



# 4th LIDER Roadmapping Workshop

Philipp Cimiano, Univ. Bielefeld  
Sebastian Hellmann, Univ. Leipzig  
Bettina Klimek, Univ. Leipzig

representing

the LIDER Project Consortium

# LIDER Project

- **Support and Coordination Action** (CSA) funded by the European Commission under FP7
- **Duration:** 2 years
- **Goal:** Define Use Cases, Roadmap and Architecture for Content Analytics based on Semantic Technologies (Linked Data) for 2020 and beyond

# LIDER Roadmapping

## Workshop: **Goals**

- **Goal:** define a roadmap report to guide R&D investments by European Commission
- Methodology:
  - **gather use cases and requirements** from providers and content analytics users
  - **prioritize use cases, assemble them into a R&D agenda** to be delivered to the EC

# LIDER Roadmapping Workshop: **Program**

- 9:00 – 9:15: Language Meets Knowledge in Digital Content Management – **Tatiana Gornostay, Tilde**
- 9:30 – 9:45: Generating Multilingual Lexicographic Resources – **Ilan Kernerman, K Dictionaries**
- 9:45 – 10:00: Marin Dimitrov, **Ontotext**
- 10:00 – 10:15: Resources! Resources! Resources! – **Heiko Ehrig, Neofonie**
- 10:15 – 10:30: Software-supported Bibliographic Recording and Linked Data - **Mark Zöpfgen, DNB**
- 10:30 - 11:00 Coffee/Tea Break

# LIDER Roadmapping Workshop: **Program**

- 11:00 – 11:15: Social Media Monitoring: from Sentiment to Intention – **Massimo Romanelli, Attensity Europe GmbH**
- 11:15 – 11:30: Non-reactive Concept Mapping to Elicit Consumer Perception – **Marc Egger, Insius**
- 11:30 – 11:45: Linked Data for Content Analytics in CELI- **Alessio Bosca, CELI**
- 11:45 – 12:00: Content Analytics for Media Agencies – **Oscar Muñoz, HAVAS Media**
- 12:00 – 12:15: Applied Insights: Computational linguistics and semantic analysis as part of business workflows – **Andreas Nickel, Ferret Go**
- 12:30 – 12:45: Multilingual Situation-based Mobile Conversation and MLOD – **Sven Schmeier, Yocoy**
- 12:45 – 14:00: Lunch und Networking Session

# LIDER Roadmapping Workshop: **Program**

- 14:00 – 14:15 Setting them up for failure – How customer expectations collide with economic realities of text analytics – **Patrick Bunk, CEO, uberMetrics**
- 14:15 – 14:30 Introduction to the German Wortschatz Project – **Dirk Goldhahn, University of Leipzig**
- 14:30 – 14:45 Towards the Single Digital Market – Processing Knowledge, Independent from Language – **Michael Wetzel, Coreon**
- 14:45 – 15:30 Discussion & Round-Up of Roadmapping
- 15:30 – 16:00 Coffee Break

# Discussion

## **I Resource creation and sharing (Heiko, Massimo, ...)**

- Share that at business-friendly licenses. What are current bottlenecks and how can they be overcome?
- Cooperative creation and joint exploitation of data
- Is there potential for a B2B data market?
- Middle-level quality resources are sufficient for most applications, but some applications require high-quality curated resources (e.g. lexica). How to address the quality vs. cost tradeoff? Should we invest in algorithmic support? Better? Human-machine collaboration workflows?
- Important to have resources for micro-domains that generalize and can be shared.
- Should EC invest in consortia / joint partnerships of relevant industrial that create high-quality domain-specific resources?
- Linking of lexicographic and terminological resources across languages, semi-automatically supported
- Important resources: terminologies, lexicographic resources, NL datasets, POS tagging / dependency parsing for UGC, sentiment datasets, in all relevant languages

## **II Standardization of APIs**

- to access/integrate data more easily
- to support cross-vendor comparison of technology
- support benchmarking and Q/A

# Discussion

## III (Open) Linked Data Publishing and Consumption

- Not always useful / no added value
- Emphasize more the linking, how to overcome knowledge silos
- Implement suitable access control on top of LD so that different levels of data access can co-exist
- Important of high-quality open data (e.g. government data)
- Which information is worth being published / released? **Small and reusable building blocks**
- Linked Multilingual Concepts / Multilingual Knowledge Infrastructure (MKS)
- Standardization (Different worlds mean inefficiency!): do we need more standards or less standards?

## IV) Multilingual Semantic Content Analysis and Search

- Robust and accurate analysis at different levels in frequently spoken languages
- Deeper analysis / more semantics
- Cross-lingual normalization (not only a translation problem)
- Semantic search (not for the string, but for the concept / the thing)



# Discussion

## **V) Big text & data Analytics**

- Big data analytic solutions to integrate fast and big data streams including analysis of unstructured data
- Integration of knowledge across languages, formats at large scale

## **VI Human factor (perception of technology and expectations)**

- How to cope with failure?
- What accuracy is needed (statisticians, humanities majors)?
- Do we need to educate end users better to tell us what kind of insights or analytics to they want?
- Evangelization required
- Privacy and legal aspects