

mlider

LIDER: FP7 – 610782

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

Deliverable number	D4.4.3
Deliverable title	Updated Project Fact Sheet, Phase II
Main Authors	Nieves Sande, 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	+34-91-336-7417
Fax	+34-91-3524819

Document Identifier	D4.4.3
Class Deliverable	LIDER EU-ICT-2013-610782
Version	1.0
Document due date	31 October 2015
Submitted	22 October
Responsible	W3C/ERCIM
Reply to	fsasaki@w3.org
Document status	Final
Nature	O(Other)
Dissemination level	PU(Public)
WP/Task responsible(s)	Felix Sasaki, DFKI / W3C Fellow
Contributors	-
Distribution List	Consortium Partners
Reviewers	Reviewed by the project consortium
Document Location	http://lIDER-project.eu/?q=doc/deliverables

Executive Summary

This document, the updated Project Fact Sheet – Phase II, summarizes basic information about the LIDER project and highlights the outcome of the second 12 months of the project.

The main pages of this deliverable provide the textual content of the Project Fact Sheet as a text document. The final pages provide the Project Fact Sheet visualized as a flyer for printing.

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	D4.4.3	Title	Update Project Fact Sheet, Phase II
Workpackage	Number	4	Title	Community building and dissemination

Date of Delivery	Contractual	31 October 2015	Actual	22 October
Status	version 1.0		Final ■	
Nature	prototype <input type="checkbox"/> report <input type="checkbox"/> dissemination ■			
Dissemination level	public ■ consortium <input type="checkbox"/>			

Authors (Partner)	Felix Sasaki, DFKI / W3C Fellow			
Responsible Author	Name	Felix Sasaki	E-mail	fsasaki@w3.org
	Partner	DFKI / W3C Fellow	Phone	+49-30-23895-1807

Abstract (for dissemination)	This document, the updated Project Fact Sheet – Phase II, summarizes basic information about the LIDER project and highlights the outcome of the second 12 months of the project. The main pages of this deliverable provide the textual content of the Project Fact Sheet as a text document. The final pages provide the Project Fact Sheet visualized as a flyer for printing.
Keywords	LIDER, project fact sheet

Version	Modification(s)	Date	Author(s)
01	Initial version	20/10/15	Nieves Sande, DFKI; Felix Sasaki, DFKI / W3C Fellow

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: felix.sasaki@dfki.de
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	BACKGROUND AND THE IMPORTANCE OF LLOD FOR BIG DATA.....	7
2	REFERENCE ARCHITECTURE AND ROADMAPPING	7
3	REACHING OUT TO COMMUNITIES, BUILDING BUSINESS USE CASES	7
4	GUIDELINES AND BEST PRACTICES FOR MULTILINGUAL LINKED OPEN DATA.....	8
5	LINGHUB AND LLOD CLOUD	8
6	GUIDELINES AND REFERENCE CARDS	8
7	IMPACT	9
8	STEPS AFTER LIDER.....	9
9	ADDITIONAL INFORMATION.....	9

1 Background and the importance of LLOD for Big Data

The explosive growth in the volume, velocity and variety of content on the Web demands new approaches to content analytics.

Language technology is the key to efficient understanding and analysis of multilingual content present as unstructured text and the linguistic content in diverse media streams. The effectiveness of language technology for content analytics is however critically dependent on the availability of relevant language resources, i.e. in the right languages, addressing the domains of interest and at sufficient volumes. Linked Open Data (LOD) based on the standards from the World Wide Web Consortium (W3C) offer a robust and sustainable solution to the problems of publishing, discovering, exchanging and managing language resources.

Language technology is the key to efficient understanding and analysis of multilingual content present as unstructured text and the linguistic content in diverse media streams.

The LIDER project is completing a two-year programme of industry consultation, community building and development of technical guidance that defines a clear path for the widespread adoption of LOD to support multilingual content analytics. The result is an ecosystem of free, interlinked, and semantically interoperable resources from the realm of both language (“Linguistic Linked Data” representing corpora, dictionaries, lexical and syntactic metadata and conceptual models) and media (image, speech, video and its metadata).

2 Reference architecture and roadmapping

LIDER has published 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 of existing tools and initiatives that can be used to implement the architecture.

LIDER analysed input from surveys, roadmapping workshops, various market reports, funding agencies priorities and input from various research communities. This resulted in a detailed roadmap for the development and adoption of linguistic linked data.

The roadmap is formulated around three key application areas: global customer engagement, public sector / civil society, and linguistic linked data lifecycle and data value chain. For each of these areas, the roadmap provides several use cases including time lines, predictions and relevant actors, and these have been fed into broader Strategic Research and Innovation Agenda setting activities for H2020.

