LIDER: FP7 – 610782

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

Deliverable number	**D4.5**
Deliverable title	**Report of the 1st Roadmapping Workshop**
Main Authors	**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	D5.1
Class Deliverable	LIDER EU-ICT-2013-610782
Version	1.1
Document due date	31 May 2014
Submitted	9 June 2014
Responsible	W3C/ERCIM
Reply to	fsasaki@w3.org
Document status	draft
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 report summarizes the main aspects of the LIDER roadmapping workshop, held 21st March in alignment with the European Data Forum in Athens.

The workshop offered several short talks around the topics of linguistic resources available in the linked data format and their potential impact on research and industry.

We explored the advantages of promoting free, open and interoperable language resources as Linked Open Data (LOD) to lower the entry cost to innovating in the use of such resources and technologies in research and industry. We also discussed how Linguistic Linked Licensed Data open new opportunities for market-based linguistic data exchange.

The event brought together stakeholders involved in the linguistic data value chain in order to explore new use cases and business models for reusing existing and future language (open or licensed) linked data resources and platforms for the joint development and hosting of multi–language data sets and services in different types of applications such as content analytics, machine translation, sentiment analysis, term disambiguation, etc.

More information about the event is available via the Ld4LT wiki, see

https://www.w3.org/community/ld4lt/wiki/LD4LT_Group_Kick-Off_and_Roadmap_Meeting

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	D4.5	Title	Report of the 1st Roadmapping Workshop
Workpackage	Number	4	Title	Community building and dissemination

Date of Delivery	Contractual	31 May 2014	Actual	9 June 2014
Status	version 1.1		final ■	
Nature	prototype □ report □ dissemination ■			
Dissemination level	public ■ consortium □			

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)	
Keywords	LIDER, roadmapping workshop, report

Version	Modification(s)	Date	Author(s)
01	First Draft	15/05/14	Felix Sasaki, DFKI / W3C Fellow
02	Final version	09/06/14	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 Goals of the Workshop

1.1 *Background: Linked Data and Language Technology*

Linked Data (LD) has proven beneficial in many new and unforeseen ways for Language Technology (LT) and the newly gained interoperability and availability of LT data and services is currently receiving industry adoption. This report summarizes the findings in this area from the first LIDER roadmapping workshop, held 21 March 2014 and aligned with the European Data Forum 2014 event. The workshop offered several short talks around the topics of linguistic resources available in the linked data format and their potential impact on research and industry.

We explored the advantages of promoting free, open and interoperable language resources as Linked Open Data (LOD) to lower the entry cost to innovating in the use of such resources and technologies in research and industry. We also discussed how Linguistic Linked Licensed Data open new opportunities for market-based linguistic data exchange.

The event brought together stakeholders involved in the linguistic data value chain in order to explore new use cases and business models for reusing existing and future language (open or licensed) linked data resources and platforms for the joint development and hosting of multi–language data sets and services in different types of applications such as content analytics, machine translation, sentiment analysis, term disambiguation, etc.

1.2 *The LD4LT Community Group*

The workshop was held primarily as the kickoff meeting of the W3C "Linked Data for Language Technology" (LD4LT) community group:

https://www.w3.org/community/ld4lt/

The LIDER project uses this group as its industry board. The nature of W3C community groups allows people to participate without fees or other financial requirements. As of writing the group already has 62 participants, including 35 persons from outside the LIDER consortium. The LD4LT group will continue to be used as the forum for organizing LIDER roadmapping workshops and for presenting the outcome of these workshops to the public. As one part of this exercise, the workshop reports will be made available through the community group wiki.

2 **The Workshop Sessions**

2.1 *Session 1: "Challenges"*

The first session was chaired by **Dave Lewis** (Trinity College Dublin, Ireland), co-chair of the LD4LT community group. In his opening presentation, he introduced the purpose and plans for roadmapping workshops and how their outcomes will feed into a roadmap for linguistic linked data and content analytics. A major goal of the workshops and the LD4LT group is to bring the communities around linked data and language resources closer together. The alignment of this roadmapping workshop with the European Data Forum is an ideal means to achieve this.

