much on the architectural approach or the individual roles of the parties taking part in the data consumption and production.

Overall, the EUCLID curriculum represents a more practice-oriented approach towards covering some of the topics, which are part of the PlanetData curriculum. Each of our modules directly points out tools, technologies and approaches and goes in depth, when it comes to addressing specific challenges that a data practitioner might face. It seems that the PlanetData curriculum focuses on providing more insight and detail on the introductory topics and takes the two more advanced sections only up to a certain theoretical level. Therefore, we can conclude that it is foreseen for a rather broader and inexperienced audience, in contrast to the EUCLID’s curriculum, which addresses firstly the needs of data professionals, while still be in suitable for a more general training.

3.2 Open Data Institute (ODI)

One of the important developments that have taken place in the context of teaching and Linked Data is the launching of activities of the Open Data Institute (ODI)\(^6\). The ODI is an independent, non-profit, non-partisan, limited by guarantee company, which aims to “catalyse an open data culture that has economic, environmental and

\(^6\) http://www.theodi.org

22
social benefits”. In the context of this goal, the ODI offers a number of training events such as the Open Data in Practice7, which is a five-day course and the Introduction to Open Data for Journalists: Finding Stories in Data8.

The Open Data in Practice course is of particular relevance to EUCLID’s curriculum, since it provides a very business-oriented and practical introduction to open data. In particular, it gives an introduction to the technical, commercial and legal aspects of open data. Furthermore, it aims to support the uptake of the principles by highlighting key opportunities for working with open data and how they can be exploited across government, business and society. The course is designed to enable anyone to understand how to publish, consume and exploit open data as well as give an understanding of best practice, law, licensing and policy issues.

The description of the course states that it is targeted towards open data practitioners, policy officials and advisors, account and project managers, statisticians and analysts, strategists, entrepreneurs, business developers, ICT suppliers, knowledge managers, policy owners, developers, information architects, journalists, research and intelligence. Similarly to the EUCLID curriculum, no previous experience is required but good computer skills and familiarity with information technologies is of benefit.

The outline of the Open Data in Practice course is the following:

1. What is Open Data? What are the benefits? Considering personal data?
2. Let’s make some data. How does it work on the web?
3. Licensing, the law and best practice
4. From publishing to consuming open data, including open data standards
5. Tools for analysing data: cleaning, validating, and enriching data.
6. Analysing data continued: Establishing trust Visualising data
7. Business and open data: benefits, applications, value propositions
8. Innovating with open data
9. Hack Day in teams

A it can be seen the course is very practice oriented, talking specific topics such as licensing, law and best practice. Furthermore, it also includes implementation sessions every day, so that the participants can directly apply the theoretical knowledge. Section 2 is quite similar to module 3 on providing Linked Data. Furthermore, sections 5 and 6 relate to our module on interaction with Linked Data, including analysing and visualising data. What we have not reflected on in EUCLID’s curriculum is the framework around dealing with data, which includes legal issues and business models that can be applied. Therefore, it seems that the ODI addresses not simply data practitioners but also managers and decision makers. This is naturally quite important when designing a course that tries not only to provide an introduction to the technology but also to encourage its adoption. In contrast, our curriculum is focused on developing practical skills and not so much on motivating the need for Linked Data and demonstrating how it can be economically exploited.

As stated by the course description the objectives and expected competencies are the following:

- Have an overview of open data, law, web technologies and its application potential
- Have an understanding of the architecture and openness of the web
- Understand the considerations of publishing personal data
- Have a practical understanding of how to publish, consume, and exploit open data
- Understand processes required to release large data sets
- Have developed the ability to evaluate open data strategies
- Have an increased knowledge of how to commercialise and innovate using open data

7 http://www.theodi.org/courses/open-data-practice

8 http://www.journalism.co.uk/introduction-course-open-data-for-journalists/s382/
- Have worked with others to produce data
- Understand the vocabulary around data, such as linked data and the semantic web
- Have developed an ability to share and brief others on benefits of open data

These objectives once again emphasise the practical orientation of the course and also confirm that further topics such as the legal issues or business setting should be included in curricula that support training that supports the adoption of a particular technology.

