Multilingual Linked Data Technologies for the Single Digital Market

Philipp Cimiano
(representing the LIDER Project)
LD4LT Teleconference
April 2nd, 2015

Digital Single Market: initiative by the European Commission to exploit existing growth opportunities and remove barriers to e-Commerce

The Single Market Act from 2012 summarizes 12 key action points:
1. Access to finance for small businesses
2. Citizens' mobility
3. Intellectual Property Rights
4. Consumer Protection
5. Services
6. Networks
7. Digital Single Market
8. Social Business
9. Taxation
10. Social Cohesion
11. Regulatory Environment for Businesses
12. Public Procurement

Essence: Digital Single Market is about Economic Growth, in particular stimulating cross-border commerce

Motivations:

- 72% of all the e-commerce transactions are of a domestic nature, with only 20% being cross-border within the EU and 9% with countries outside of the EU
- **Traders:**
  - slightly more than half of the EU retailers use e-commerce as a channel, but only about 20% engage in cross-border transactions
  - high compliance costs for traders to comply with different national regulations
  - 15% of the SMEs have applied restrictions in e-commerce with members from other states and 10% had to refuse service due to differences in consumer protection rules in the contract laws of other member states
- **Consumers:**
  - lack of comparability
  - concerns about data protection and security

Main Problem: foregone trade volume

- European Commission currently in the process of establishing the legislative framework in key areas (see "*Roadmap to the Single Digital Market*"):
  - **Common European Sales Law**
  - **Data Protection**
  - **E-Identification and Authentication**
  - **Alternative/Online Dispute Resolution**
  - **Collective Rights management**
  - **Orphan Works**
  - **Re-use of Public Sector Information**

Proposal	Size of Enabled Market	Savings in costs of compliance	Price Efficiency-related Consumer Surplus
Common European Sales Law	EUR 13-90 billion	EUR 1.5-3 billion	EUR 1.8 billion?
Data Protection	n.a.	EUR 4.3 billion	n.a
E-Identification & Authentication	> EUR 15-30 billion	n.a.	EUR 0.5-1.5 billion
ADR/ODR	n.a.	EUR 3—13 billion	n.a.
Collective Rights Management	> EUR 4-5 billion	EUR 5-10 mln	n.a.
Orphan Works	EUR 1-5 billion	EUR 1.5 billion	n.a.
Re-use of Public Sector Information	EUR 40 billion	n.a.	n.a.
Total	> 100 EUR billion	EUR 10-20 billion	> 5 billion

From „Roadmap to the Single Digital Market"

Current estimates: 340 EUR Growth in GDP overall
(Factsheet „Why we need a Digital Single Market", March 2015)

- Priorities for Juncker Commission (Commission Work Programme 2015)
  1. A New Boost for Jobs, Growth and Investment
  2. **A Connected Digital Single Market**
  3. A Resilient Energy Union with a Forward-Looking Climate Change Policy
  4. A Deeper and Fairer Internal Market with a Strengthened Industrial Base
  5. A Deeper and Fairer Economic and Monetary Union
  6. A Reasonable and Balanced Free Trade Agreement with the U.S:.
  7. An Area of Justice and Fundamental Rights Based on Mutual Trust
  8. Towards a New Policy on Migration
  9. A Stronger Global Actor
  10. A Union of Democratic Change

Mission of Vice President for the Single Digital Market:

– Modernizing copyright rules
– Reform of data protection rules
– Supporting citizens to benefit from the digital economy
– Supporting development of creative industries
– Supporting ways to make public administration more open and effective.
– Improve interaction with member states' administrations, for example on **reporting and exchanging information on systems needed to make the EU function**, and facilitating **cross-border interoperability between them**.

- Common European Sales Law (CESL):
  - How do we support SMEs in providing information about products, their features in 24 languages?
  - How do we ensure consistency and homogeneity of information so that cross-vendor comparison becomes possible?

- Online Dispute Resolution:
  - How do we support online dispute resolution in 24 languages effectively?
  - How do we support finding previous online dispute resolution similar cases, also across languages?