Asun Gómez Pérez (Universidad Politécnica de Madrid, Spain), coordinator of the LIDER project, introduced LIDER to the audience. Currently there is a huge heterogeneity of linguist resources, in terms of formats for resource representation, licenses for (re-) using resources, and relevant organizations & initiatives in the field (CLARIN, ELDA, ELRA, META-NET, etc.) Both language resource metadata and the data itself are hard to discover, integrate and process. Linked Data has the potential to serve as a language resource data integration layer. To this end, LIDER will
   •   agree on linked data vocabularies for describing resource metadata and content;
   •   provide standardized non-proprietary APIs for (meta)data access; and
   •   enable linkage with other resources.
Asun emphasized the need and opportunity of language resource related communities to work together with LIDER and use the LD4LT group as the common forum. These communities have a long term experience in working with language resources data and metadata. LIDER provides key players in the linked data field.
When representing linguistic resources as linked data, they should become "Linguistic Linked Licensed Data", that is: use standardized linked data technologies for describing the linguistic data and also machine readable license data.

Hans Uskorzeit (DFKI, Germany) talked about potential synergies between the fields of language technology and linked data. Hans introduced the META-NET initiative, which has worked for a long time on bringing the two communities closer together. The LIDER project is in a very good position to move this community building process forward.
Currently, more and more data is being published in vast amounts, including many new types of data. Big data becomes available as smart, linked and more or less clearly licensed data.
Such new resources have the potential to serve language technology applications in many areas, like: factoring linguistic knowledge for applications, interlinking semantic resources, providing repositories with semantic facts and improving machine translation.
Hans introduced the concept of a knowledge graph that can play a role in many such applications. However the manual creation and maintenance of such a graph is time-consuming and error-prone. Here technologies like information extraction can help to populate the knowledge graph. The so-called SAR-Graphs are a means to store information about every entity and how this is expressed in multiple languages. A next step may be to include hypotheses about uncertain knowledge on which users can apply filters.
Hans conclude the presentation by emphasizing that showcases with existing and new types of data could be quite helpful in demonstrating the mutual benefit of using linked

data and language technology together. Also, for community building purposes, it may help to broaden the goal of the LD4LT group, that is: not only linked data will be a means for (improving) language technology, but there is a benefit vice-versa: one should talk about "Linked Data and Language Technology".

Nicoletta Calzolari (CNR, Italy) talked about language resources, in particular about meta data for language resources. There have always been challenges of openness, interoperability, and collaboration. Infrastructural Issues play a major role for all natural language processing applications. The two communities of linked data and language technology need to work together to resolve such issues.

The "FlareNet recommendations" around dimensions like interoperability and sharing / distributions provide an important input in this development. One recommendation is to invest more in standardization to foster interoperability.

Currently more and more language resources are made available via the Web. Also, resources themselves are made available not as static entities but as services. An analysis of the "LRE map" reveals that 26% of the resources are provided as files, but already 74% are identified via URLs. The language resource community has taken this development into account by starting work on an international standard language resource number. This will contribute to raising interest in depositing and curating linguistic resources, and to the sustainability of linguistic resources.

An important aspect of this development is sustainability. Funding agencies who support the development of resources need to be made aware that persistent hosting, that is hosting of resources after the lifetime of the project, has to be considered and supported before the project planning.

Stelios Piperidis (Athena, Greece) described the META-SHARE architecture. META-SHARE aims at providing an open, distributed, secure, and interoperable infrastructure for the language technology domain. There are many different types of resources to take into account: monolingual and multilingual data sets, structured data (lexica, terminological databased, thesauri), unstructured data (raw text data) etc. META-SHARE provides access to these resources via an open infrastructure, in the form of a network of distributed repositories for language resources.