It seems that the ODI is becoming very prominent in the area of open data and will continue to offer courses and trainings. Therefore, we will include an updated list of covered topics in the following versions of this deliverable.

### 3.3 Lean Semantic Web

Jie Bao\(^9\) offers a tutorial on Lean Semantic Web\(^{10}\) including a list of the covered topics. The curriculum is quite extensive and in addition to covering the semantic technologies, also provides a lot of details on data management, data storage and databases. Overall the sections are very data-centric and focused on explaining in detail the related technologies. A list of all the sections is provided below.

#### Section 1 Data Representation
- What is data and structure?
- Value of unstructured data
- Value of structured data
- Cost of data modeling
- The art of readable knowledge
- Naming resources
- Locating resources
- Relating resources
- Syntax: XML, JSON, YAML, RDF, Python etc.
- Implementation: some Python (RDFLib...)

#### Section 2 Databases
- Cost of modeling and indexing semantics in database
- Use relational DB for semantic modeling
- Document database: MongoDB, Elastic Search etc
- Graph database: TinkerPop stack, Neo4j, OrientDB
- Graph batch processing: Pregel, Hama, GraphChi etc
- Querying triples (with RDB, document db, graph db, or dedicated triple store)
- Implementation: some Python

#### Section 3 Search and Findability
- Database vs search engine
- Cost of inverted index
- Extend inverted index to model semantic relations
- Understanding user queries: from keywords to sentences
- Faceted search: Elastic Search and Solr
- Graph search
- Implementation: some Python (ESClient...)

#### Section 4 Data Exchange and Integration
- Portability
- Protocol Buffers and Thrift
- Email and MIME

---

\(^9\) [http://www.baojie.org/](http://www.baojie.org/)

\(^{10}\) [https://github.com/baojie/leansemanticweb/blob/master/Syllabus.md](https://github.com/baojie/leansemanticweb/blob/master/Syllabus.md)
- JSON-RPC
- XMPP and Google Wave Protocol
- REST API design
- Some most important data APIs
- Implementation: some Python (JSON...)

**Section 5 Inference**
- Cost of reasoner and index (IR, DB, KB)
- Just-in-time knowledge
- Practical rule modeling
- Inference as graph operations
- Inference using databases
- Inference using functional programming
- Implementation: some Python (Pydatalog, Fuxi)

**Section 6 Knowledge Extraction**
- Cost of knowledge extraction
- Data cleaning
- Structure extraction
- Shallow parsing
- Entity extraction
- Relation extraction
- Implementation: some Python (NLTK...)

**Section 7 Visualization**
- Cognitive background
- Exhibit and others
- D3 (and other JavaScript lib)
- NetworkX and (and other Python lib)
- Implementation: some Python (mplotlib...) and Javascript

**Section 8 User Interaction**
- It's about people, not machine
- Guided data exploration and discovery: why semantics is part of the solution
- Query formulating
- Faceted Browser
- Mobile search, Voice interface and personal assistants
- Implementation: some Python (ElasticSearch, Redis, RDF store based)

**Section 9: Big Data and Lean Data**
- Measuring semantics in data
- Small is beautiful in knowledge
- The rule of knowledge growth on Web
- Small knowledge: in-memory graph models
- Big knowledge: knowledge bases on clusters
- Big data
- Datasets: Freebase, DBPedia, LOGD, Factual etc.
- Platforms: EC2, and some others
- Implementation: some Python (boto, starcluster)

**Section 10 Lean Application Development**
- Build, measure, learn
- Lean Canvas
- MVP
- Build: mockup strategies, prototyping, prototyping
- Measure: key metrics, but not vanity metrics
- Learn: why^5
• Semantic Wordpress/Drupal/Wiki, etc,

As it can be seen, the list of covered topics is quite long and there are a lot of details related to the specific underlying technologies. Overall, the curriculum is targeted towards developers who have to deal with different challenges related to data. The sections are not necessarily based on the different data-related tasks that have to be completed but also based on the development that can be done for and with data (such as databases or applications).

