



LIDER: Building Free, Interlinked, and Interoperable Language Resources

Asunción Gómez-Pérez
Philipp Cimiano

Multilingual Web Workshop
Riga, 28th of April. 2015



POLITÉCNICA



NUI Galway
OÉ Gaillimh

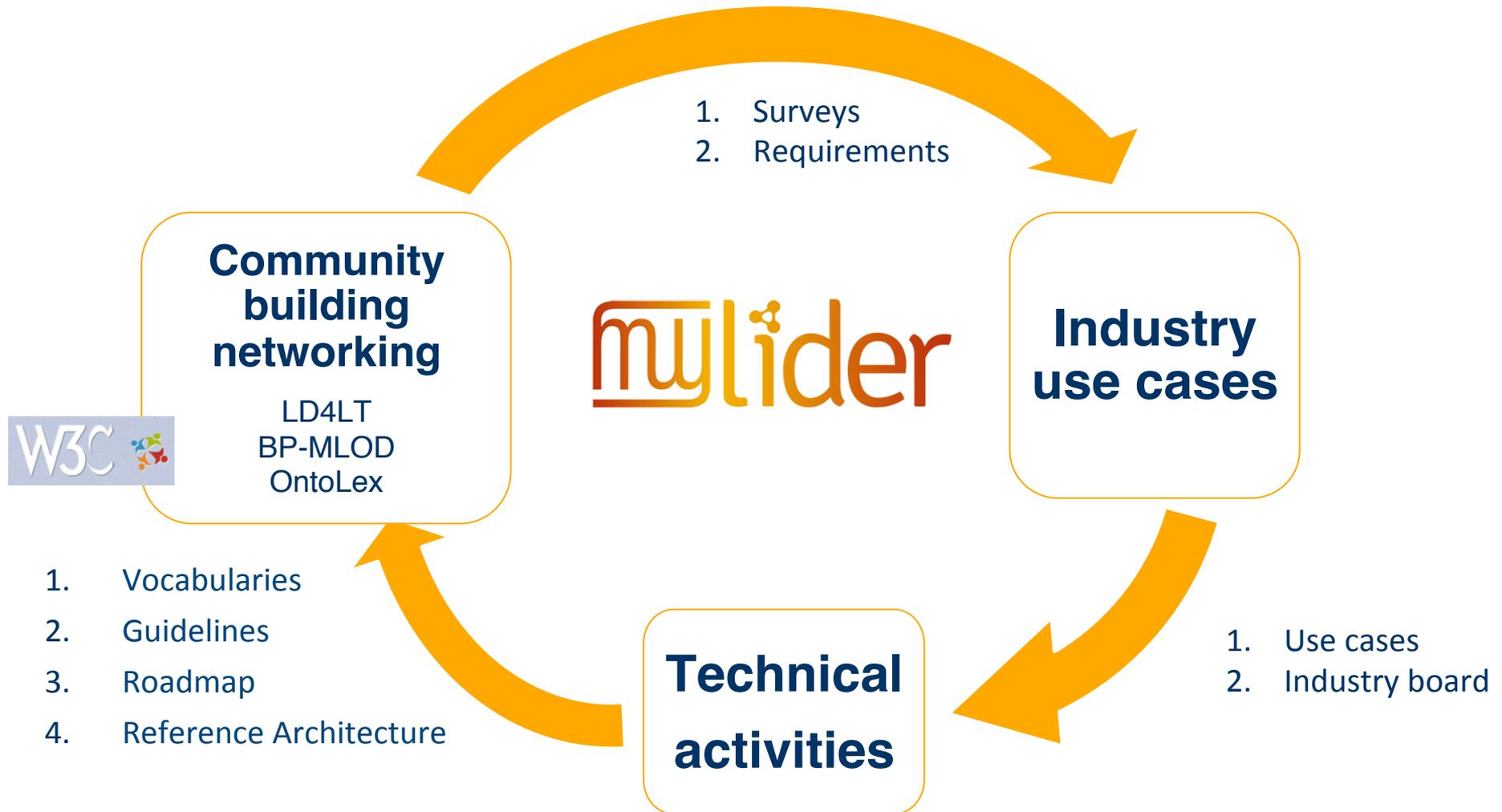
UNIVERSITÄT LEIPZIG



Universität Bielefeld



W3C®



3LD

Linguistic Linked Licensed Data



*Language resources
such as:*

- Lexica
- Corpora
- Dictionaries ..