3 Reaching out to communities, building business use cases

LIDER has held seven roadmapping workshops assembling representatives from industry, the public sector, the voluntary sector and research to develop use cases and requirements for linguistic linked data. The workshops targeted different communities of practitioners including those working in: data management, multilingual web content,

localization and content analysis. The feedback was gathered via the W3C's "Linguistic Linked Data for Language Technology" (LD4LT) community group.

The results have been documented in workshop reports and use case and requirements analyses. They have led to the development of a research and innovation roadmap for linguistic linked data. LIDER has promoted these results on linguistic linked data and content analytics at a wide range of research and industry events.

4 Guidelines and best practices for Multilingual Linked Open Data

LIDER has led the W3C "Best Practices for Multilingual Linked Open Data" (BPMLOD) community group in developing best practices for creating multilingual linked data sources. Conversion guidelines between existing language resources, e.g. multilingual dictionaries, and linguistic linked data have been developed and documented in best practice reports.

LIDER has worked closely with the META-NET and the CLARIN communities on specifications for the migration of existing language resource metadata to linked data. To ease the adoption of guidelines, LIDER has developed support tools such as LingHub for discovering language resources, and an TBX2RDF service for converting terminological resources to linked data.

5 LingHub and LLOD Cloud

LIDER has developed LingHub, an open repository of metadata in the domain of language resources and linguistic data.

It is a central access point for data discovery that aggregates metadata coming from other metadata repositories. LingHub allows people to discover the right data that are needed for a particular application.

The Linguistic Linked Open Data Cloud (LLOD Cloud) is a diagram automatically generated from the data contained in LingHub, and shows the current status of the LLOD Cloud.

The LLOD Cloud is a collaborative effort with the general goal to develop a Linked Open Data (sub-)cloud of linguistic resources.

LIDER has defined a clear path for the widespread adoption of LOD to support multilingual content analytics.

6 Guidelines and Reference Cards

LIDER has developed guidelines around the following topics:

- Bilingual dictionaries
- Multilingual dictionaries (BabelNet)
- WordNets
- Terminologies in TBX
- Linked Data corpus creation using NIF
- NIF-based NLP Web Services
- LLOD aware services
- LLD exploitation

In addition LIDER has developed reference cards to provide guidance for working with LLD:

- How to publish Linguistic Linked Data
- Language Resource Licensing - ODRL Reference Card
- Inclusion in the LLOD Cloud
- Data ID
- Discovering Language Resources with Ling
- NIF corpus
- How to represent crosslingual links
- Documenting a language resource in Datahub

LIDER roadmapping and further activities are summarized at https://www.w3.org/community/ld4lt/wiki/Lider_roadmapping_activities

7 Impact

- LIDER fostered a shared understanding about the representation of language and media specific information on a semantic level.
- LIDER laid the groundwork for reducing costs of adapting existing analytics solutions to multiple languages and across media boundaries.
- LIDER contributed to scientific excellence and coordination while assuring the industrial relevance of research activities.

LIDER has produced an ecosystem of free, interlinked and semantically interoperable language and media resources that will allow free and open exploitation of multilingual, crossmedia content across the EU and beyond.

8 Steps after LIDER

LIDER outcomes have been taken up by other European projects and by various industry players. The LIDER community will continue to promote and grow this uptake via the LD4LT and BPMLOD W3C community groups and the Open Knowledge Foundation Working Group on Open Data in Linguistics. The uptake of linguistic linked data will be focused on specific vertical domains including medicine, cultural heritage and the news sector.

LIDER is also collaborating with other projects in the linguistic linked data community in new pan-European infrastructures, including the EU's planned Connecting Europe Facility (CEF), specifically services for automated translation.

9 Additional information

Check our developments ...

<http://linghub.lider-project.eu>

<http://linguistic-lod.org/llod-cloud>

<http://lider-project.eu/guidelines>

... and our video:

<http://lider-project.eu/video>

LIDER coordinator

Prof. Dr. Asunción Gómez-Pérez

Universidad Politécnica de Madrid (UPM)

asun@fi.upm.es

Project 610782, CSA



FP7-610782

The following pages contain the content of the factsheet visualized as a flyer for printing.

Providing the basis for the Linguistic Linked Data Cloud

The logo for the LIDER project. It features the word "lider" in a lowercase, sans-serif font. The letters are colored in a gradient from orange to yellow. The letter 'i' is stylized with a small cluster of three dots above it, resembling a molecular or network structure. To the left of the word "lider" is a stylized graphic element consisting of three overlapping, rounded rectangular shapes in the same orange-to-yellow gradient.

