

2nd Call for Papers

## **3rd Workshop on Linked Data in Linguistics (LDL-2014):**

### **Multilingual Knowledge Resources and Natural Language Processing**

Tuesday, May 27, 2014, Reykjavik (Iceland), <http://ldl2014.org>

Collocated with the 9th Language Resources and Evaluation Conference (LREC-2014)

The explosion of information technology has led to a substantial growth in quantity, diversity and complexity of linguistic data accessible on the Web. The lack of interoperability between linguistic and language resources represents a major challenge that needs to be addressed, in particular, if information from different sources is to be combined, such as machine-readable lexicons, corpus data and terminology repositories. The Linked Data in Linguistics (LDL) workshop series provides a forum to discuss these types of resources, strategies to address issues of interoperability between them, protocols to distribute, access and integrate this information and technologies and infrastructures developed on this basis.

The goal of the workshop is twofold. First, we will assemble researchers from various fields of linguistics, natural language processing, knowledge management and information technology to present and discuss **principles, case studies, and best practices** for representing, publishing and linking mono- and multilingual linguistic and knowledge data collections, including corpora, grammars, dictionaries, wordnets, translation memories, domain specific ontologies etc. In this sense, we particularly invite contributions discussing the application of the **Linked Open Data paradigm** to linguistic data as it might provide an important step towards making linguistic data: i) easily and uniformly queryable, ii) interoperable and iii) sharable over the Web using open standards such as the HTTP protocol and the RDF data model [1]. The adaptation of some processes and best practices to **multilingual linguistic resources and knowledge bases** acquires also new relevance in this context. Some processes may need to be modified to accommodate the publication of resources that contain information in several languages. Also the linking process between linguistic resources in different languages poses important research questions, as well as the development and application of freely available knowledge bases and crowdsourcing to compensate the lack of publicly accessible language resources for various languages.

Secondly, we will provide researchers on natural language processing and semantic web technologies a platform to present case studies and best practices on the exploitation of linguistic resources exposed on the Web for **Natural Language Processing** applications, or other content-centered applications such as content analytics, knowledge extraction, etc. The availability of massive linked open knowledge resources raises the question how such data can be suitably employed to facilitate different NLP tasks and research questions. Following the tradition of earlier LDL workshops, we encourage contributions to the **Linguistic Linked Open Data (LLOD) cloud**<sup>1</sup> [2] and research on this basis. In particular, this pertains to contributions

---

<sup>1</sup> <http://linguistics.okfn.org/resources/lod/>

that demonstrate an added value resulting from the combination of linked datasets and ontologies as a source for semantic information with linguistic resources published according to as linked data principles. Another important question to be addressed in the workshop is how Natural Language Processing techniques can be employed to further facilitate the growth and enrichment of linguistic resources on the Web.

The **intended audience** includes linguists, NLP engineers and researchers from any field of computer science interested in the application of Semantic Web formalisms and related technologies to language data, empirically-working linguists and lexicographers interested in the representation, exchange and interlinking of knowledge resources, linguistic data and metadata, and developers of infrastructures for linguistic data and other researchers with an interest in both aspects.

## Background and History

This workshop brings together two community efforts, the Open Linguistics Working Group of the Open Knowledge Foundation (OWLKG), and the W3C Ontology-Lexica Community Group. LDL-2014 is also supported by a recently started EU Support Action: LIDER (Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe), which aims to provide an ecosystem for the establishment of linguistic linked open data, as well as media resources metadata, for a free and open exploitation of such resources in multilingual, cross-media content analytics across Europe.

The workshop is continuing a series of workshops on the application of the Linked Data paradigm to linguistic data that have been initiated and organized by the Open Linguistics Working Group: The *First Workshop on Linked Data in Linguistics (LDL-2012)* was conducted in March 2012 at the University of Frankfurt am Main/Germany, and collocated with the 34th Annual Meeting of the German Linguistics Society (DGfS-2012). The *Workshop on Multilingual Linked Open Data for Enterprises (MLODE-2012)* was conducted in September 2012 at the University of Leipzig/Germany, and collocated with the 3rd Conference on Software Agents and Services for Business, Research and E-Science (SABRE-2012). The *Second Workshop on Linked Data in Linguistics (LDL-2013)* was conducted in Sep 2013 at CNR in Pisa/Italy, and collocated with the 6th International Conference on the Generative Lexicon (GL2013).

## Linguistic Linked Data Challenge

There is a data challenge associated to the Linguistic Linked Data Workshop. In addition to regular workshop papers, we will accept dataset description of 4-6 pages describing new linguistic dataset published on the web as linked data. These linguistic datasets include, but are not limited to, lexica, terminologies, semantic networks, annotated and parallel corpora, multimodal resources, typological databases and linguistic metadata. The data challenge committee will review and evaluate data according to four dimensions, with prizes of up to €700, funded by the LIDER project, awarded to the highest scoring datasets.

