# Deliverable 2.3.2

## Interim Real-world Community Engagement Report

Coordinator: Maria Maleshkova  
*With contributions from: Maribel Acosta, Alexander Mikroyannidis, Elena Simperl, John Domingue*

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:</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>21</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Community engagement, dissemination, training</td>
</tr>
</tbody>
</table>
Executive summary

Alongside with producing a comprehensive curriculum and high-quality training materials, EUCLID aims to actively engage with the Linked Data and data community. This is important in order to be able to actively disseminate the results and make them available to a wide audience. Furthermore, this enables the fathering of feedback, which is a crucial step for improving both the structure of the educational curriculum, as well as the content of the materials.

In this deliverable we report on the delivery and feedback from both participating in real-world (face-to-face) events and presenting the current progress of the project. In particular we give details on the community engagement during the first year of the project, which is an important part of exposing our results but is also a necessary step towards the curriculum and training materials development and continuous improvement to take place within the EUCLID project.

We engaged the community through different channels, depending on the individual target groups. First, the Linked Data community has a very strong developer and 'hacker' core, which we addressed trough events such as hackathons and meet-ups, and paly a significant role in spreading knowledge and best practice. Furthermore, we targeted a number training and educational events, putting in practice EUCLID’s materials and testing how well they address the competency goals and the needs of the selected audience. As an educational curriculum EUCLID aims to achieve take-up precisely at events where people are looking for education or education providers. Therefore, conferences and summer schools, as well as further educational opportunities, are an important route to promotion. We also engaged in a number of conferences related to Linked Data, where we presented the curriculum. Finally, we address the community through collaboration with collaboration with educational institutions and companies with expertise the field.
Document Information

<table>
<thead>
<tr>
<th>IST Project Number</th>
<th>FP7 - 296229</th>
<th>Acronym</th>
<th>EUCLID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Title</td>
<td>Educational curriculum for the usage of Linked Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project URL</td>
<td><a href="http://www.euclid-project.eu/">http://www.euclid-project.eu/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document URL</td>
<td><a href="http://www.euclid-project.eu/resources/deliverables/">http://www.euclid-project.eu/resources/deliverables/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU Project Officer</td>
<td>Werner Janusch</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deliverable Number</th>
<th>Title</th>
<th>Interim Real-world Community Engagement Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Package Number</td>
<td>Title</td>
<td>Community Building and Outreach</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date of Delivery</th>
<th>Contractual</th>
<th>M12</th>
<th>Actual</th>
<th>M12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>version 0.1</td>
<td></td>
<td>final - ✓</td>
<td></td>
</tr>
<tr>
<td>Nature</td>
<td>other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissemination level</td>
<td>public</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Abstract (for dissemination) Alongside with producing a comprehensive curriculum and high-quality training materials, EUCLID aims to actively engage with the Linked Data and data community. This is important in order to be able to actively disseminate the results and make them available to a wide audience. Furthermore, this enables the fathering of feedback, which is a crucial step for improving both the structure of the educational curriculum, as well as the content of the materials.

In this deliverable we report on the delivery and feedback from both participating in real-world (face-to-face) events and presenting the current progress of the project. In particular we give details on the community engagement during the first year of the project, which is an important part of exposing our results but is also a necessary step towards the curriculum and training materials development and continuous improvement to take place within the EUCLID project.

