Development of a Linked Data curriculum

Tutorial at WWW 2014

Dr. Alexander Mikroyannidis
Knowledge Media Institute
The Open University, UK
The Open University

The Open University (OU) is the largest academic institution in the UK with:

• more than 250,000 students
• close to 7,000 tutors
• more than 1,200 full-time academic staff
• more than 3,500 support and administrative staff

Most OU courses are available throughout Europe and some are available worldwide.

Since its launch in 1969 more than 1.6 million people worldwide have achieved their learning goals by studying with the OU.
Agenda

• The EUCLID project
• The EUCLID curriculum & module production process
• The EUCLID learning materials
• Best practices for curriculum design & delivery
The EUCLID project

- A European project facilitating *professional training for data practitioners*, who aim to use Linked Data in their daily work.
- EUCLID delivers a curriculum implemented as a combination of *living learning materials and activities* delivered online & face-to-face.
- The EUCLID curriculum is *validated by the user community* through continuous feedback.
The EUCLID curriculum

A series of modules, each targeting a different crucial task related to Linked Data:

1. Introduction and Application Scenarios
2. Querying Linked Data
3. Providing Linked Data
4. Interaction with Linked Data
5. Creating Linked Data Applications
6. Scaling up
The EUCLID learning materials

- Presentation slides
- Webinars
- Screencasts
- Exercises & quizzes
- eBook
- Online courses
The EUCLID module production process

Collection of Raw Chapter Materials

Slides – First Version

Webinar – First Recording

Initial Version of eBook Chapter

Slides – Final Version

Pre-final Version of eBook Chapter

Webinar – Final Recording

Final eBook Chapter and online course
Presentation slides

• These are the first training materials produced for each module. They provide an overview of the main concepts covered in each module.
• They also contain an extensive set of examples, so that the concepts of the module are explained to practitioners more effectively.
Webinars

- The webinars are conducted based on the slides for each module.
- They are broadcasted live; the audience can ask questions via a chat facility.
- A recording is made available through the EUCLID channel in Vimeo.
Screencasts

• They demonstrate the use of tools and platforms related to the EUCLID modules, e.g.:
  – Sig.ma
  – Data.gov.uk
  – BBC Music
  – Seevl
  – MusicBrainz
  – Sesame
  – OpenRefine
  – …and more
Exercises

Exercise 1

RDF syntax: Turtl

RDF input:

```
@prefix vocab: <http://www.euclid-project.eu/examples/module1#> .
@prefix xsd: <http://www.w3.org/2000/01/XMLSchema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

vocab:consortiumMember rdf:type vocab:ProjectMember .

<Barry> a vocab:Person ;
    vocab:givenName "Barry" ;
    vocab:familyName "Barlow" .

<euclid rdf:label "The Euclid Project" .
    vocab:hasConsortiumMember ?member ;
    vocab:hasConsortiumMember "Barry" .

Vocabulary (RDF/XML) import URL:
https://xmlns.com/foaf/spec/index.rdf

Inference: RDF5

SPARQL query:

PREFIX csv: <http://www.euclid-project.eu/examples/module1#>
PREFIX vocab: <http://www.euclid-project.eu/ns#>
PREFIX xsd: <http://www.w3.org/2000/01/XMLSchema#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT * WHERE {
    ?member a vocab:Person ;
    vocab:givenName "Barry" ;
    vocab:familyName "Barlow" ;
    vocab:hasConsortiumMember "Barry" .

    ?member a vocab:ProjectMember ;
    vocab:ResearchProject vocab:hasConsortiumMember "Barry" .
}

SPARQL Query

Query:

1. Cf. Chapter 2, Slide 25
2. PREFIX dbpedia: <http://dbpedia.org/resource/>
5. PREFIX event: <http://purl.org/NET/c4dm/event.owl#>.
7. SELECT DISTINCT ?album_title ?record_duration
8. WHERE {
10.  ?release event:factor ?album ;
12.  {SELECT ?record (SUM(?track_duration) AS ?record_duration)
15.     ?track mo:duration ?track_duration } }
16.  GROUP BY ?record
17.  HAVING (?record_duration > 3600000)
18. }
19. ORDER BY DESC(?record_duration)
Question 1 of 10
Which of the following are true of knowledge represented in RDF (select all that apply)?

- A. The subject specifies the subject domain for the knowledge
- B. The object specifies any objections that may overturn the knowledge
- C. The predicate specifies the relationship between subject and object
- D. Relationships between any URI-identified resources can be specified

Question 2 of 10
Which of the following is among the bound results (select all that apply) when the following SPARQL query is executed over Graph 1?

- A. :a
- B. :b
- C. :c
- D. :d
- E. :e
- F. :f
eBook

• The EUCLID eBook encompasses the content for each module in a structured and interactive way.

• It is available for:
  – Web browsers (HTML format)
  – Apple iPad (iBook format)
  – Other tablets (ePUB format)
  – Amazon Kindle devices (MOBI format)
CHAPTER 1: INTRODUCTION AND APPLICATION SCENARIOS

Introduction

The last decade has seen a growing interest in the Semantic Web, which extends the web to include a machine understandable layer of data. This technology applies web-based standards for encoding datasets and linking them using a new layer of datasets, so that applications can exploit data from many different sources. It also provides a means of annotating data sets with additional knowledge in ontologies, allowing enhancements based on automatic reasoning (as an example).

This chapter introduces Linked Data and related semantic technologies, and shows how these can be used to build applications. As an example, we target the development of a music portal (based on the Midori project). This portal facilitates access to a wide range of information and multimedia resources relating to music.

Part I: Semantic Technologies and Linked Data Foundations

We will describe a set of technologies that allows datasets to be published over the web, and queried effectively by applications. Compared with search engines such as Google and Yahoo, which are based on text-based matching, these technologies are “semantic”. This means that information is represented not in a natural language like English or Spanish, but in a graph-based data model that facilitates extension, integration, inference and uniform querying. As a result:

- Which of the following are also true of RDF literals (select all that apply)?

Choose:

- Datatypes can be applied to literals
- Datatypes must be applied to literals
- Language tags can be applied to literals
- Language tags must be applied to literals
CHAPTER 2
Querying Linked Data
Online courses

• The EUCLID online courses provide an integrated overview, structured as a learning pathway, of all the learning materials produced in the project.

• Learners can study them at their own pace, as there is no predetermined start or end date.

• They are available for:
  – Web browsers (HTML)
  – iPad, iPhone and iPod touch (iTunes U)
4. Test your knowledge

How much have you learned from this learning pathway? Test your knowledge by completing the following quiz and exercise.

- Take the Quiz (10 multiple-choice questions).

Study the following RDF statements, expressed in the Turtle syntax, then attempt the exercises that follow. For question 3 you should use the Euclid SPARQL endpoint.