*Using **RDF** and
standard data
models
(vocabularies):*



- Lexica lemon
- Corpora NIF

NIF
NLP Interchange Format

*Published along with
a **machine-readable
license.***

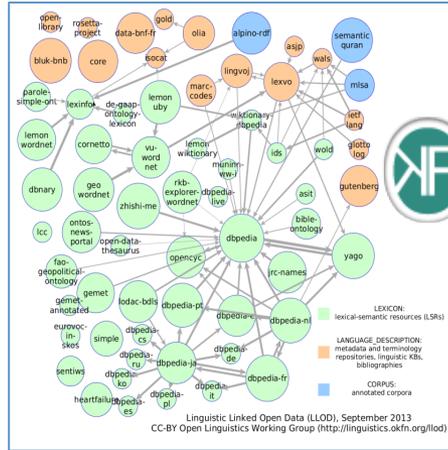
ODRL
Open Digital Rights Language



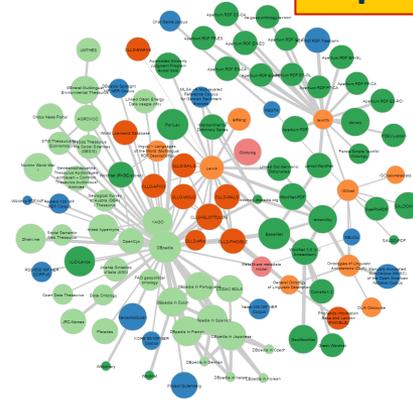
Jan. 2013



Sept. 2013



Sept. 2014



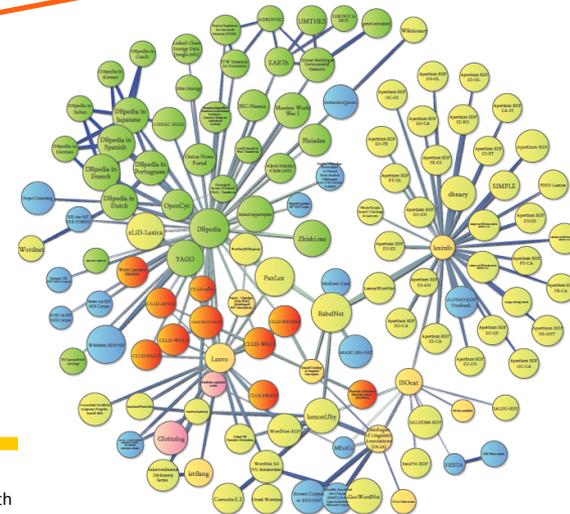
Legend

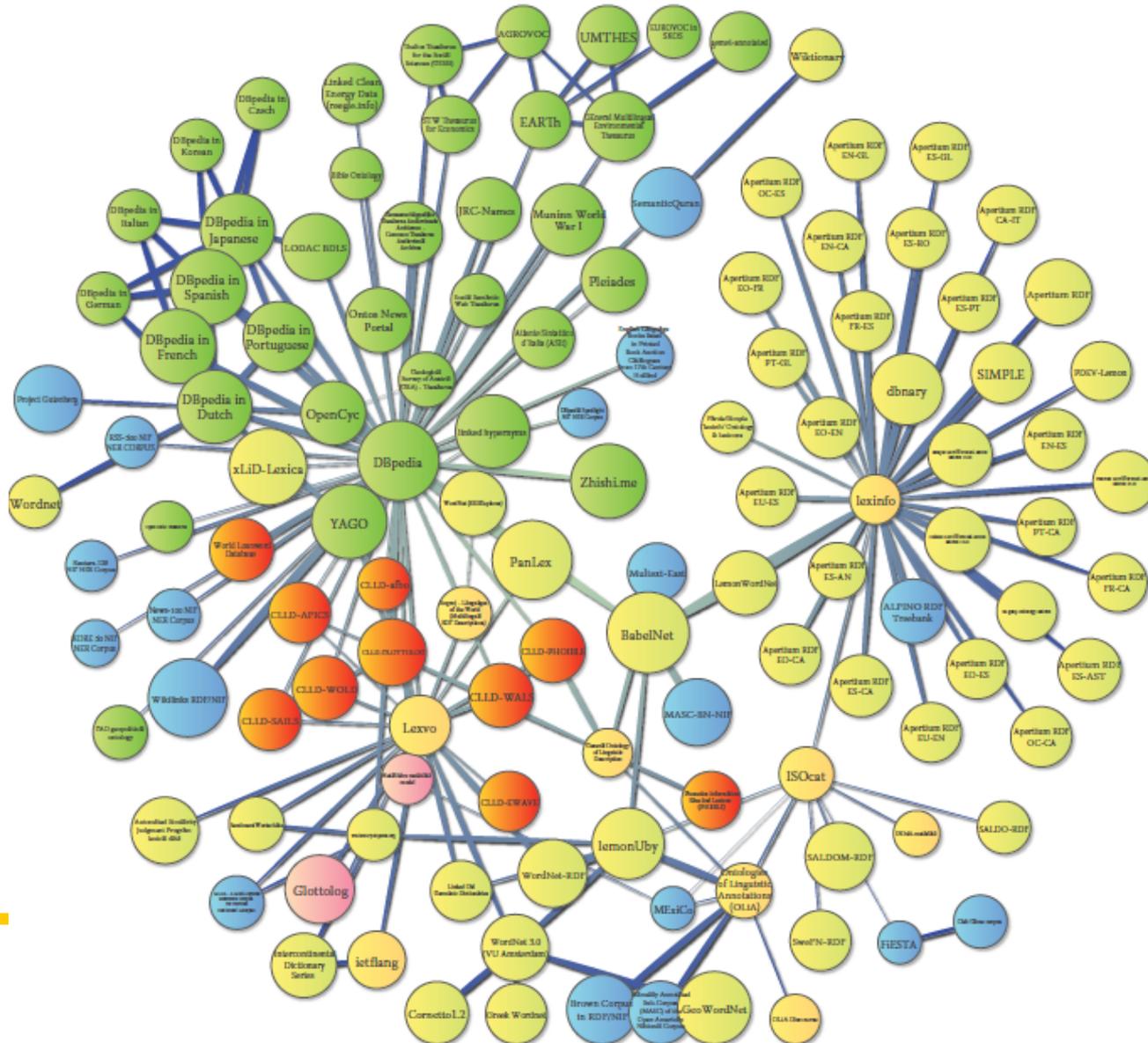
- Linguistic Resource Metadata Repositories
- Linguistic Categories
- Lexicons and dictionaries
- Corpora
- Other metadata
- Terminologies, thesauri and knowledge bases
- Other lexical resources

April. 2015



- Corpora
- Terminologies, Thesauri and Knowledge Bases
- Lexicons and Dictionaries
- Linguistic Resource Metadata
- Linguistic Data Categories
- Typological Databases