Keywords Community engagement, dissemination, training

Version Log

<table>
<thead>
<tr>
<th>Issue Date</th>
<th>Rev. No.</th>
<th>Author</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.03.2012</td>
<td>0.1</td>
<td>Maria Maleshkova (KIT)</td>
<td>Initial deliverable draft</td>
</tr>
<tr>
<td>17.04.2013</td>
<td>0.2</td>
<td>Maria Maleshkova (KIT)</td>
<td>Improved sections</td>
</tr>
<tr>
<td>28.04.2013</td>
<td>1.0</td>
<td>Maria Maleshkova (KIT)</td>
<td>Complete draft</td>
</tr>
</tbody>
</table>
# Table of Contents

EXECUTIVE SUMMARY .................................................................................................................. 2
DOCUMENT INFORMATION ............................................................................................................. 3
TABLE OF CONTENTS ..................................................................................................................... 5
LIST OF FIGURES AND TABLES ..................................................................................................... 6
ABBREVIATIONS ............................................................................................................................ 7

1 INTRODUCTION .......................................................................................................................... 8

2 EVENTS FOR REAL-WORLD COMMUNITY ENGAGEMENT .......................................................... 10
   2.1 EVENTS FOR DEVELOPERS ..................................................................................................... 10
   2.2 EDUCATION AND TRAINING EVENTS .................................................................................. 11
       2.2.1 Conference Tutorials ..................................................................................................... 11
       2.2.2 Summer Schools .......................................................................................................... 13
       2.2.3 Educational Events ...................................................................................................... 14
       2.2.4 Dedicated Training Events .......................................................................................... 14
   2.3 DISSEMINATION EVENTS ..................................................................................................... 15
   2.4 ACTIVE COLLABORATIONS ................................................................................................. 17
   2.5 FUTURE RELEVANT EVENTS .............................................................................................. 18

3 CONCLUSION .............................................................................................................................. 20

REFERENCES .................................................................................................................................. 21
List of Figures and Tables

Figure 1: ESWC Summer School 2013 ................................................................. 13
Figure 2: Session on Educational Curricular for Data Science at the EDF .................. 15
Figure 3: The Data Science Concept Cube, shown by Marko Grobelnik ....................... 16
Figure 4: The audience at the Data Science Curricula session at EDF .......................... 17
**Abbreviations**

DL – Distance Learning
DoW – Description of Work
EU – European Union
ESWC - European/Extended Semantic Web Conference
ISWC - International Semantic Web Conference
IT – Information Technology
KR – Knowledge Representation
LD - Linked Data
LOD – Lined Open Data
SemTech - Semantic Technology (& Business) Conference
SPARQL – SPARQL Protocol and RDF Query Language
SQL – Structured Query Language
WP – Work Package
WWW - World Wide Web (Conference)
XML - Extensible Markup Language


1 Introduction

Alongside with producing a comprehensive curriculum and high-quality training materials, EUCLID aims to actively engage with the Linked Data and data community. This is important in order to be able to actively disseminate the results and make them available to a wide audience. Furthermore, this enables the gathering of feedback, which is a crucial step for improving both the structure of the educational curriculum, as well as the content of the materials.

In this deliverable we report on the delivery and feedback from both participating in real-world (face-to-face) events and presenting the current progress of the project. In particular we give details on the community engagement during the first year of the project, which is an important part of exposing our results but is also a necessary step towards the curriculum and training materials development and continuous improvement to take place within the EUCLID project.

This Interim Real-world Community Engagement Report describes the direct engagement with the Linked Data community, listing the attended and organised events, and also pointing some of the upcoming planned activities. Overall, the active community engagement in the context of EUCLID is important for:

- Making the project and its results known;
- Encouraging the adoption of the curriculum and the reuse of the training materials;
- Collecting feedback on the curriculum and training materials.

As previously mentioned, an important factor in the current success of the Linked Data movement - otherwise defined by four simply-stated principles and the set of standards (W3C/IETF Web/Internet standards and W3C Semantic Web standards) - is the large and enthusiastic community and the best practice they define. The Linked Data community has a strong 'hacker ethic' where the spread of knowledge and best practice is facilitated by events such as:

- **meet-ups** – demand- and theme-driven meetings organised on informal websites such as meetup.com;
- **hackathons**, where often a particular range of datasets, and domain experts, who can give advice on their content and significance, are present and participants develop 'mash-ups' and other types of prototypical applications;
- **vocamps**, where domain experts and enthusiasts come together to define or extend semantic vocabularies to model a given domain.

