



# The LIDER Roadmap in a nutshell



Philipp Cimiano  
(representing the LIDER Project)  
LD4LT Teleconference  
February 19th, 2015



POLITÉCNICA



NUI Galway  
OE Gaillimh

UNIVERSITÄT LEIPZIG



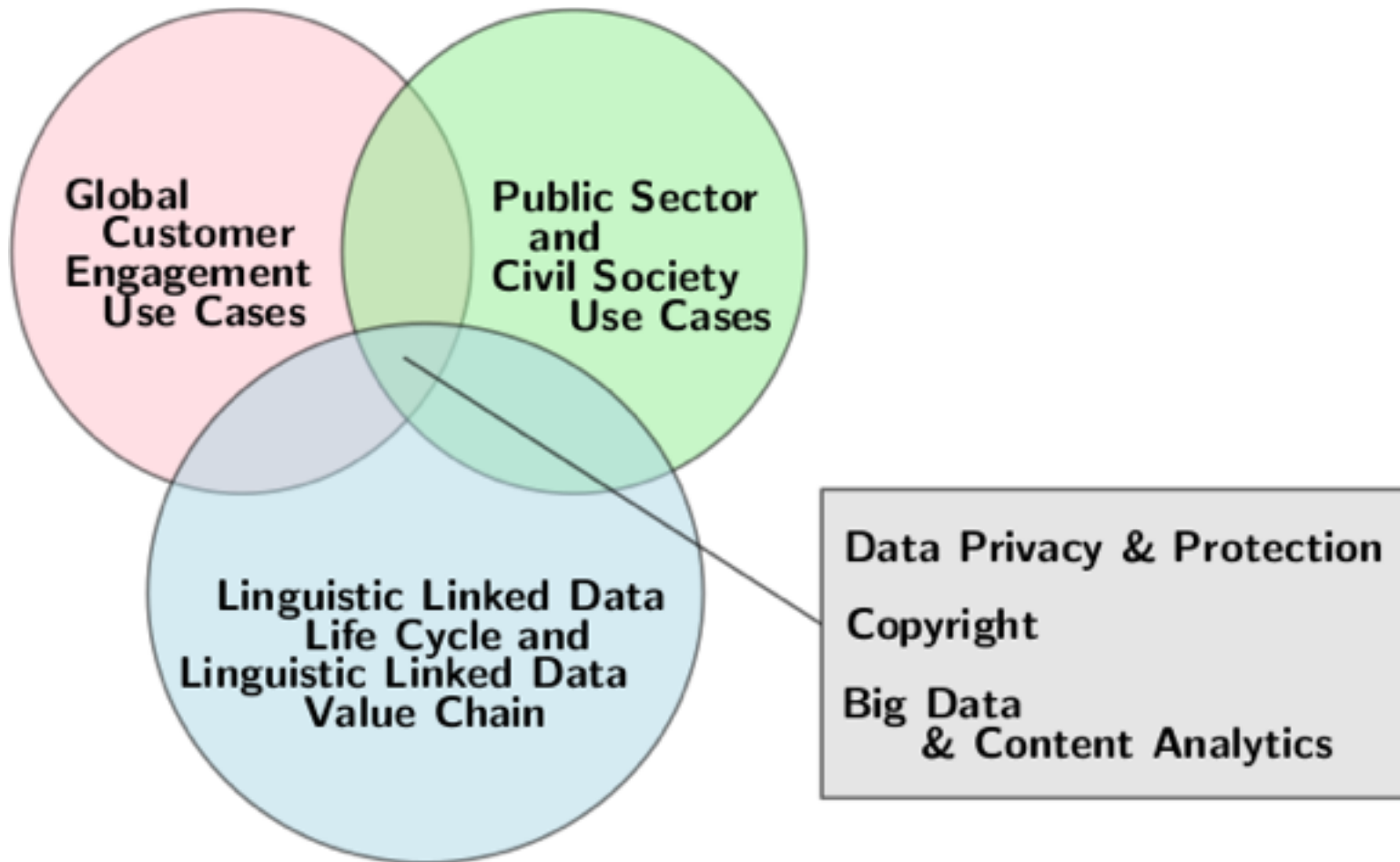
Universität Bielefeld



W3C®

- **LIDER**: Support Action funded by EC under FP7 to develop use cases, roadmap and reference model for linked data enhanced content analytics
- **Roadmap**:
  - LIDER deliverable D3.2.1
  - Product of **4 roadmapping workshops** organized by the LIDER project:
    - EDF@Athens (43 part.)
    - MW@Madrid (44 part.)
    - LocalizationWorld@ (40 part.)
    - Leipzig@SEMANTICS (51 part.)
  - includes the perspective of three **W3C community groups** (LD4LT, BMLOD and ontalex), with between 80 and 90 participants.
  - brings together the perspectives of hundreds of stakeholders (industrial and academic)
  - **three main application fields** and three orthogonal topics
  - **Horizon**: 3, 5 and 10 years