This curriculum provided us with some interesting insights about topics or motivation that needed to be included in EUCLID’s curriculum as well. Section 1 is similar to module 1, giving an introduction to the basic underlying principles. The main difference is that we did not motivate the need for structured data and the resulting benefits in so much detail. This has again to do with the target audience of EUCLID’s curriculum and with the fact that it is very practically-oriented, thus assuming that the participants are already interested in Linked Data and do not have to be persuaded first. Section 2 is a bit out of the scope of our curriculum and there is no module that focuses explicitly on databases. Instead databases are mentioned as tools towards achieving certain tasks such querying, publishing, analyzing, visualizing, etc.

Section 3 was actually taken into account while determining the content of modules 2 and 4 on querying Linked Data and on interacting with Linked Data. Indexing plays an important part for supporting search and we included examples of different systems (e.g. Solr). Still the overall detail of the section is a bit more than we aim to cover while providing practical training on how to query and search Linked Data. The technologies mentioned in Section 4 are used as a basis for explaining the development of applications on top of Linked Data, which will be covered in EUCLID in module 5. Yet again, the level of detail here is not reflected in the module, since we only aim to describe the fundamentals and focus on the development scenarios.

Section 5 is included in module 2 as part of more complex ways of querying Linked Data, while Section 6 is partially reflected as part of the steps for providing Linked Data. Similarly to previous sections, our curriculum focuses on describing how Linked Data can be created and while knowledge extraction topics are relevant, we do not reflect on them in so much detail.

Sections 7 and 8 are covered as part of the module on interaction with Linked Data. EUCLID’s curriculum covers more tools and approaches than the here mentioned ones. In fact, we are aiming to provide a catalog of tools for manipulating Linked Data, including visualization. The module is very important, since it gives the basics for exploring the data and benefiting from it. Therefore this is one of the few cases where EUCLID will provide more details than the here discussed sections.

We envision to cover the topics listed in Sections 9 under module 6 on scaling-up. In fact this section was quite useful, since we were able to update our initial outline. Finally Section 10 will be integrated as part of module 5 on building applications on top of Linked Data.

Overall, this curriculum was very helpful in making sure that we have covered a sufficient set of topics, in the required details. It also became evident that it might be difficult to finding the balance between developing a curriculum based on theoretical approaches and one focused on implementation and specific technologies. By focusing on the tasks that need to be completed by data practitioners, we always aim to provide the required knowledge and skills in the corresponding modules.

3.4 Further Training Initiatives

In this section we describe some further training initiatives and materials, related to Linked Data and data in genera.

The Linked Data community website\footnote{http://linkeddata.org} itself provides a collection of reference tutorials and presentations. Unfortunately, it does not offer a training curriculum or a list of topics that are relevant and should be covered.
Instead under “Guides and Tutorials”\(^\text{12}\) there is a list of relevant documents, tutorials, slides and frequently asked questions. The provided resources are given below:

- **Key Reference Documents**
  - Design Issues: Linked Data, by Tim Berners-Lee
  - Linked Data: Evolving the Web into a Global Data Space, by Tom Heath and Christian Bizer

- **Textual Guides/Tutorials**
  - Linked Data: Evolving the Web into a Global Data Space, by Tom Heath and Christian Bizer
  - How to Publish Linked Data on the Web (Tutorial), by Chris Bizer, Richard Cyganiak, Tom Heath (superseded by the Linked Data book, by Tom Heath and Christian Bizer)
  - Introducing Linked Data and the Semantic Web
  - Introducing Graph Data
  - Introducing RDF/XML
  - Semantic Modelling
  - Introducing RDFS and OWL
  - Deploying Linked Data using OpenLink Virtuoso
  - Querying Semantic Data
  - Learn Linked Data - A growing collection of tutorials, essays, links and discussion about Linked Data and related topics.

- **Video Tutorials**
  - How to Publish Linked Data on the Web tutorial by Tom Heath, Michael Hausenblas, Chris Bizer, Richard Cyganiak, Olaf Hartig, from ISWC2008, Karlsruhe, Germany.
  - The Web, one huge database ... screen-cast by Michael Hausenblas (see also examples and slides).