- Re-use of Public Sector Information:
  - 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.?
  - How do we support finding and accessing data intuitively?
  - How do we ensure interoperability of public services across borders? (see Ansip letter)
  - How do we ensure quality and trust in data?
  - How do we support clearance of licensing issue to ensure compliance in data use?

- Customer Engagement:
  - How do we support SMEs in interacting with customers in 24 languages and in listening to the voice of customers in 24 languages?

How can LD4LT contribute?

Philipp Cimiano

Common European Sales Law:
- **Localized product and feature catalogues** agreed upon vendors, no monolithic top-down approach, but agreed-upon vocabularies that are linked and for which information about how they are realized in different languages Good example: Good Relations Ontology
- **Small and highly reusable vocabularies** that are collaboratively created and localized by relevant stakeholders
- **Framework and platform to support collaboration** on Pan-European vocabulary creation and evolution
- **Pay-per-use model** to provide flexibility, also **non-monetary incentives** for contributing vocabulary elements, translations etc.
- Support bottom-up standardization of vocabularies, terminologies and catalogues
- Mechanisms for exploiting existing Linguistic Linked Data sources
- Ecosystem that involves relevant stakeholders: i) vendors, ii) consumers, iii) translators etc. to grow a shared vocabulary that is localized into all 24 languages
- Linked Data techniques for harmonizing and linking vocabularies in the background

Online Dispute Resolution:
- **Multilingual ontology-based generation and authoring**, to support issuing claims and responding to claims in different languages using a semi-controlled language approach (focusing on domain semantics)
- **Cross-lingual similarity** to find similar/related cases

Re-use of Public Sector Information:

- **Discovery:** technical solutions need to be implemented to actually discover relevant datasets. Linked Data technologies can significantly contribute to this on the basis of the extensive experience in developing data portals.

- **Exploitation:** The terms and conditions under which the data can be used and for which purpose need to be made explicit so that machines can automatically discover data with terms and conditions of use that can be exploited for a particular purpose in a compliant fashion.

- **Data Processing:** Infrastructure that supports filtering and processing data to generate valuable insights are required. This comes with the necessity to have appropriate e-infrastructure or cloud infrastructure where the data can be processed. An ecosystem of services that can be composed to realize complex processing chains on the data will need to be realized. Techniques for simplifying the creation of cross-border data value chains are needed.

- **Integration and Harmonization:** As relevant datasets will come in a variety of formats, with different semantics and be expressed in different natural languages major efforts to make sure that data comes in uniform formats and that semantics can be made explicit (e.g. by linking to existing vocabularies) and content in different languages can be linked. Otherwise, data will be open and accessible but not be exploitable.

Data Protection and Compliance:

- Attach provenance and licensing information to data, vocabularies etc. throughout the whole lifecycle
- Ensure compliance with user-defined constraints along life-cycle of data

Customer Engagement:

–   Support localization of messages to customers, homogeneous, localized and meaning-preserving communication across languages based on existing closed and open terminologies

–   Techniques for robustly analyzing the voice of the customer in 24 languages beyond polarity to recognize intention of customers

–   Create semantic user profiles by linking information across services and applications as a basis to personalize and contextualize interaction

–   Ensure compliance with data protection regulations and personal preferences of users

- Key Performance indicators;
  - How much open linked datasets of public sector information published?
  - How much datasets feature licensing information in machine-readable version?
  - How much lexical resources published as open linked data?
  - How many lexicalizations per language per vocabulary element?
  - How much links across multilingual datasets exist?

- Mission Letter by Jean-Claude Juncker to Andrus Ansip, Vice-President for the Digital Single Market, Nov. 2014
- "A Digital Agenda for Europe", Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions, Mai 2010
- "Single Market Act: Together for New Growth", Internal Market and Services
- "Roadmap to Digital Single Market", DG for Internal Policies, Note, 2012
- Factsheet: Why we need a Single Digital Market, March 2015, European Commission

- Thanks for your attention!