- **General IT Trends** (Gartner Reports)
- Global discourse on **information society** (Global Information Technology Report 2014)
- **Needs in Content Analytics** (Alta Plana Survey, 4th LIDER Roadmapping Workshop)
- Perspective of **public R&D funders** (Horizon 2020, CEF)
- Perspective of **European Commission: Digital Single Market (DSM)**
- **Content production and delivery industry** (TechRadar Report by Forrester)
- **Data Protection and Privacy** (Rethinking Personal Data: A New Lens for Strengthening Trust, World Economic Forum)
- **Research Perspective**



- robust analysis and **understanding** of voice of the customer across languages and modalities, including **intention recognition**
- **support cross-channel, localized and consistent customer experience and communication**
- **agile content creation and integration**
- **contextualization and personalized content delivery**

## Challenges:

- **repurposing of content and storytelling feasible and practical**, and to lower the cost for doing so
- **multimodal story generation from multiple sources** including text, video and other modalities as well as new methods for re-purposing and composing heterogeneous content for different challenges, natural languages and audiences.
- **best practices for linked data based content publishing** as well as experience reports on the adoption of such best practices in verticals
- **potential of linked data to connect different media beyond separate annotations** (cross-media links) and the potential of linked data to share such data across companies.
- creation of a **seamless network of data and knowledge that spans multiple modalities, languages, structured and unstructured data**, as well as open and closed datasets in a way that is respectful to intellectual property (IP) and corresponding licenses

In the future, techniques will be needed to analyze and interpret the voice of the customer with:

1. a **high level of accuracy**,
2. **across natural languages and modalities**, and
3. analyzing sentiment at **deeper levels beyond mere polarity** in order to also recognize **the intent of a user** as a basis to generate actionable knowledge

We will need techniques to foster the creation of deep personalized profiles of users to support personalized, situated and contextualized content delivery:

- exploit domain-specific background knowledge available as linked data to create **semantic user profiles**
- Linked data technologies can also play a key role in **linking profiles of users across channels and sites**.

- **Creation of a single digital market:**
  - cross-border public and government services that interoperate
  - data value chains across the borders of countries
  - techniques for localization and translation
- **Requires:**
  - identification of **key domains** in which cross-border communication is essential
  - methodologies for development of **shared terminologies and vocabularies**



## Key issues/challenges:

- The development of **shared ontologies** of key administrative and legal concepts across Europe
- **Linking of vocabularies and ontologies** existing in different countries and jurisdiction to foster interoperability
- Development of **declarative specifications of workflows and processes**, so that tools can reason about them and compose them to achieve some task
- Methods for **collaborative ontology creation across languages and countries**
- Exploitation of **terminologies and ontologies to ensure consistency of communication** in public administration

Relevant for DSM:

- 1) **Common European Sales Law** => localized product portfolio
  - 2) **Alternative/Online Dispute Resolution (ADR/ODR)** => mediation between players in different countries
  - 3) **Reuse of Public Sector Information** => linking of data across languages
- Linked data technologies have the potential to impact the current localization and translation market and processes by providing **more flexible ways of publishing and exploiting multilingual datasets** including parallel corpora, terminologies and translation memories
  - **Best practices and standards for publishing parallel texts** as part of the linguistic linked data cloud
  - **New paradigms in which content creation and translation are intertwined** in the sense that machine translation can be exploited in bootstrapping content creation and vice versa will become feasible in the near future.
  - **Linked data as an enabler of high-quality personalized translation**
  - **Translation of terminologies as a way to ensure terminological consistency** across national players and stakeholders.

Main challenges by content analytics sector:

- **lack of data**
- **heterogeneous formats and lack of standard APIs**
- **high effort** to customize analytics solutions
- **policy-compliant exploitation** of data

**Requires:**

- Development of **ecosystem and market:**
  - supporting **discovery** language resources
  - compliant **exploitation of open and closed** resources
  - data and services easily **combinable, exchangeable** and **repurposeable**
  - **trust and certification**
  - **benchmarking**

Data Privacy and Protection

Copyright

Big Data

## Key issues/challenges:

- Ensure that users can specify their **privacy rules and levels** and that **compliance to these rules** can be monitored over the whole data lifecycle
- **Best practices and standards for licensing terms and conditions of use as machine readable RDF** data, attached to data through the whole lifecycle
- **Access control to Linked Data** (views, access levels, etc.)
- **Architectures for scalable, parallelizable NLP services** that can be run on the cloud

Thanks for your attention!  
Any comments, questions,  
...?