EUCLID: EdUcational curriculum for the usage of LInked data

Conference or Workshop Item

EUCLID: EdUcational Curriculum for the usage of LInked Data

Elena Simperl, Maribel Acosta, Marin Dimitrov,
John Domingue, Peter Haase, Maria Maleshkova,
Alexander Mikroyannidis, Barry Norton, and Maria Esther Vidal

http://euclid-project.eu/

1 EUCLID in a nutshell

Linked Data has established itself as an emerging standard for the publication of structured data over the Web, enjoying amazing growth in terms of the number of organizations committing to use its best practices and technologies for exposing and interlinking data sets for seamless exchange, integration, and reuse. More and more ICT ventures offer innovative data management services on top of Linked (Open) Data, creating a demand for data practitioners with a background in semantic technologies. Ensuring the availability of such expertise will prove crucial if European businesses are to reap the full benefits of the advanced data management technology, and the know-how accumulated over the past years by researchers, technology enthusiasts and early adopters in various European countries. EUCLID had a major contribution to this goal by providing a comprehensive educational curriculum, supported by multi-modal training materials and state-of-the-art eLearning distribution channels, tailored to the real needs of data practitioners. Building upon experience reports from over twenty Linked Data projects with over forty companies and public offices in more than ten countries, complemented by feedback from hundreds of training events, and an in-depth analysis of the community discourse through mailing lists, discussion forums, Twitter, and the blogosphere, the curriculum focuses on techniques and software to integrate, query, and visualize Linked Data, as core areas in which practitioners state to require most assistance. It is realized as a combination of multi-modal learning resources, including an iBook published on iTunes U, and evaluated through webinars, f2f training, and continuous community feedback. By providing these key knowledge-transfer components, EUCLID will not only promote the industrial uptake of Linked Data best practices and technologies, but, most importantly, will contribute to their further development and consolidation, and support the sustainability of the community, all essential aspects given the rapid pace at which the field has recently advanced.

2 Main outcomes of the project

Learning resources The EUCLID educational curriculum consists of a series of modules, each containing multi-format learning materials, such as presentations, webinars, screencasts, exercises, eBook chapters, and online courses.

These learning materials complement each other and are connected to deliver a comprehensive and concise training programme to the community. Learners are guided through these materials by following learning pathways, which are sequences of learning resources structured appropriately for achieving specific learning goals. A significant part of the learning materials consists of examples referring to real-world data sets and application scenarios, code snippets and demos that developers can run on their machines, as well as best practices and how-tos. Different types of eLearning distribution channels are targeted by each type of learning materials, including Apple and Android tablets, Amazon Kindle, as well as standard Web browsers. Instead of toy Linked Data examples, we use a collection of data sets and tools that are deployed and used for real, most importantly, MusicBrainz and DBpedia. Our collection of tools includes Seevl, Sesame, Open Refine and GateCloud, all of which are used in real-life contexts. We also showcase scalable solutions, based upon industrial-strength repositories and automatic translations, e.g., by using the W3C standard R2RML for generating RDF from large data contained in standard databases.

Best practices for curriculum design and delivery EUCLID has distilled the lessons learned during the production of the learning materials, and their evaluation and improvement through webinars and community feedback into a list of ten principles covering aspects such as industrial relevance, collaboration (both internal and external), explicit learning goals, use of real-world data sets and tools, open formats, and self-testing. The principles target future designers of training programs on similar topics such as Open Data, or Data Science.

Community monitoring platform We monitor communication and engagement with the Linked Data community through W3C email lists, in the social network channels LinkedIn and Twitter as well as content dissemination channels such as Vimeo and SlideShare. We transform the monitoring results into RDF and make these available at a SPARQL endpoint. This can be thought of as a first step in using Linked Data to support learning analytics.

In the networking session we will demonstrate how each of the materials can be used, in particular the integrated courses and the iBook (on an iPad). We will also discuss the best practices for curriculum delivery and design, as well as lessons learned in community engagement. Finally, we will show a prototype of the topic monitoring platform which is based on Information Workbench and Ontotext storage technology.

3 Further information

The project consortium is coordinated by STI Research (AT) and consists of the Karlsruhe Institute of Technology (DE), Ontotext (BG) and The Open University (UK). The project has a number of associate partners, including Fluid Operations AG (fDE), University Simn Bolvar (VZ) and the University of Southampton (UK). It is supported by $FP7$ under the grant agreement $FP7 - 296229$. The project has started in May 2012 and has a duration of two years.