



LIDER – Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe

Background

The aim of LIDER is to address the challenge of the explosive growth of content in volume, velocity and variety on the Web, which demands new approaches to content analytics. A key element is the understanding and interpretation of language in content, both as unstructured textual content and linguistic content present in the context of different media streams. LIDER will build a pathway for the establishment of a new Linked Open Data (LOD) based ecosystem of free, interlinked, and semantically interoperable resources from the realm of both language (“Linguistic Linked Data” representations of corpora, dictionaries, lexical and syntactic metadata, etc.) and media (image, video, etc. metadata). LIDER will study how these resources can serve as an enabler technology for enterprise content analytics on the Multilingual Web, including multilingual content delivered in multiple media.

LIDER: Main achievements in Year 1

➤ **Reaching out to communities, building business use cases.** LIDER has held four roadmapping workshops to gather input for business use cases and requirements for linguistic linked data. The workshops targeted four groups: the data community, the Web community at large, the localisation community and the analytics community. The feedback was gathered via a dedicated W3C community group LD4LT (“Linguistic Linked Data for Language Technology”), has been documented in workshop reports and has fed into the LIDER roadmap. In addition, LIDER has been present at a large number of research and industry events, held

tutorials and gave presentations about linguistic linked data and content analytics.

➤ **Guidelines and best practices.** LIDER has used a dedicated W3C community group BPMLOD (“Best Practices for Multilingual Linked Open Data”) to gather best practices for creating multilingual linked data sources. Conversions for existing language resources into linked data have been developed. This concerns first metadata about language resources. LIDER worked closely with the META-NET and the CLARIN community on the conversions, to assure broad acceptance and distribution of existing metadata schemes. Second, LIDER described best practices for various types of resources, like bilingual or multilingual dictionaries, and started conversions for other resources like terminology (TBX) data or corpus data.

➤ **Reference architecture and roadmapping.** LIDER has created a first draft of a linguistic linked data reference architecture for content analytics. The architecture defines a general model for building linguistic linked data aware services and several patterns for building content analytics applications. It also provides an overview tools and initiatives relevant for the architecture. LIDER also has created a first version of a roadmap for linguistic linked data for content analytics. The roadmap interprets input from various sources: LIDER activities like surveys or roadmapping workshops, various reports, funding agencies and various research communities. The roadmap is formulated around three key application areas:



LIDER – Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe

global customer engagement, public sector / civil society, and linguistic linked data life cycle and linguistic linked data value chain. For each of these areas, the roadmap provides several use cases including time lines, predictions and relevant actors.

Next steps

LIDER will validate its year one outcomes by discussing them with the community at large. This concerns conversion of linguistic resources and its metadata, the linguistic linked data reference architecture for content analytics, and the roadmap for linguistic linked data for content analytics. In addition, LIDER will reach out to new groups that have not been targeted in year one, especially in the area of public sector information.

Coordinator

Contact Person: Prof. Dr. Asunción Gómez-Pérez

- Tel:+34 913367417
- E-mail: asun@fi.upm.es
- Organisation: Universidad Politécnica de Madrid (UPM)

Participants

- Universidad Politécnica de Madrid (UPM)
- Trinity College Dublin (TCD)
- German Research Center for Artificial Intelligence (DFKI)
- INSIGHT National Center for Data Analytics, National University of Ireland, Galway
- Institute for Applied Informatics (InfAI), University of Leipzig
- University of Bielefeld
- Sapienza University of Rome
- W3C / Geie ERCIM

Project setup

- Duration: 24 months
- Start date: 1 November 2013
- End Date: 31 October 2015
- Project Officer: Susan Fraser
- Project 610782, CSA

How to contribute and how to be informed about LIDER

- Project website <http://www.lider-project.eu/>
- Community portal <http://www.multilingualweb.eu/>
- Upcoming W3C Community Group on “Linked Data for Language Technology”
- Mailing list lider-community@listas.fi.upm.es for gathering anonymous feedback and for getting information. Use <http://www.lider-project.eu/get-involved> to subscribe
- Social media presence via hashtag #LiderEU e.g. on twitter <https://twitter.com/multilingweb>