Currently there are 26 META-SHARE repositories. The underlying software and META-SHARE metadata model is identical; the repositories vary in number of resources or completeness of metadata. The metadata is harvested, and handling the data itself is the responsibility of the data provider.

META-SHARE has legal provisions for sharing of resources. There are licensing templates and standard licenses based on Creative Commons, the META-SHARE Commons licenses, or META-SHARE „No distribution" licenses.

To assure the quality of linguistic data services and linking, the accurateness and completeness of metadata records is critical. One has to avoid manual checks and develop automatic methods for filling in missing metadata values. Workflows are needed that allow such processes and fed back processing results into the language resource repositories.

In the discussion around META-SHARE, the importance of aligning a metadata model for language resources with other vocabularies like DCAT became clear. To allow re-use of linguistic linked licensed data, provenance information is crucial. Here the W3C provenance data model could serve as a means to identify data providers thoroughly.

2.2 *Session 2: "Impulses"*

The second session was chaired by Felix Sasaki (DFKI / W3C Fellow). The session consisted of a set of impulse statements highlighting a specific technical aspect of multilingual resources.

Phil Archer (W3C, Data Activity Lead) discussed multilingual ontologies. The mission of W3C is to provide technologies for everybody. Creating documentation of ontologies in several languages is an important aspect to achieve this goal. Current linked data projects often overlook this issue. Documentation is mostly multilingual and in English. W3C is now looking into whether they can establish support for multilingual ontologies in a sustainable manner.

The premise of this approach is to believe: an ontology, or speaking more generally a vocabulary, is more useful if it is described in several languages. However it is not clear whether this is really true. The DCAT "Data Catalog Vocabulary" specification is the first W3C vocabulary that is localized into several languages, to gather experience and to understand whether this is useful.

When translating a vocabulary, questions come up like what to translate (the underlying specifications, labels, both) and how to keep different items aligned. Some requirements cannot be applied to all languages, e.g. using CamelCase for naming does not work for scripts without upper versus lower case distinction.

Phil's presentation made clear that there is a need of best practices for ontology / vocabulary and specification localisation. Questions and comments about the presentation concerned tooling requirements, the separation of normative versus non-normative language, and the potential role of multilingual lexica. W3C community groups in the realm of linked data that can provide input are the OntoLex group and BPMLOD – both will be discussed below.

Felix Sasaki (DFKI / W3C Fellow) presented one potential solution to the ontology localisation task: the Internationalization Tag Set (ITS) 2.0 provides metadata ("data categories") to enhance localisation processes and may help both with translation of schemas documentation. ITS ready-to-use tooling is available for XML and HTML content, so ITS could be applied to the RDF/XML serialization of ontologies.

The OKAPI framework is a prototypical example of such tooling. It provides modules for building localisation workflows, various ready-to-use WYSIWG editors and workflow support including various language technology tasks like machine translation.

OKAPI shows the maturity of XML and HTML based workflows. The issue with linked data localisation is of course that linked data can be stored in many serializations. The use of RDF/XML is rather decreasing. One might define ITS 2.0 metadata in a linked data serialization independent manner, but this has the drawback of missing tool support.

Arno Scharl (WebLyzard and Modul University) described various applications WebLyzard is working on. A long term focus was on unstructured information (web sites, non-private email etc.). Recently structured (linked) data sources have become of interest. Applications that process various types of content are: sentiment analysis, strategic positioning and topic identification, and corporate knowledge management.

Arno presented various use cases in which linked data may play a role, like the Media Monitor Austria or NOAA (National Oceanic and Atmospheric Administration), and relevant projects like Carbonet, Sentiment quiz, or the Ucomp project.

A requirement for building analytics applications is the availability of language resources, e.g. a context-specific sentiment lexicon. The LIDER project / the LD4LT group are in a good position to formulate requirements on these and other multilingual language resources.

These presentations were followed by short briefings about three community groups.

