LIDER: FP7 – 610782

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

Deliverable number	**D4.3.1**
Deliverable title	**Preliminary report on LIDER Community and community portal – Phase I**
Main Authors	**David Lewis, Felix Sasaki**

Grant Agreement number	610782
Project ref. no	FP7-610782
Project acronym	LIDER
Project full name	Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe
Starting date (dur.)	1/11/2013 (24 months)
Ending date	31/10/2015
Project website	http://www.lider-project.eu/

Coordinator	Asunción Gómez-Pérez
Address	Campus de Montegancedo sn. 28660 Boadilla del Monte, Madrid, Spain
Reply to	asun@fi.upm.es
Phone	+34-91-3367417
Fax	+34-91-3524819

Document Identifier	D4.3.1
Class Deliverable	LIDER EU-ICT-2013-610782
Version	2.1
Document due date	31/1/2014
Submitted	03/2/2014
Responsible	David Lewis, TCD
Reply to	Dave.lewis@scss.tcd.ie
Document status	Final
Nature	R(Report)
Dissemination level	PU(Public)
WP/Task responsible(s)	ERCIM, TCD
Contributors	David Lewis (TCD), Dominic Jones (TCD), Kevin Koidl (TCD), Felix Sasaki (DFKI/ERCIM)
Distribution List	Consortium Partners
Reviewers	Martin Brümmer (INFAI)
Document Location	http://lider-project.eu/?q=doc/deliverables

Executive Summary

This document provides an overview of the set up of the LIDER community portal, which has been implemented as a W3C Community Group called Linked Data for Language Technology.

Document Information

IST Project Number	FP7-610782	Acronym		LIDER	
Full Title	Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe				
Project URL	http://www.lider-project.eu/				
Document URL	http://lider-project.eu/?q=doc/deliverables				
EU Project Officer	Susan Fraser				

Deliverable	Number	D4.3.1	Title	Preliminary report on LIDER Community and community portal –Phase I
Workpackage	Number	WP4	Title	Community building and dissemination

Date of Delivery	Contractual	31/01/2014	Actual	03/02/2014
Status	version 2.1		final ■	
Nature	prototype □ report ■ dissemination □			
Dissemination level	public ■ consortium □			

Authors (Partner)	David Lewis (TCD), Dominic Jones (TCD), Kevin Koidl (TCD), Felix Sasaki (DFKI/ERCIM)			
Responsible Author	Name	David Lewis	E-mail	Dave.lewis@scss.tcd.ie
	Partner	TCD	Phone	+35318968428

Abstract (for dissemination)	This document provides an overview of the set up of the LIDER community portal, which has been implemented as a W3C Community Group called Linked Data for Language Technology.
Keywords	LIDER Community portal

Version	Modification(s)	Date	Author(s)
01	First draft	15 jan 14	D.Lewis TCD
02	Revised template used, review comments integrated	31 jan 14	D.Lewis TCD

Project Consortium Information

Participants		Contact
Universidad Politécnica de Madrid		Asunción Gómez-Pérez Email: asun@fi.upm.es
The Provost, Fellows, Foundation Scholars & The Other Members of Board of The College of the Holy & Undivided Trinity of Queen Elizabeth near Dublin (Trinity College Dubl, Ireland)		David Lewis Email: dave.lewis@cs.tcd.ie
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI, Germany)		Felix Sasaki Email: fsasaki@w3.org
National University of Ireland, Galway (NUI Galway, Ireland)		Paul Buitelaar Email: paul.buitelaar@deri.org
Institut für Angewandte Informatik EV (INFAI, Germany)		Sebastian Hellmann Email: hellmann@informatik.uni-leipzig.de
Universität Bielefeld (UNIBI, Germany)		Philipp Cimiano Email: cimiano@cit-ec.uni-bielefeld.de
Universita degli Studi di Roma La Sapienza (UNIVERSITA DEGLI STU, Italy)		Roberto Navigli Email: navigli@di.uniroma1.it
GEIE ERCIM (ERCIM, France)		Felix Sasaki Email: fsasaki@w3.org

Table of Contents

1 Objectives

The Objective of the LIDER Community Portal is to provide an online infrastructure to support the external community that LIDER assembles to engage on requirements and use cases for developing guidelines, best practice and the research roadmap for Linguistic Linked Data.