EUCLID has actively engaged in this type of events, presenting some of our results and participating in discussion. Since Linked Data is, to a large degree, a movement encouraging the development and maintenance of best practice guidelines, it is essential that the EUCLID project engage with these events.

We have also actively engaged the community by participating in educational and training events, applying the developed curriculum and using the completed materials in practice. Some of the organised and attended events include:

- **Tutorials at conferences** are an adequate way for putting in practice some of the training materials, which have already been completed. Sessions at conferences, both those oriented towards researchers and industrial adopters of Linked Data technologies, are an opportunity for encouraging the take-up of the training offering as a whole and also for gathering valuable feedback for improving the curriculum and the used content.
- **Summer schools** – similarly to the tutorials, the summer schools provide an excellent opportunity to trial, and receive rich feedback on the learning materials developed in the EUCLID project. Since summer schools usually cover a number of days, larger sections of the curriculum can be used.
- **Professional training events** – partners in the consortium already conduct regular stand-alone training activities for the benefit of their customers. This way of engaging the community is especially important, since it targets data practitioners, not from an academic environment, but rather from a business one.
As an educational curriculum EUCLID aims to achieve take-up precisely at events where people are looking for education or education providers. Therefore, conferences and summer schools, as well as further educational opportunities, are an important route to promotion. We also engaged in a number of conferences related to Linked Data, where we presented the curriculum. Finally, we address the community through collaboration with educational institutions and companies with expertise the field.
2 Events for Real-World Community Engagement

In this section we provide details on the events that were attended and organised in order to actively engage with the Linked Data community. This is an important part of the project, not only because such interactions are a valuable source of feedback but also because of the nature of the Linked Data community, which is largely driven by the development and propagation of best practices. We cover events targeted towards addressing developers, general educational and training events, conferences and further collaboration activities.

2.1 Events for Developers

As already mentioned, the Linked Data initiative has largely been driven by the developer community and the fundamental best practices have largely been formed, not out of corporate IT, but out of informal hacker networks. Therefore, it is important to engage in interaction with the community in order to discuss difficulties and advances in the state-of-the-art through a series of relatively informal events. EUCLID addresses precisely these events in order to communicate its results and gather input on the developed educational curriculum. In particular we have been engaged in meet-ups and hackathons.

The Semantic Web meet-ups are an established and well-suited channel for addressing challenges, presenting new results and demonstrating the latest technologies to the community. Meet-ups tend to appeal especially to start-ups who view them as a combination of education and networking. EUCLID will engage, in particular, with the Lotico [1]-organised Semantic Web meet-ups in London [2]. For EUCLID this is an opportunity to pick up new relevant topics but also to communicate the achieved results. Therefore we are involved in the London Semantic Meet-ups. In particular there are a few presentations that were made related to EUCLID results:

• London Semantic Web Meet-up, 11 April 2013¹.
  o Presentation on Mapping Relational Databases: MusicBrainz
  o Abstract: Late last year the W3C released two recommendations for publishing Linked Data from relational data sources: the Direct Mapping and the RDB-to-RDF Mapping Language, R2RML. The mapping from the MusicBrainz dataset, which includes many 'advanced relationships' missing from previous mappings, is one of the largest applications of R2RML to date, and presents a near-unique opportunity for large-scale examples and benchmarking. Dr. Barry Norton, who was part of the team originally mapping the MusicBrainz Next Generation Schema with the MusicBrainz Foundation, and maintains the updated mappings and public SPARQL endpoint will present this case study and provide insight into the provision of a large dataset using the developing technologies supporting the standard.

• New York Semantic Web East Coast Meetup, 17 April 2013².
  o Presentation on Applying Ontologies to Linked Data
  o Abstract: A technical presentation on the semantic technologies stack and how ontologies can be applied on Linked Data, in order to improve the quality and interlinking of the data.