<http://www.lider-project.eu>

LIDER aims to address the challenge of multilingualism in data and content to make them accessible any time, anywhere and in any language using linked data technologies



Background and the importance of LLOD for Big Data

The explosive growth in the volume, velocity and variety of content on the Web demands new approaches to content analytics. Language technology is the key to efficient understanding and analysis of multilingual content present as unstructured text and the linguistic content in diverse media streams. The effectiveness of language technology for content analytics is however critically dependent on the availability of relevant language resources, i.e. in the right languages, addressing the domains of interest and at sufficient volumes. Linked Open Data (LOD) based on the standards from the World Wide Web Consortium (W3C) offer a robust and sustainable solution to the problems of publishing, discovering, exchanging and managing language resources.

„Language technology is the key to efficient understanding and analysis of multilingual content present as unstructured text and the linguistic content in diverse media streams.„

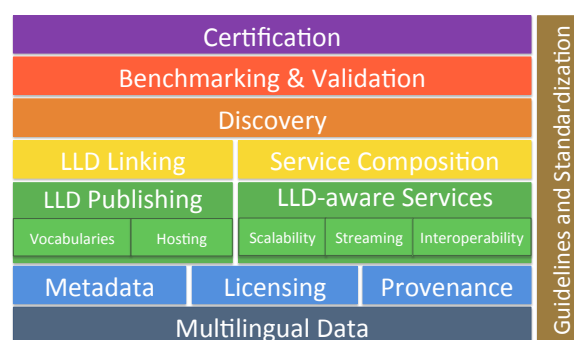
The LIDER project is completing a two-year programme of industry consultation, community building and development of technical guidance that defines a clear path for the widespread adoption of LOD to support multilingual content analytics. The result is an ecosystem of free, interlinked, and semantically interoperable resources from the realm of both language (“Linguistic Linked Data” representing corpora, dictionaries, lexical and syntactic metadata and conceptual models) and media (image, speech, video and its metadata).

Reference architecture and roadmapping

LIDER has published 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 of existing tools and initiatives that can be used to implement the architecture.

LIDER analysed input from surveys, roadmapping workshops, various market reports, funding agencies priorities and input from various research communities. This resulted in a detailed roadmap for the development and adoption of linguistic linked data.

Reference Architecture



The roadmap is formulated around three key application areas: global customer engagement, public sector / civil society, and linguistic linked data lifecycle and data value chain. For each of these areas, the roadmap provides several use cases including time lines, predictions and relevant actors, and these have been fed into broader Strategic Research and Innovation Agenda setting activities for H2020.

Reaching out to communities, building business use cases

LIDER has held seven roadmapping workshops assembling representatives from industry, the public sector, the voluntary sector and research to develop use cases and requirements for linguistic linked data. The workshops targeted different communities of practitioners including those working in: data management, multilingual web content, localization and content analysis. The feedback was gathered via the W3C’s “Linguistic Linked Data for Language Technology” (LD4LT) community group.

The results have been documented in workshop reports and use case and requirements analyses. They have led to the development of a research and innovation roadmap for linguistic linked data. LIDER has promoted these results on linguistic linked data and content analytics at a wide range of research and industry events.

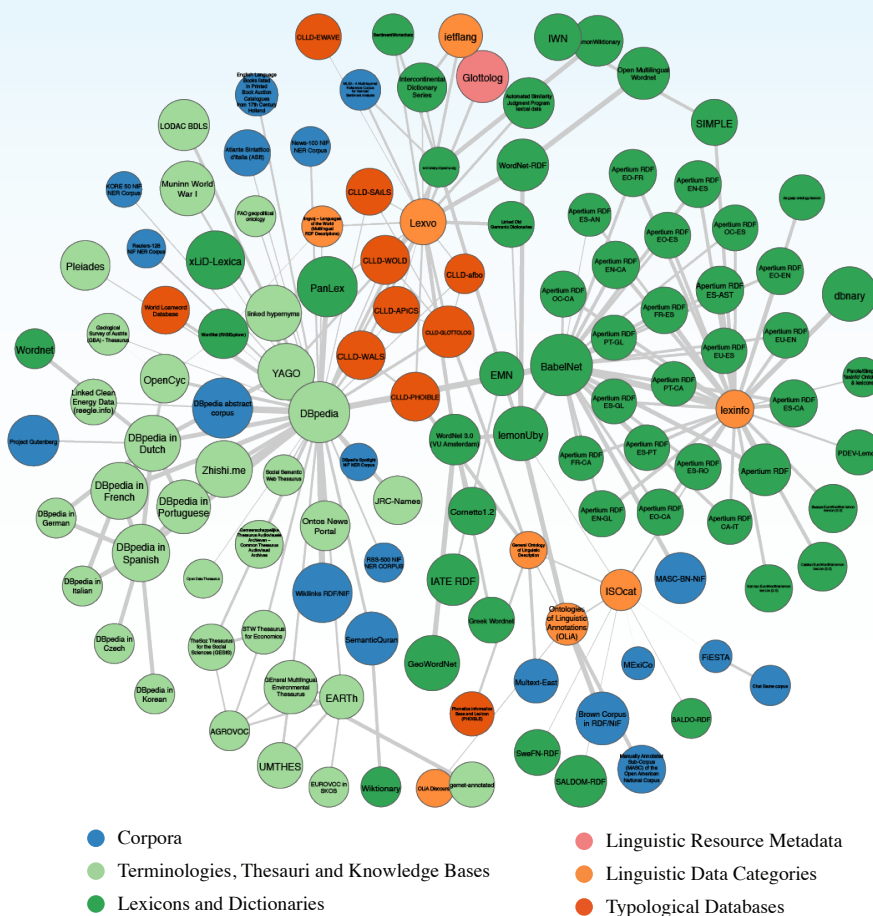
Guidelines and best practices for linguistic linked data generation: <http://lider-project.eu/guidelines>