Philipp Cimiano (Bielefeld University) is co-chair of the forehand metioned W3C Ontolex community group. There are M:N relations between items in ontologies and lexical entries. For example, different lexicalizations can be made for the same ontology. The discussion of appropriate & multilingual labels is just one first step: one should create linguistic objects with their own semantic, internal structure.
The OntoLex group is working on the underlying model, taking a lot of standardization work for linguistic lexicon models into account. A first model will be released later this year. A crucial step to assure adoption of the OntoLex model is to provide implementations, like APIs for accessing ontology-lexicon resources or validation tools and services.

Jorge Gracia (UPM) is co-chair of the "Best Practices for Multilingual Linked Open Data Community Group". The group has been founded after the March 2013 MultilingualWeb workshop. Jorge reported on the current state of activities. Main topics and use cases have been identified. Patterns about using and creating multilingual linked data properly are being created.
Main topics are: naming of URIs, textual information, linking within or between languages, ontologies and vocabularies, quality and tools of multilingual linked data. Re-occurring topics both in the BPMLOD group and at the LD4LT workshop are licensing and legal aspects of resources.
Example use cases are localisation workflows and ontology or application localisation. So far the group has worked on patterns related to naming and textual information.

Gabriela Vulcu (Insight, National University of Galway, Ireland) reported on the Sentiment and Emotion community group. The goal of the group is to bring together initiatives working on representation of emotion and sentiment information. Building on existing formats like EmotionML, Marl and Onyx, a linked data representation will be developed. The emphasis will be on sentiment polarity and emotion categories.
The group has started recently. Participants include members both from academia and industry.

2.3 *Brainstorming: Work Items in the LD4LT Group*

The workshop participants were asked about key activities for the LD4LT group. One purpose in this exercise was to detect industrial needs. For the translation and localisation industry, a key goal is to achieve faster, less cost, and efficient translations. For this various requirements on linked data resources can be formulated: they need to be of good quality, in standard formats and inherently multilingual.
For all use cases discussed at the workshop it is a strong requirement that language resources easily can be discovered and accessed. The aim is not to replace existing language resource infrastructure; working with linked data = RDF representations an SPARQL as a query language is just an added value.

A set of common goal for the language resource and the linked data community could be identified:

- Provide proper information and documentation, and language resource metadata accessible as linked data as a first step.
- Assure that the past and current work undertaken in the language resource community becomes more visible in the linked data community.
- Provide unique identifiers for language resource as a first step towards using HTTP URIs.
- Especially from the perspective of companies, take into account that translators and companies are not keen on sharing resources freely. For them the main gain by linked data concerns interoperability.

The last goal makes clear that LD4LT needs to work on understanding the real needs of workers in the language industry and in the language technology industry domain.

2.4 *Session 3: "Position Statements and Industry Perspective"*

The session was chaired by Philipp Cimiano (Bielefeld University). It consisted of short position statements conveying use cases and requirements from various industry domain.

Athena Vacali (Aristotle University, Greece) described an application of linked data for social network analytics. The main idea is to identify important themes through patterns in Web 2.0 content, enrich it with existing linked data representations, and provide the results in a machine-readable form. This data then can be exploited in further services.

Martin Benjamin (EPFL: Swiss Federal Institute of Technology) introduced kamusi, a dictionary project with the aim to provide linguistic data for all languages, both for humans and in a machine readable form. Kamusi constitutes a network of lexical items, taking also the ambiguity of terms into account. Various linguistic information is provided, related to morphology, grammar patterns, family relations etc. Also, Kamusi allows to interlink with other resources like WordNet.

Tatiana Gornostay (Tilde, Latvia) introduced TaaS ("Terminology as a Service") project. TaaS addresses challenges in Web-based terminology services, like: instant access to up-to-date terms, user participation in acquisition and sharing, and efficient terminology reuse. TaaS also eases the domain adaptation of statistical machine translation (SMT) systems by dynamic integration with terminological data. The exchange of terminology information currently relies on the TBX format. A next step is to work on linked data representations of TBX, to allow for combining terminology information with other linguistic or general linked data resources.