As already mentioned in D2.3.2, within the consortium there are connections with Lotico and experience presenting at their native New York meetup [3]. Furthermore, we continue to liaise with existing contacts relating to the EUCLID motivational scenario, for instance at the Centre for Digital Music (C4DM) at Queen Mary University of London (QMUL) [4] and the BBC, who use the music ontology and MusicBrainz dataset in their music reviews [5].

At events such as the SemTech, we also spontaneously attended hackathons. Events such as hackathons are also relevant in the context of our collaboration with Ontotext and provide access to the newest technologies that are

¹ http://www.meetup.com/LondonSWGroup/events/109969672/
² http://www.meetup.com/semweb-25/events/112237022/
being put in practice and are also therefore, relevant for EUCLID’s curriculum. In general hackathons, encourage the attendance of those with limited experience with Linked Data technologies, aiming to solve existing particular problems they are trying having. Another common group of are technology experts, very much at the cutting-edge of Linked Data best practice. It is feasible that EUCLID could organise its own hackathon, during the second year. With the upcoming release of EUCLID’s monitoring platform we are planning to engage even more actively in developer-oriented community engagement.

### 2.2 Education and Training Events

We have also actively engaged in conducting a series of educational and training events. These types of events provide valuable information about how the training materials can be improved and how they can be made more suitable for the target audience. Furthermore, they are a natural way of disseminating the results of the project. Therefore, by giving tutorials at conference and as part of summer schools, EUCLID has the chance to engage a different community of Linked Data adopters. This will increase the amount and the diversity of feedback offered, in turn leading to an even greater improvement over time in the comprehensiveness and timeliness of the curriculum and the quality of the associated teaching materials.

#### 2.2.1 Conference Tutorials

As envisioned by our community engagement plan, we have actively targeted conferences relevant for the Linked Data and Semantic Web (SW) community. These two communities nicely complement each other and there are a number of SW academic conferences that have, which have included Linked Data topics and have attracted a large audience of practitioners. In particular, we made submissions at the World Wide Web conference (WWW), the International Semantic Web Conference (ISWC) and the European Semantic Web Conference (ESWC).

In addition to academic conferences, we also addressed more technology and business-oriented events, such as the Semantic Technology and Business (SemTech) conference series. The SemTech has both European and US editions and provide a platform for addressing Linked Data practitioners from the business environment.

Independent of their orientation, whether academic or practice-focused, conference tutorials offer only a limited amount of time for introducing a single module or parts of individual modules. They cannot be used to cover the complete educational curriculum or conduct extensive training in a particular application area. Therefore, the limited amount of material that can be covered, is used as a way promoting the online materials, the eBooks and the upcoming webinars.

Up-to-date we have made presentations and submissions to the following conferences:

- **Semantic Technology and Business Conference (SemTechBiz) 2012**[^3], 19-20 September in London
  - Training and Introduction to Linked Data[^4]
  - Content: The presentation was based on the EUCLID introduction (Module One) to Semantic Technologies and Linked Data. It covered the Semantic Web technologies stack (RDF, RDFS, OWL, SPARQL) and the Linked Data Principles and best practice (with respect to HTTP, URIs, etc.) The target group was newcomers to these technologies: developers, IT managers, etc., regardless of domain of application. However, in future training will also be targeted at the community of music data workers in order to obtain feedback from domain experts on the motivating scenario.

- **Tutorial on Online Learning and Linked Data at the World Wide Web Conference (WWW) 2013**[^5], 14 May in Rio de Janeiro

Online Learning and Linked Data

Content:

- Part I: New online learning methods for teaching Linked Data
  1. Introduction
  2. How new online learning methods can be used to support the teaching of Linked Data
  4. Plenary discussion on the outcomes of the first group activity.