The LIDER community is also referred to as the Industrial Board, to indicate the intention that it should be largely made up of industrial representatives, and in particular SMEs.

2 Context: LIDER Public Facing Infrastructure

The LIDER Community Portal sits within a broader infrastructure for public engagement which additionally consists of:

- The Multilingual Web portal operated by the W3C for making announcements and publishing material from the Multilingual Web conference series. This consists of the www.multilingualweb.eu web site together with the associated email list and social media feeds.
- The LIDER project portal hosted by UPM at www.lider-project.eu. This integrates the project's public announcement mailing list and social media feeds together with public GitHub and data repository.

3 W3C Community Group Infrastructure

The LIDER Community Portal is implemented technically and organizationally as a W3C Community Group. These groups are supported by the W3C to enable community discussion about technical issues that could potentially contribute to standardization on the Web[1]. Members require a freely available W3C member account. Community members are free to structure their groups as they see fit amongst themselves, selecting group chairs, discussion/meeting frequency, setting objectives and work timetables and are not subject to control or supervision by the W3C.

The technical infrastructure available for W3C Community Groups consists of:

- Wiki
- Event calendar
- Online chat
- Blog
- RSS feed
- Mailing list
- Programmable survey/questionnaire infrastructure
- Online group and membership administrative interface

---

[1] http://www.w3.org/community/about/#cg

4 Linked Data for Language Technology Community Group

LIDER has therefore established a Community Group called Linked Data for Language Technology, LD4LT for short. The name was chosen to provide widespread recognition of the goals of the community, as opposed to being an adjunct to the LIDER project. The community is focused on the capture of use cases and requirements for Linguistic Linked Data in support of content analytics applications.

For the development of best practice the existing 'Best Practices for Multilingual Linked Open Data', BPMLOD for short[2], will be used as the focus. This W3C community group was established based on community consensus established at the Multilingual Web conference in Rome 12-13 March 2013[3]. This group was already co-chaired by LIDER partners but benefits from a pre-established mandate and recognition, and had therefore good potential to provide a public conduit for the best practice work in LIDER that can continue its work beyond the lifespan of LIDER. The LD4LT community on the other hand was initially intended to support the development of and engagement with the LIDER industrial board with the aim of capturing requirements and use cases up to month 18 (April 2015). It was anticipated that such a focused, fixed term group with a concrete outcome date would prove more attractive for industry partners compared to a broader, possibly more open-ended activity. Requirements and use cases captures via LD4LT would then form a source of input into BPMLOD as well as other LIDER activities in WP2 (guidelines and best practices) and WP3 (Reference Architecture and Roadmap for content analytics).

There is also strong potential for requirements and use cases gathered by the LD4LT community to provide input to Ontology-Lexica Community Group, OntoLex for short[4]. OntoLex has been established since July 2011. It is also co-chaired by LIDER partners and has a well established work-programme that is progressing to a lexical model that uses semantic web representations.

The LD4LT community is described on the community portal as follows:

This group aims to consult with current and potential users of linguistic data to assemble user cases and requirements for Language Technology Applications that use Linked Data. The results will be used to guide future interoperability, research and development activities spanning the language technology and linked data domains, including via the OntoLex and BPMLOD community groups.

Potential users are companies and public bodies involved in content management, the language services and localisation industry and other applications of content analytics techniques used in search, recommender systems, sentiment analysis and terminology management.

The group will engage with users through surveys and international road-mapping events organised by an EU-funded R&D consortium called LIDER. Contributions can be provided openly through engagement with the Community Group or anonymously if preferred. The anonymised outcomes of this exercise,

---

[2] http://www.w3.org/community/bpmlod/
[3] http://www.multilingualweb.eu/documents/rome-workshop/
[4] http://www.w3.org/community/ontolex/

in terms of use case and requirements priorities, technology gaps and interoperability roadblocks will be made publicly available via this community group by May 2014."

5 Community Engagement Plan via Portal

The LIDER community portal will therefore form a hub for the engagement with the LIDER Industry Board, and in particular will provide the medium via which requirements and use cases will be presented, discussed and refined. The operation of the portal will therefore consist of close collaboration between LIDER WP1 which is implementing the capturing of business use cases and LIDER WP4 which is concerned with community engagement. For the latter, the LD4LT community will provide a focus for community engagement in the development of the road-mapping workshops.