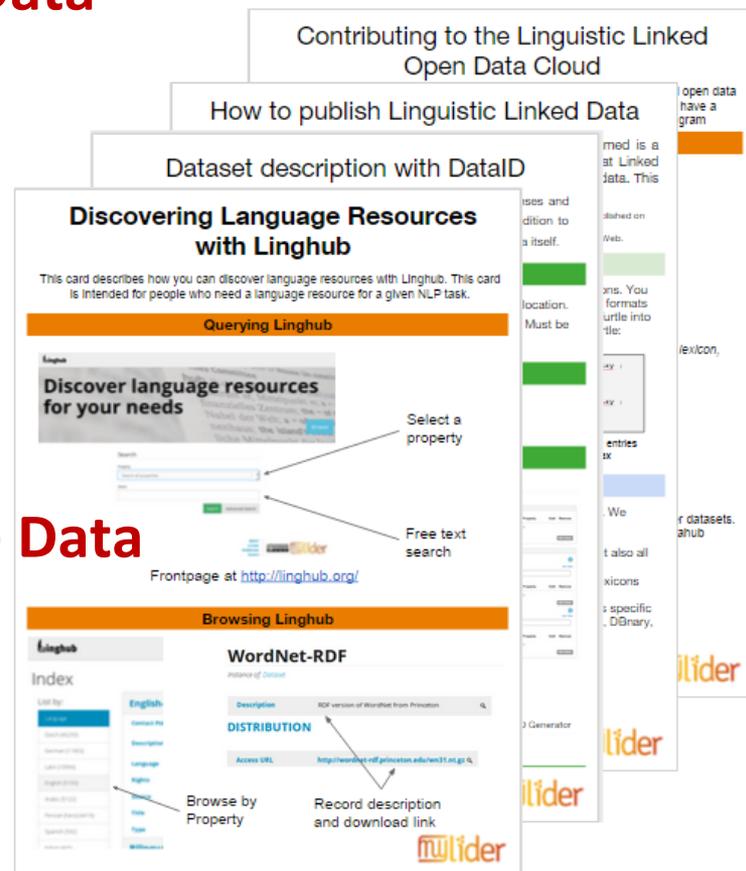
- Corpora
- Terminologies, Thesauri and Knowledge Bases
- Lexicons and Dictionaries
- Linguistic Resource Metadata
- Linguistic Data Categories
- Typological Databases

1. Best practices for Multilingual Linked Data Publication

- Practices for Naming (URIs)
- Practices for Dereferencing
- Practices for Textual Information
- Practices for Linking
- Practices for Language Identification

2. Guidelines for Linguistic Linked License Data

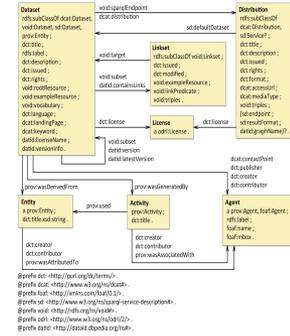
- Wordnets
- Multilingual Lexicographic resources
- Bilingual Dictionaries
- Terminologies in TBX
- NIF-based NLP Web services



1. Metadata (LD4LT@W3C)

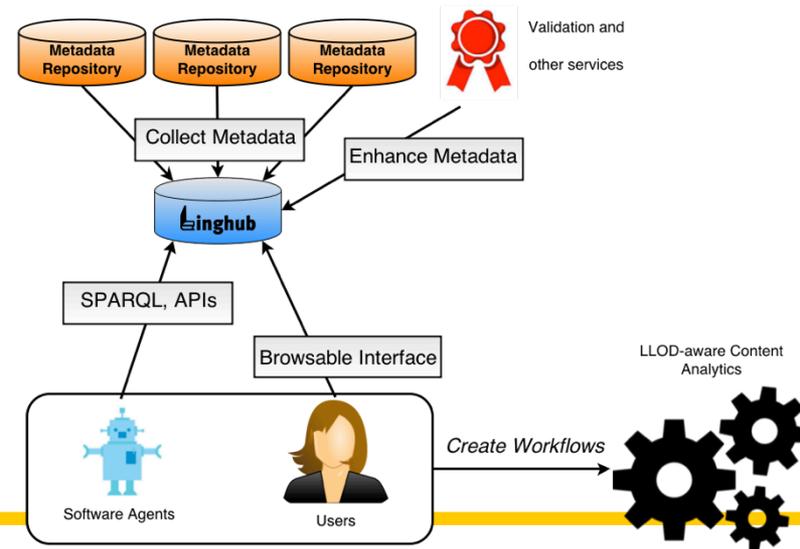
- Metadata OWL ontology @ LD4LT W3C group
 - based on UPF's Metashare model
 - Expanded with PROVENANCE and LICENSE
 - Backwards compatible with MS and LREMap models
 - In agreement with members of LD4LT W3C group