- Part II: Supporting online learning with Linked Data
  1. How Linked Data technologies can be used to publish online education resources
  2. The Linked Education Cloud
  3. Using the Linked Education Cloud for educational services
  4. The future of linked learning (group discussion)

- RDB2RDF Full-Day Tutorial at ESWC 2013⁵, 26-30 May in Montpellier
  - The new R2RML and Direct Mapping Standards, from Semantics to Practice
  - Content: This tutorial will introduce participants to RDB2RDF. We will start out with the historical context and then present an overview of the new W3C RDB2RDF standards: Direct Mapping and R2RML and how they complement each other. We will continue to present recent scientific results and then present an overview of different RDB2RDF tools, such as D2R, Karma, SparqlMap and Ultrawrap. In the afternoon, we will share several case studies of integrating relational databases in semantic systems. We will present the MusicBrainz’s case study, where RDB2RDF tools have been evaluated in order to ETL relational data to RDF and load it into OWLIM. Additionally, we will present the case study of rCAD, an RNA database that has been mapped to the Gene Ontology and RNA Ontology in order to support semantic search.

- Semantic Technology and Business Conference (SemTechBiz) 2013⁷, 2-4 June in San Francisco
  - Relational Database to RDF (RDB2RDF): The R2RML and Direct Mapping Standards⁸
  - Content: Relational Database to RDF (RDB2RDF) bridges silos of relational data with Linked Data. This tutorial will introduce, contrast and explain the preferred solution organizations embodied by the new W3C RDB2RDF standards: Direct Mapping and R2RML. In particular, the suitability of each approach will be characterized by both architectural concerns of the larger system, and the skill sets that exist in the supporting organizations. Architectural aspects include comparing ‘Extract, Transform, Load’ (ETL) of relational data to RDF into a large cached, or warehoused triplestore versus loosely coupled, distributed linked-data by answering SPARQL queries against a live relational database. Organizational issues concern the scope of in-house ontology development and accompanying Semantic Web components. The material will be organized around case studies spanning reference architectures suggested by a number of vendors and open-source contributors. This tutorial will also include an introductory hands-on session.

The objectives of the tutorial are the following: 1. Present an overview of the W3C Direct Mapping and R2RML standards; 2. Communicate case studies of RDBMS in Semantic Web systems; 3. Deliver practical experience in two scenarios (dynamic querying and ETL) for providing SPARQL

⁵ http://www2013.org/2013/02/26/accepted-tutorials/
⁷ http://semtechbizsf2013.semanticweb.com
⁸ http://semtechbizsf2013.semanticweb.com/programDetails.cfm?ptype=K&optionID=206&pgid=4
capabilities over real-world data. 4. Provide introductory hands-on experience with RDB2RDF tools

We plan to make submissions to the same events during the final year of the project.

2.2.2 Summer Schools

Summer schools offer the possibility for a more focused and intensive training, covering a wider range of topics and allowing for exercises and implementation sessions. While conferences have a wide range of participant backgrounds, including many commercial practitioners, summer schools usually have students with similar background and common interests. Therefore, we aim to use EUCLID material in conducting summer schools, since they provide the perfect setting for conducting EUCLID-based training and the gathered feedback is invaluable.

This year EUCLID will be involved in the 3rd ESWC Summer School, and will present training materials and gather detailed feedback. The 3rd ESWC Summer School will run in southern Crete from the morning of Monday September 2nd to midday Saturday September 7th 2013. The overall goal for this event is to provide intensive training and networking opportunities for the next EWC generation. In particular we wish to facilitate the creation of a new cohort of ESWC: Master's, Ph.D. students and junior researchers who will in time become the next leaders of the ESWC conference and of the Semantic Web research area in general.