Guidelines and best practices for Multilingual Linked Open Data

LIDER has led the W3C “Best Practices for Multilingual Linked Open Data” (BPMLOD) community group in developing best practices for creating multilingual linked data sources. Conversion guidelines between existing language resources, e.g. multilingual dictionaries, and linguistic linked data have been developed and documented in best practice reports.

LIDER has worked closely with the META-NET and the CLARIN communities on specifications for the migration of existing language resource metadata to linked data. To ease the adoption of guidelines, LIDER has developed support tools such as LingHub for discovering language resources, and an TBX2RDF service for converting terminological resources to linked data.

LLOD Cloud October 2015



LingHub and LLOD Cloud

LIDER has developed LingHub, an open repository of metadata in the domain of language resources and linguistic data.

It is a central access point for data discovery that aggregates metadata coming from other metadata repositories. LingHub allows people to discover the right data that are needed for a particular application.

The Linguistic Linked Open Data Cloud (LLOD Cloud) is a diagram automatically generated from the data contained in LingHub, and shows the current status of the LLOD Cloud.

The LLOD Cloud is a collaborative effort with the general goal to develop a Linked Open Data (sub-)cloud of linguistic resources.

„LIDER has defined a clear path for the widespread adoption of LOD to support multilingual content analytics.“

Guidelines

- Bilingual dictionaries
- Multilingual dictionaries (BabelNet)
- WordNets
- Terminologies in TBX
- Linked Data corpus creation using NIF
- NIF-based NLP Web Services
- LLOD aware services
- LLD exploitation

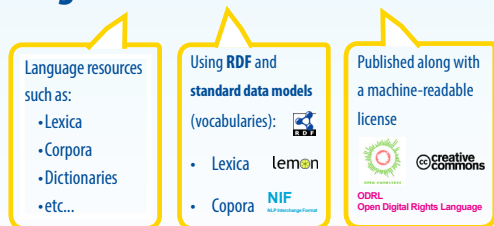
Reference Cards

- How to publish Linguistic Linked Data
- Language Resource Licensing - ODRL Reference Card
- Inclusion in the LLOD Cloud
- Data ID
- Discovering Language Resources with Ling
- NIF corpus
- How to represent crosslingual links
- Documenting a language resource in Datahub

Roadmapping and further activities: https://www.w3.org/community/ld4lt/wiki/Lider_roadmapping_activities

Overcome language and media barriers in content analytics

3LD Linguistic Linked Licensed Data



Impact

- LIDER fostered a shared understanding about the representation of language and media specific information on a semantic level.
- LIDER laid the groundwork for reducing costs of adapting existing analytics solutions to multiple languages and across media boundaries.
- LIDER contributed to scientific excellence and coordination while assuring the industrial relevance of research activities.

„LIDER has produced an ecosystem of free, interlinked and semantically interoperable language and media resources that will allow free and open exploitation of multilingual, crossmedia content across the EU and beyond.“

Steps after LIDER

LIDER outcomes have been taken up by other European projects and by various industry players. The LIDER community will continue to promote and grow this uptake via the LD4LT and BPMLOD W3C community groups and the Open Knowledge Foundation Working Group on Open Data in Linguistics. The uptake of linguistic linked data will be focused on specific vertical domains including medicine, cultural heritage and the news sector.

LIDER is also collaborating with other projects in the linguistic linked data community in new pan-European infrastructures, including the EU's planned Connecting Europe Facility (CEF), specifically services for automated translation.

Check our Developments...

<http://linghub.lider-project.eu>
<http://linguistic-lod.org/lod-cloud>
<http://lider-project.eu/guidelines>

... and our Video

<http://www.lider-project.eu/video>



mlider

<http://www.lider-project.eu>

Coordinator

Prof. Dr. Asunción Gómez-Pérez
Universidad Politécnica de Madrid (UPM)
asun@fi.upm.es

Project 610782, CSA

Join us!

www.lider-project.eu/get-involved

www.multilingualweb.eu



twitter.com/multilingweb



#LiderEU

