



## LIDER – I Linked Data come veicolo di contenuto multilingue e multimediale per le imprese europee

### La sfida

La crescita esponenziale in termini di volume, velocità di diffusione e varietà del Web richiedono la definizione di nuovi approcci per l'analisi dei contenuti. Questo è il prerequisito per affrontare i problemi di analisi su larga scala e nell'interpretazione di insiemi di dati eterogenei, provenienti da differenti supporti multimediali, in diverse lingue e da diverse organizzazioni.

Tuttavia, un elemento chiave mancante è la comprensione e l'interpretazione dei contenuti, sia a livello di testo non strutturato, sia a livello linguistico presente nel contesto di diversi flussi multimediali.

LIDER costruirà un percorso per la creazione di un nuovo ecosistema *Linked Open Data* (LOD) in modo da permettere l'accesso a risorse di tipo linguistico semanticamente interconnesse (rappresentazioni "*Linguistic Linked Data*" di corpora, dizionari, metadati lessicali e sintattici, ecc.) e su differenti supporti (metadati di immagini, video, ecc.). LIDER studierà come queste risorse possono contribuire alla realizzazione di una tecnologia per l'analisi di contenuti aziendali sul Web in diverse lingue, ivi compresi i contenuti multilingue distribuiti sui vari media.

LIDER contribuirà a costruire una comunità attorno al *Linguistic Linked Licensed Data* (3LD), includendo risorse linguistiche aperte e regolate da licenze nel formato linked data.

### LIDER: Principali risultati nel primo anno

➤ **Raggiungimento di comunità, costruzione di casi d'uso aziendali.** LIDER ha tenuto quattro *roadmapping workshops* al fine di raccogliere materiale per casi d'uso aziendali e requisiti per dati collegati linguistici. I workshop erano rivolti a quattro gruppi: la comunità dei dati, la comunità Web nel suo insieme, la comunità della localizzazione e la

comunità dell'analisi dei contenuti. Sono state dapprima raccolte opinioni attraverso il gruppo LD4LT ("*Linguistic Linked Data for Language Technology*"), un gruppo dedicato della comunità W3C, poi documentate nei report di workshop ed infine introdotte nella *roadmap* di LIDER. Inoltre, LIDER è stato presente ad un grande numero di eventi di ricerca e industriali, ha organizzato tutorial e tenuto presentazioni su dati collegati linguistici e sull'analisi di contenuto.

➤ **Linee guida e best practice.** LIDER si è avvalso di un gruppo W3C dedicato, chiamato BPMLOD ("*Best Practices for Multilingual Linked Open Data*"), al fine di raccogliere le linee guida per la creazione di nuove risorse linked data multilingua. Sono state inoltre effettuate conversioni di risorse linguistiche già esistenti in linked data. In primo luogo si è guardato soprattutto ai metadati di risorse linguistiche. LIDER ha collaborato costruttivamente con le comunità di META-NET e CLARIN per la conversione, al fine di assicurare un largo consenso e riutilizzo degli schemi di metadati già esistenti. In secondo luogo, LIDER ha descritto delle linee guida per vari tipi di risorse quali, ad esempio, dizionari bilingua e multilingua, e ha avviato la conversione di altre risorse, quali dati terminologici (TBX) o dati provenienti da collezioni testuali.

➤ **Architettura di riferimento e roadmapping.** LIDER ha creato una prima bozza di una architettura di riferimento per dati collegati linguistici per l'analisi di contenuto. L'architettura definisce da una parte un modello generale per la costruzione di servizi linguistici che operano a livello di *linked data* e dall'altra diversi schemi per la costruzione di applicazioni per l'analisi di contenuto. Essa include altresì una panoramica sugli strumenti e le iniziative rilevanti.

LIDER ha anche creato una prima versione di una



*roadmap* per dati collegati linguistici per l'analisi di contenuto. Nel concreto, vengono integrate informazioni provenienti da varie sorgenti: attività di LIDER quali *survey* o *roadmapping workshop*, report vari, agenzie di finanziamento e varie comunità di ricerca. La *roadmap* è formulata attorno a tre aree di applicazione chiave: coinvolgimento globale del cliente, settore pubblico / società civile, il ciclo di vita e la catena del valore di *linked data* linguistici. Per ciascuna di queste aree, la *roadmap* fornisce numerosi casi d'uso che includono linee temporali, previsioni e fattori rilevanti.

### **I prossimi passi**

LIDER validerà i risultati del primo anno attraverso un'interazione con l'intera comunità. Questa includerà la conversione di risorse linguistiche e dei loro metadati, l'architettura di riferimento e la *roadmap* su *linked data* linguistici per l'analisi di contenuto. Infine, LIDER si aprirà a nuovi gruppi di utenti non ancora raggiunti nel primo anno, specialmente nell'area dell'informazione del settore pubblico.

### **Coordinatore**

Contatto: Prof. Dr. Asunción Gómez-Pérez

Tel:+34 913367417

E-mail: [asun@fi.upm.es](mailto:asun@fi.upm.es)

Organizzazione: Universidad Politécnica de Madrid (UPM)

### **Partecipanti**

- Universidad Politécnica de Madrid (UPM)
- Trinity College Dublin (TCD)
- German Research Center for Artificial Intelligence (DFKI)
- INSIGHT National Center for Data Analytics, National University of Ireland, Galway
- Institute for Applied Informatics (InfAI), University of Leipzig
- University of Bielefeld
- Università degli studi di Roma “La Sapienza”
- W3C / Geie ERCIM

### **Organizzazione del progetto**

Durata: 24 mesi

Inizio: 1 novembre 2014

Fine: 31 ottobre 2015

Responsabile: Susan Fraser

Progetto 610782, CSA

### **Come contribuire e come essere informati su LIDER:**

- Sito Web del progetto: <http://www.lider-project.eu/>
- Portale della comunità: <http://www.multilingualweb.eu/>
- W3C Community Group su “Linked Data for Language Technology”: <http://www.w3.org/community/ld4lt/>
- Mailing list [lider-community@listas.fi.upm.es](mailto:lider-community@listas.fi.upm.es) per la raccolta di commenti anonimi e per ottenere informazioni
- Utilizzare <http://www.lider-project.eu/get-involved> per l'iscrizione.
- Nei social media tramite hashtag #LiderEU, per esempio su twitter: <https://twitter.com/multilingweb>