Figure 1: ESWC Summer School 2013

As stated by the website of the summer school: Recent press have highlighted the value of Open Educational Resources (OERs) within Massive Online Open Courses. The summer school will make extensive use of OERs from the Euclid and PlanetData EU projects allowing participants to:

- View materials related to the technical part of the Summer School as high quality eBooks (available on eBook readers, on iPads, Kindle) in web form (see 'Learning Materials' menu item).
- View the technical presentations as Webinars (see 'Learning Materials' menu item).
- Provide feedback on all materials to the course authors and Summer School tutors for the benefit of both educators and learners.

We plan to remain involved in the summer school and hopefully the complete EUCLID curriculum can be used during the summer school in 2014.

---

2.2.3 Educational Events

EUCLID materials can be also effectively integrated as part of existing university courses. In addition to conference events, university activities and lectures can be used to promote the EUCLID by frequent reference during the training, and students will be encouraged to read beyond the material delivered in a classroom situation. Furthermore, the students will be encouraged to engage with the webinar content produced by the project, as this can serve to gain further related knowledge.

Currently EUCLID materials are being reused in the following educational courses:

- **Semantic Web Technologies II lecture during the Summer Semester 2013 at KIT**
  - Modules 1, 2 and 4 will be reused as part of the lectures, including their corresponding exercises.

- **Service Oriented Computing II lecture during the Summer Semester 2013 at KIT**
  - Modules 1 and 5 will be reused as part of the lectures, including their corresponding exercises.

- **Integration in teaching at the Open University**

We will continue to look for possibilities to integrate EUCLID materials in actual educational courses.

2.2.4 Dedicated Training Events

Through the involvement of Ontotext in the project consortium, we already have a very strong partner with experience in professional training. In particular, Ontotext organises a bi-monthly training event in London [6], covering Linked Data technologies and several specific technologies covered in the EUCLID curriculum. Furthermore, Ontotext is partially integrating EUCLID teaching materials within this course. They will also offer to gauge customer interest in the inclusion of advanced topics from the EUCLID curriculum both in this open training event, and within the in-house training for which they are contracted by a number of major customers, particularly in the media and publishing industries.

Some of the events that have already taken place and are based on integrating EUCLID materials include:

- **Internal and external training on Semantic Technologies with OWLIM - UK:**

- **Free training on Semantic Technologies Introduction collocated with the London SemWeb Meetup 2013:**
  - Semantic Technologies Introduction is a free introductory course aimed at people who use or intend to use semantic technologies for data management, publishing, content analysis. It is targeted at senior managers, IT professionals and individuals who are interested in gaining a thorough understanding of the basic concepts and standards around the Semantic Web. It is also a foundation for the Semantic Technologies with OWLIM course offered the following day on April 12th.


- **Semantic Technologies with OWLIM - SemWeb Training 2013**
  - Semantic Technologies with OWLIM is aimed at people who use or intend to use semantic technologies for data management, publishing, content analysis, data integration and warehousing. Our condensed, one-day course is intended to accommodate participants with a wide range of interests and backgrounds, but assumes some knowledge of semantic technologies, such as the material covered in the preceding free training event at London SemWeb Meetup 2013.
2.3 Dissemination Events

We had a very successful session on a data science curriculum for professionals at the European Data Forum 2013 in Dublin\textsuperscript{10}. The session was a joint venture between the EUCLID and the PlanetData projects.

The side track was chaired by John Domingue and included the following themes:

- John Domingue (The Open University) spoke about the need also for learning materials for the wider community to ensure citizen engagement around the main issues, including privacy, associated with Big Data. He also discussed the need for a constructivist learning approach based upon sound pedagogical theories and supported by easy-to-use environments. Slides available at: http://www.slideshare.net/EUCLIDproject/data-science-curriculum-v-32

- Nick Campbell (Trinity College Dublin) spoke about his wide experiences in speech understanding and its relationship to Big Data especially from the perspective of privacy. Slides available at: http://www.slideshare.net/EUCLIDproject/speech-technology-and-big-data