The plan for engagement via the LD4LT portal is documented in LIDER deliverable D4.1: "Community building and dissemination plan"

Appendix I: Activities planned for 2014 (current state)

The following list provides activities that are already confirmed to take place in the first year of the project. For the roadmapping workshops, the project will use the shared budget to invite speakers.

- December 2013: press release, based on the fact sheet (see deliverable D4.4.1). The press release about the start of the LIDER project has been created and translated in several languages. It is being released by the LIDER consortium partners, see e.g. http://www.dfki.de/web/presse/pressemitteilungen_intern/2013/lider-connects-language-and-multimedia-resources
- February 2014: XML Prague
  - A pre-conference session at XML Prague about content analytics, XML and Semantic Web http://www.xmlprague.cz/preconf2014/#sem
  - A demo of content analytics functionality in the oXygen user session
- March 2014: European Data Forum
  - Several LIDER partners have submitted papers to the EDF event.
- March 2014: The first LIDER roadmapping workshop
  - Will take place aligned with EDF. Target community: industry, research, policy makers, and community initiatives around open / linked / big data in Europe.
- May 2014: MultilingualWeb workshop
  - Several LIDER participants are taking part in the workshop program committee to assure that LIDER topics will be taken into account during the workshop.
- May 2014: The second LIDER roadmapping workshop
  - Will be organized as part of the MultilingualWeb workshop. Target community: Web industry at large, various language related industries (language technology, localization), content (e.g. CMS) related industries.
- May 2014: LREC conference
  - A workshop on Linked Data in Linguistics at the LREC conference has been organized by several LIDER partners, see http://ldl2014.org/index.html
  - A tutorial "Linked Data for Language Technology" http://lrec2014.lrec-conf.org/media/filer_public/2013/12/23/t10-tutorialoutline.doc has been organized by UPM.
  - A workshop on "Linked Data in Linguistics: Multilingual Knowledge Resources and Natural Language Processing" has been co-organized by UNIBI and UPM.
  - The LIDER consortium has assured a LIDER presence in the LREC EC village http://lrec2014.lrec-conf.org/en/hlt-village/
- June 2014: The third LIDER roadmapping workshop
  - Will be organized as part of LocalizationWorld event in June 2014, Dublin. Target community: localization industry, machine translation.
- September 2014: The fourth LIDER roadmapping workshop:
  - Will be organized as part of the e-semantics conference in September 2014, Leipzig. Target community: German industry.
- November 2014: Unicode conference
  - A partner of the LIDER consortium will submit a tutorial on linked data and content analytics to the Unicode conference.

Appendix II: Mailing Lists and Communities

Community	Contact partner
ELRA	DFKI
META-NET / META-SHARE	DFKI
ETSI	DFKI
W3C MLW list	ERCIM
W3C ITS IG, I18N groups	ERCIM
DBpedia community	INFAI
Wikidata community	INFAI
NLP2RDF	INFAI
LOD2	INFAI
Open Data in Linguistics WG	INFAI
OKFN	INFAI
Clarin	NUI Galway
FlaReNet	NUI Galway
CNGL, GALA,Localization World, LT-Innovate, TAUS	TCD
OASIS XLIFF TC	TCD
W3C OntoLex community group	UNIBI
WordNet community	UNIBI
BabelNet community	UNIRM
Corpora list	UNIRM
Elsnet	UNIRM
Siglex	UNIRM
AI*IA (Italian Association for Artificial Intelligence)	UNIRM
LN (Language Naturel) French NLP list	UNIRM
W3C BPMLOD community group	UPM
Europeana	UPM
MAVIR	UPM
Linked Data Spain	UPM
AEPIA	UPM
Real Academia de la lengua	UPM
AENOR (Spanish Association for Normalization)	UPM
SEPLN (Spanish Association for NLP)	UPM
AETER (Spanish Association of Terminology)	UPM
Real Academia de Ingeniería	UPM
Fundeu (Fundación Español Urgente)	UPM
DARIAH/CLARIN	DFKI

GSCL (German Society for Computational linguistics	DFKI
INTEROP VLab (European Virtual Laboratory on Interoperability)	UNIRM