- Exposure of Metashare, Clarin, LREMap, datahub metadata as LD

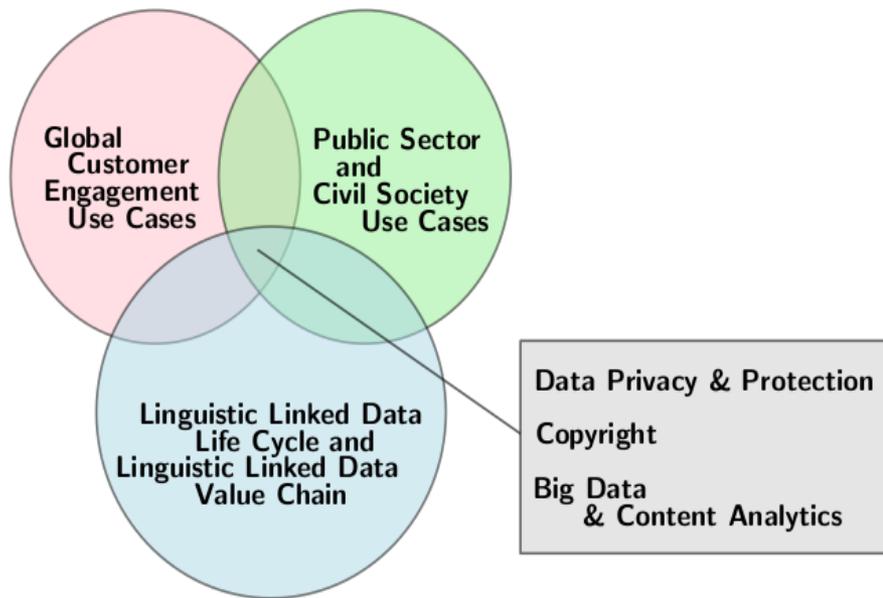


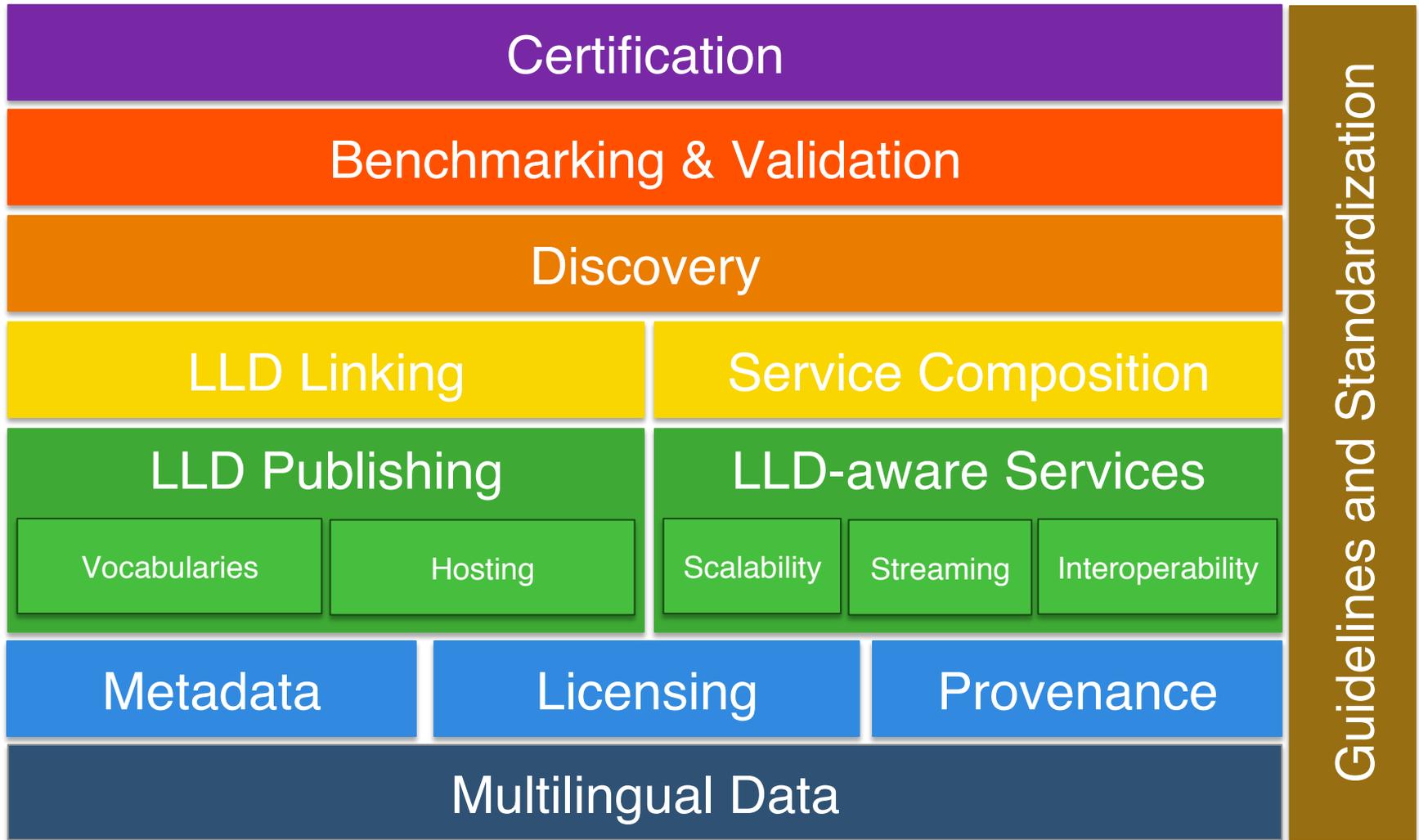
2. Linghub helps you to search LRs

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



- LIDER deliverable D3.2.1
- Product of 6 roadmapping workshops organized by the LIDER project:
 - EDF@Athens 2014 (43 part.)
 - MW@Madrid 2014 (44 part.)
 - LocalizationWorld@Dublin 2014 (40 part.)
 - SEMANTICS@Leipzig (51 part.)
 - PSI @Lisbon 2014
 - PSI @ Timisoara, 2015
- Includes the perspective of three W3C community groups (LD4LT, BMLOD and ontolex), with between 80 and 90 participants.
- Brings together the perspectives of hundreds of stakeholders (industrial and academic)
- Three main application fields (see figure) and three orthogonal topics
- Horizon: 3, 5 and 10 years





1. How do we support SMEs in providing information about products, their features in 24 languages?
2. How do we ensure consistency and homogeneity of product information so that cross-vendor comparison becomes possible?
3. How do we support online dispute resolution in 24 languages effectively?
4. How do we support finding previous online dispute resolution similar cases, also across languages?
5. How do we link datasets (e.g. business registry data, open government data) etc. across languages effectively to allow for mashups, cross-country comparisons, repurposing of data etc.?
6. How do we support the discovery of relevance public sector datasets?
7. How do we ensure interoperability of public services across borders?
8. How do we ensure quality and trust in data?
9. How do we support clearance of licensing issues to ensure compliance in data use?
10. How do we support interaction with customers in multiple languages?



www.lider-project.eu



Join the community

www.w3c.org/community/ld4lt



twitter.com/multilingweb

Hashtag: **#LiderEU**

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