- **Introductory Slide Sets**
  - Quick Linked Data Introduction, by Michael Hausenblas
  - An Introduction to Linked Data, by Tom Heath, from the Semantic Web Austin Linked Data tutorials
  - Tutorial on Linked Data - A Practical Introduction, by Michael Hausenblas
  - 30 Minute Guide to RDF and Linked Data, by Ian Davis, from code4lib2009

- **Frequently Asked Questions**
  - Linked Data FAQ at linkeddata.org
  - Linked Data FAQ (enterprise focus)

As it can be seen, there is no particular structure to the materials and they are not grouped according to topic or application area. Still, we used part of the resources as further reading in some of the modules, while others were helpful in defining the fundamentals.

Another collection of related training courses is given by the Data Science Academy (DSA)\(^\text{13}\). The Data Science Academy is a new project from Data Science London, which offers Data Science courses and workshops. Currently, the website provides a list of relevant courses, which are given below:

The Little List of Free Online Data Science Courses\(^\text{14}\)

\(^{12}\) [http://linkeddata.org/guides-and-tutorials](http://linkeddata.org/guides-and-tutorials)

\(^{13}\) [http://datascienceacademy.com](http://datascienceacademy.com)
As it can be seen, the courses are not specifically tailored to Linked Data but rather cover data topics in general. We have taken some input from the introduction to data science courses. Still currently, we cannot directly align EUCLID’s curriculum with the training offered by DSA. However, since this initiative is currently gaining on popularity, we will continue to monitor the developments and report on new curricula in the upcoming deliverables.

Another channel that is worth mentioning is Coursera\(^\text{15}\), which provides a collection of official university courses available online. Currently, there are no specific courses on Linked Data. However, training specific in the context of data is already being offered, for example through the Introduction to Data Science course\(^\text{16}\). We will continue to monitor this channel and report on any newly offered training in the area of Linked Data.

### 3.5 Video Lectures, Summer Schools, Tutorials and Related Events

In this section we cover further sources that might serve as input in improving and updating EUCLID’s curriculum and training materials. There are a number of online course organisations, tutorials at conferences 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.

As already mentioned we are following videolecures\(^\text{17}\) and Coursera for new training in the context of Linked Data. Relevant materials are also available through the Webinar series of the LOD2 project\(^\text{18}\).

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 researches that we identified are:

- 2\(^\text{nd}\) and upcoming 3\(^\text{rd}\) ESWC Summer School\(^\text{19}\)
- SUMMER SCHOOL ON ONTOLOGY ENGINEERING AND THE SEMANTIC WEB (SSSW)\(^\text{20}\)

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. We also have a list of key conferences in the area, which we are monitoring for tutorials and related training events:

\(^{15}\) [https://www.coursera.org/](https://www.coursera.org/)
\(^{16}\) [https://www.coursera.org/course/datasci](https://www.coursera.org/course/datasci)
\(^{17}\) [http://videolectures.net/](http://videolectures.net/)
\(^{19}\) [http://summerschool2013.eswc-conferences.org/](http://summerschool2013.eswc-conferences.org/)
- European Data Forum 2013
- ESWC 2013
- Social Linked Data for Science and Education - SOLISE 2013
- Semtech 2013
- ISWC 2013
- International Conference on Dublin Core and Metadata Applications

For example the ESWC 2013 offers a tutorial on analyzing and visualizing Linked Data with R\(^2\), while the EDF 2013 included a dedicated sidetrack on developing training curricula for data science.

\(^{21}\) [http://linkedscience.org/events/lodr2013/](http://linkedscience.org/events/lodr2013/)
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 an updated and refined version of EUCLID’s training curriculum. It is based on six main modules, which aim to gradually build up the trainee’s knowledge in the filed. The main covered topics include – Introduction to Linked Data and Application Scenarios, Querying Linked Data, Providing Linked Data, Interaction with Linked Data, Creating Linked Data Applications, 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 current state of the main topics and content, which are to be covered by EUCLID.