\textsuperscript{10} http://2013.data-forum.eu/
• Barry Norton (Ontotext) gave an overview of the main principles underlying the teaching approach adopted in Euclid:
  
  o **Show realistic solutions** – Euclid’s learning materials are based around a data set containing more than 200M triples at the moment and Euclid aims to grow this to around 0.5B triples.
  
  o **Use real data** – Euclid has adopted the MusicBrainz dataset which is part of the Linked Open Data Cloud and is used in a number of real applications.
  
  o **Use real tools** – which are explained in screencasts, webinars as well as being available for students to interact with. We also include the latest W3C standards such as R2RML in our portfolio.
  
  o **Show scalable solutions** – all the technologies we show work at the 100M+ triples scale. Moreover we make these available to students within our iBook formats integrated with the multi-media learning material.
  
  o **Eat the dog’s food** – Euclid monitors community reaction to our material on public email lists, on Twitter, in LinkedIn, SlideShare and Vimeo, aggregates the results as Linked Data and provides the results in a public SPARQL endpoint. Slides available at: [http://www.slideshare.net/EUCLIDproject/big-linked-data](http://www.slideshare.net/EUCLIDproject/big-linked-data)

Marko Grobelnik gave an overview of the main concepts associated with data science based on the cube shown below. **Data modalities** represent different data formats, which vary from rich representations based on ontologies to raw data associated with signals. **Data operators** capture different approaches, which exist to how one can manage and manipulate the data to support understanding and decision-making. **Additional issues** incorporate other essential criteria required to ensure that the mechanisms work in practice. Marko stressed in his presentation the issue that at present there is little collaboration and communication between the research disciplines associated with the cube. Part of this is due to the different viewpoints and languages used in each distinct research area. He gave as an example how he only managed to get statisticians to listen to him when he mentioned that semantics can de-sparse data. Slides available at: [http://www.slideshare.net/EUCLIDproject/data-science-curriculum-approaches-towards-preparing-a-curriculum](http://www.slideshare.net/EUCLIDproject/data-science-curriculum-approaches-towards-preparing-a-curriculum)

---

**Figure 3:** The Data Science Concept Cube, shown by Marko Grobelnik.

The session was well attended and after the presentations a lively debate followed which, on audience insistence, took 20 minutes of the 30 minute coffee break. Our Project Officer Stefano Bertolo was in the audience and he tweeted positively about Barry’s (“…[@BarryNorton](https://twitter.com/BarryNorton) makes me weep of joy when he explains what it takes to teach scalanility (sic) …”) and Marko’s (“Semantics de-sparify data says [@marko_grobelnik](https://twitter.com/marko_grobelnik) and that's why it should be in machine learning curricula…””) presentations resulting in several immediate requests for the slides.
Overall, this session at the EDF represented a very successful way of sharing our experience, gathering feedback and discussing current issues related to developing educational and training curricula. Therefore, we will continue to pursue involvement in such events in the future.

2.4 Active Collaborations

In addition to providing trainings and actively sharing EUCLID’s materials, we also engage in collaborations with companies and organisations with interests or expertise in the Linked Data area.

- fluidOps is a company active in the field of employing semantics for data management. We use some of the developed tools in order to provide examples within the modules, while we receive feedback on generated training materials. In particular we use the Information Workbench. As stated by the product description:
  - Pursuing the goal to lower the entry barrier into the world of Linked Data and to leverage its benefits, the Information Workbench™ provides a Web-based open platform for Linked Data and Big Data solutions in the enterprise. With its flexible and extensible interfaces, data from different sources can be integrated and seamlessly connected. A Linked Data layer on top of the content facilitates semantic access and semantic search in the integrated data across the borders of domains. The Information Workbench also provides comprehensive and timely access to large-scale data sets.

11 http://www.fluidops.com/
12 http://www.fluidops.com/information-workbench/
The Open Data Institute (ODI)\textsuperscript{13} is an organisation that aims to support the wider adoption and availability of open data both in the context of governmental setting but also for business and commercial use. Currently, we are following their training activities and aligning EUCLID’s curriculum the ODI’s courses.