John McCrae (Bielefeld University, Germany) described the conversion of WordNet into linked data. There are WordNets in 70+ languages available. So far interlinkage to the English WordNet is not stable e.g. in terms of identifiers. The WordNet-RDF representations provides this stability, labels in several languages, and linkage with other WordNet representations.

Roberto Navigli (Sapienza University of Rome, Italy) introduced BabelNet 2.0, a multilingual semantic network including encyclopedic (from Wikipedia) and lexicographic (from WordNet and OmegaWiki) entries. The network covers 50 languages and provides 50 million word senses. A representation of BabelNet as an RDF lemon model has been

created, consisting of more than 1.026 billion triples. One issue to be resolved for re-use is licensing. BabelNet does this via an onion-like model with BabelNet core being unrestricted data. The Babelfy service allows to analyze textual content both in terms of entities metioned and word denoting senses.

Costas Nadalis (TMServe, Greece) provided a perspective from the daily work situation of language workers. Language data comes in lots of proprietary formats. For usage in e.g. statistical machine translation training, data cleansing is inevitable. Various levels need to be taken into account: character encoding, markup, segmentation and quality (e.g. resolve misspellings). A definition of data types including application scenarios and cleansing procedures would help to get clear understanding of data quality in a given application context.

Giannis Stoitsis (AgroKnow Technologies, Greece) described best practices for providing online multilingual services. These have been developed in the Organic-Lingua project as a basis for a multilingual European online service: Organic.Edunet. Some best practices are related to general Web service architecture, e.g. "provide a RESTful interface". Others are specific to multilingual technologies, e.g. "select the best machine translation service". Giannis emphasized also the need to take user requirements for multilingual content into account and to apply analytics about user behavior to get the relevant feedback.

Phil Ritchie (VistaTEC and Digital Linguistics, Ireland) made three statements from the point of view of a language service provider, as input to the LD4LT group. First, provenance information is crucial in translation processes for all actors: content creators, translators, and linguistic reviewers. Second, for translating marketing material, metaphors are crucial to translate (or rather adapt) across languages. Translators could benefit from getting suggestions for previously translated / suggested metaphors and analogy models. Third, identifying semantic similarity between content items in one or across languages could help to improve semi-automatic quality assurance.

Daniel Vila-Suero (UPM, Spain) provided details of an approach towards Linguistic Linked Licensed Data (3LD). An analysis of the linked data cloud reveals that its openness in terms of licenses for data sets is quite heterogeneous. The concept of 3LD is to publish linguistic linked data along with machine-readable licenses. A prototype processing 3LD has been developed, including a license-aware linked data server and a data policies and license manager.

Ilan Kernermann (kdictionaries, Israel) described trends in the dictionary domain, e.g. from print to digital, from passive readers to interactive users, or from tangible to virtual. kdictionaries develops dictionary data in 40+ languages and cooperates with lexicographers, translators, publishers and many others. Ilan provided details about the kdictionaries global series, a lexicographic data set for several languages. One aspect of the global series is the integration with natural language processing and language technology tasks.

Uroš Milošević (Institute Mihajlo Pupin) introduced the Rozeta tool. It provides automatic extraction of dictionaries in STRUTEX representation as linked data, semantic enrichment through link discovery, and manual revision and authoring. Using the lexical information, document similarity and document classification can be applied. Rozeta has been developed in the LOD2 project.

2.5 *Session 4: "Requirements and Use Case Discussion"*

Georgeta Bordea (INSIGHT, Ireland) provided a summary of the LD4LT online survey about use cases and requirements for linguistic linked data. Details about the survey and about the current state of use cases are provided in the LIDER deliverable "Business use cases for the use of Linguistic Linked Data in content analytics processes - Phase I". The outcomes of this session are a key input to the work undertaken in WP1.