```turtle
@base <http://www.euclid-project.eu/examples/module1/> .
@prefix vocab: <http://www.euclid-project.eu/voc/> .
@prefix rdfs: <https://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

vocab:ResearchProject rdfs:subClassOf foaf:Group .
vocab:consortiumMember rdfs:subPropertyOf foaf:member .

<barry> a foaf:Person ;
foaf:givenName "Barry" ;
foaf:familyName "Norton" .

<euclid> rdfs:label "The Euclid Project"@en, "Das Projekt Euclid"@de ;
vocab:consortiumMember <barry> .
```

1. Re-express the statements in N-Triples (i.e. remove all prefixes and abbreviations to give full triples in absolute URI).
2. Add a resource representing yourself, attaching your name using the FOAF properties.
3. Execute the following SPARQL query and consider why the class "Agent" has members, even though none are explicitly asserted in the data: `SELECT ?agent WHERE { ?agent a foaf:Agent }`.
4. Add a property to "consortiumMember" to assert that all subjects should be research projects (members of the "ResearchProject" class).
5. Adapt the query from Q3 to ensure that "euclid" is now a research project.
6. Create a new property to relate training participants to research projects and use it to relate yourself with "euclid".

- Use the Exercise 1 form to try out your RDF and SPARQL.
Online course in iTunes U

<table>
<thead>
<tr>
<th>Learning outcomes</th>
<th>× 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating, Interlinking and Publishing Linked Data</td>
<td>× 1</td>
</tr>
<tr>
<td>Test your knowledge</td>
<td>1</td>
</tr>
<tr>
<td>Further reading</td>
<td>1</td>
</tr>
</tbody>
</table>

**Learning outcomes**

By the end of this learning pathway you should have an understanding of:

* The main stages in the Linked Data lifecycle from its creation through to its publication and use.

**Creating, Interlinking and Publishing Linked Data**

Learn what should be on your checklist for creating, interlinking and publishing Linked Data.

* Creating Linked Data
  - Watch Part I of the webinar 'Providing Linked Data'
  - View the slides of this webinar
  - Read Part I of Chapter 3 'Providing Linked Data'

**Linked Data Catalogs and Tools for Providing Linked Data**

Learn about Linked Data catalogs and the tools that can assist you with the creation and interlinking of Linked Data.

* Watch the Part II of the webinar 'Providing Linked Data'

08/05/2014
Best practices for curriculum design

• Industrial relevance
• Team curriculum design
• External collaboration
• Explicit learning goals
• Show realistic solutions
• Use real data & tools
• Show scalable solutions
• Eating our own dog food
Best practices for curriculum delivery

• **Open to format**
  Our learning materials are available in a variety of formats including: HTML and as an eBook, as an Apple iBook and on Amazon Kindles.

• **Addressibility**
  Every concept in our curriculum is URI-identified so that HTML and RDF(a) machine-readable content is available.

• **Integration**
  The main textual content, relevant webinar clips, screen casts and interactive components are placed into one coherent space.
Best practices for curriculum delivery

• **High quality**
  We have a formalised process where all materials go through several iterations to ensure quality, e.g. for each module we run both a practice and a full webinars facilitating critique and commentary.

• **Self-testing and reflection**
  In every module, we include inline quizzes formulated against learning goals enabling students to self-monitor their progress.
Download our learning materials:

www.euclid-project.eu

**eBook**
- HTML
- iBooks
- ePUB
- Kindle
- Amazon Kindle

**Online courses**
- HTML
- iTunes U

08/05/2014