Data Science London\textsuperscript{14} – Building a Community of Data Scientists, Passionate about Data Science. As the motto states, this a community around data science. Currently the collaboration actions are focusing on aligning the training curricula and exchanging expertise.

European Data Forum (EDF)\textsuperscript{15} “is a meeting place for industry, research, policymakers and community initiatives to discuss the challenges of Big Data and the emerging Data Economy and to develop suitable action plans for addressing these challenges”. The goal is to build a community of parties interested in managing and using Data, with focus on SMEs. We aim to be active in the context of EDF events, where we have already presented our training curriculum.

Optique\textsuperscript{16} is a European project on supporting scalable end-user access to Big Data. In particular the goal is to provide a semantic end-to-end connection between users and data sources, and to enable users to rapidly formulate intuitive queries using familiar vocabularies and conceptualisations. We will share our training materials with the project, so that they can be reused as part of the teaching and dissemination events.

INTUITEL\textsuperscript{17} is a research project co-financed by the European Commission with the aim to advance state-of-the-art e-learning systems via addition of guidance and feedback for learners. Currently, the project is in the process of developing their learning system prototype. One the system is ready, EUCLID’s training materials will be one of the first materials to be offered.

Both the University of Southampton\textsuperscript{18} and the University Simon Bolivar\textsuperscript{19} provide us with initial materials while preparing the module but more importantly they play an important role in the quality assurance process, giving us feedback on the complete eBook chapters.

Ontotext’s training program\textsuperscript{20} with all their training events, where we regularly reuse and integrate EUCLID training materials.

\section*{2.5 Future Relevant Events}

In this section we provide a list of planned activities until the end of the project, building up on the success achieved until now:

- Continuous training/awareness slot at a the Semantic Web London meet-ups;
- Involvement in EDF-related training activities;
- A hackathon, organized by, or including, EUCLID;
- Submissions of conference tutorials at the WWW, ESWC, ISWC;
- Submission of tutorials at SemTech, building on the achieved results;
- Reuse of EUCLID training materials as part of summer schools, continuous involvement in the ESWC Summer School;
- Inclusion of EUCLID training material in educational courses;

\textsuperscript{13} http://www.theodi.org
\textsuperscript{14} http://datasciencelondon.org/
\textsuperscript{15} http://www.data-forum.eu/
\textsuperscript{16} http://www.optique-project.eu/
\textsuperscript{17} http://www.intuitel.de/
\textsuperscript{18} www.southampton.ac.uk
\textsuperscript{19} http://www.usb.ve/
\textsuperscript{20} https://confluence.ontotext.com/display/training/Home
• Inclusion of EUCLID training material in Ontotext professional training early 2014.
3 Conclusion

In this deliverable we describe the current actions that we have taken in order to actively engage with the Linked Data community in particular and the Data community in general. On one side we provide trainings by reusing EUCLID’s materials and, therefore, advertise the project and disseminate its results. On the other side we gather valuable feedback on how we can improve the curriculum, extend the modules with further topics or make the trainings more suitable for a particular audience.

We have distinguished several types of event that are relevant to this approach, and have engaged the community through different channels, depending on the individual target groups. First, the Linked Data community has a very strong developer and ‘hacker’ core, which we addressed through events such as hackathons and meet-ups, and play a significant role in spreading knowledge and best practice. Furthermore, we targeted a number training and educational events, putting in practice EUCLID’s materials and testing how well they address the competency goals and the needs of the selected audience. As an educational curriculum EUCLID aims to achieve take-up precisely at events where people are looking for education or education providers. Therefore, conferences and summer schools, as well as further educational opportunities, are an important route to promotion. We also engaged in a number of conferences related to Linked Data, where we presented the curriculum. Finally, we address the community through collaboration with collaboration with educational institutions and companies with expertise the field.
References