The participants have identified the following use cases.
- Linked Data
  - Modelling the life cycle of LRs representation as LOD all the steps
  - Document (e.g., news) similarity using heterogeneous linked data resources
  - Ontological resources, linked to dynamic data communication
  - Third party quality annotation
- Machine Translation
  - Disambiguate senses on source side
  - Multilingual dictionaries - dynamic creation/aggregation
  - Parallel (MT) data discovery
  - Global translation memories
  - Live localization: Use recurring concepts to make multilingual sites on-the-fly (e.g, bookings, recipes)
- Sentiment Analysis
  - Trend analysis
  - Brand monitoring
  - Sentiment-based recommendation
  - Sentiment analysis
- Other
  - Organizational / corporate knowledge graphs
  - Build a knowledge interlinked graph in various languages
  - Organic. Edunet multilingual discovery service
  - Cross-lingual data discovery / navigation
  - Use of NLP services and language resources for data cross-lingual interlinking
  - Cross-lingual and/or cross-format resource linking
  - Accessibility and availability of language data

In addition, the participants formulated the following requirements. Some of these are related to linguistic linked data in general, others are specific to selected use cases.
- Linked Data
  - Minimal core metadata vocabulary (RDFS) for language resources
  - Common Owl vocabulary to represent language resource metadata
  - Simple service to validate metadata and current status of language resources (e.g. are they available?)
  - Unique "Ontology" metadata schema For language resources
  - Community agreement on shared vocabularies
  - Easy identification of language resource data documentation etc. in an uniform way
  - Resources should be domain agnostic

- o Ontology for support "my" language resource life cycle
- o Linkage to non language resource catalogues, to foster new usage scenarios
- o Preserving of information available in existing meta-data
- o Integration with non RDF tooling
- o Promotion of impactful sets Of language resources in linked data
- Data Modelling
  - o Model usage of data
- Business Models
  - o Find business models for sharing data
  - o Pricing models & commercial transactions
  - o Open source vs. commercial
  - o License retrieval, management and control
  - o Access control - pay-portal
  - o Licensing
  - o IPR
- Machine Translation
  - o Align entities from different language resources
  - o Disambiguation by sense
- Other
  - o Handling of inflected forms / variables / accentuation in automated way
  - o Use existing formats
  - o Conversions to industry formats like TMX, TBX, etc.
  - o Open (documental) standards
  - o Combination of heterogeneous data packages & automatic metadata description of them
  - o Compatibility with already existing and widely used cat-tools (Computer-Aided translation)
  - o Tool agnostic representations
  - o Sharing resources
  - o Usage of language resources & involvement of the community
  - o Crowdsourcing curation

3 Conclusions and Next Steps

The first LIDER roadmapping workshop was a great success in several areas. First, it helped to bring the language technology and linked data community closer together. A tangible outcome is the conversion of existing language resource metadata into linked data. This work item is now moving forward within the LD4LT group.

Second, the workshop helped to identify use cases and requirements from various industry and research perspectives. Some general best practices and a general architecture for providing linguistic linked data, both metadata and the data itself, are needed. Other W3C community groups like BPMLOD and OntoLex and the "Sentiment Analysis" group are working on selected topics. For some of these the LIDER project has related EU project work packages, tasks and deliverables. To allow broad community input, the LD4LT participants will be able to provide input to these without the need to engage into the detailed LIDER project plan. Most people liked the opportunity to interact with experts and that some asked for more discussion opportunities.

Third, the workshop has made various industries aware of the opportunities for linguistic linked data. These opportunities will be discussed in more detail in upcoming roadmapping workshops, and the results again will be disseminated in the LD4LT community group.

Participants made various concrete suggestions for future LIDER workshops:

- One should provide pointers to resources developed within the project (e.g. data sets, implementations, publications, summarizes of events etc.).
- Future workshops should use summaries of previous events as an introduction
- Legal experts should be brought onto the table.
- The work plan of the LIDER project should be made clear to the public including milestones and activities.
- Showcases on benefits of linked data in various domains like agriculture, medical, education etc. would be beneficial.