The criteria for the Linguistic Linked Data Challenge include:

- Availability
  - Use of Linked Data and RDF.
  - Hosted on a publicly accessible server.
  - Use of an open license.
- Quality of Resource
  - Represents useful linguistically or NLP-relevant information.
  - Reuses relevant standards and models.
  - Contains complex, non-trivial information, e.g., multiple levels of annotation.
- Linking
  - Links to external resources.
  - Reuse of existing properties and categories.
- Impact/usefulness of the resource
  - Relevant and likely to be reused by many researchers in NLP and wider fields.
  - Uses linked data to improve the quality of and access to the resource.
- Originality
  - Represents a type of resource or a community currently under-represented in (L)LOD cloud activities
  - Facilitates novel and unforeseen applications or use cases (as described by the authors) enabled through Linked Data technology.

Details of the challenge are announced in separate Calls for Datasets, see <http://ldl2014.org/challenge.html> for up-to-date information.

### **Topics of interest**

We invite contributions related (but not limited) to the following topics:

1. Use cases and project proposals for the creation, publication or application of linguistic data collections that are linked with other resources
2. Modelling linguistic data and metadata with OWL and/or RDF
3. Ontologies for linguistic data and metadata collections as well as cross-lingual retrieval
4. Descriptions of data sets following Linked Data principles
5. Applications of such data, other ontologies or linked data from any subdiscipline of linguistics (may include work in progress or project descriptions)
6. Application and applicability of (Linguistic) Linked Open Data in NLP
7. NLP contributions to (Linguistic) Linked Open Data
8. Challenges of multilinguality and the use of LOD and collaboratively constructed open resources for knowledge extraction, machine translation and other NLP tasks.
9. Legal and social aspects of Linguistic Linked Open Data
10. Best practices for the publication and linking of multilingual knowledge resources

## Submission & Publication

We accept submission of both **long (up to 8 pages)** and **short papers (up to 4 pages)** to be presented as long or short oral presentation at the workshop to be submitted via <http://www.softconf.com/lrec2014/LDL/>. The papers of the workshop will be published as online proceedings. In addition, we aim for a journal special issue as post-conference proceedings in which a selected amount of papers presented at the workshop will be published.

When submitting a paper, authors will be asked to provide essential **information about resources** (in a broad sense, i.e. also technologies, standards, evaluation kits, etc.) that have been used for the work described in the paper or are a new result of your research. Moreover, ELRA encourages all LREC authors to **share the described LRs** (data, tools, services, etc.), to enable their reuse, replicability of experiments, including evaluation ones, etc. If this data (or parts of it) are provided as Linked Data, then please also consider to participate in the **Linguistic Linked Data Challenge** (<http://ldl2014.org/challenge.html>).

For contact data, stylesheets, up-to-date details on submission and the workshop itself, please consult our website: <http://ldl2014.org>.

## Timeline

Submission deadline:	Fri, Feb 7, 2014
Notification of acceptance:	Fri, Mar 7, 2014
Camera-ready paper:	Fri, Mar 21, 2014
Workshop:	Tue, May 27, 2014

Please note that due to synchronization with the main conference, no extensions can be given.

## Organizers

Christian Chiarcos	(Goethe-Universität Frankfurt am Main, Germany)
John McCrae	(Universität Bielefeld, Germany)
Elena Montiel	(Universidad Politécnica de Madrid, Spain)
Kiril Simov	(Bulgarian Academy of Sciences, Sofia, Bulgaria)
Antonio Branco	(University of Lisbon, Portugal)
Nicoletta Calzolari	(ILC-CNR, Italy)
Petya Osenova	(University of Sofia, Bulgaria),
Milena Slavcheva	(JRC-Brussels, Belgium)
Cristina Vertan	(University of Hamburg, Germany)

## **Program Committee**

Eneko Agirre (University of the Basque Country, Spain)  
Guadalupe Aguado (Universidad Politécnica de Madrid, Spain)  
Peter Bouda (Interdisciplinary Centre for Social and Language Documentation, Portugal)  
Steve Cassidy (Macquarie University, Australia)  
Damir Cavar (Eastern Michigan University)  
Eric Charton (Ecole Polytechnique de Montréal, Canada)  
Walter Daelemans (University of Antwerp, Belgium)  
Ernesto William De Luca (University of Applied Sciences Potsdam, Germany)  
Gerard de Melo (University of California at Berkeley)  
Thierry Declerck (Deutsches Forschungszentrum für Künstliche Intelligenz, Germany)  
Dongpo Deng (Institute of Information Sciences, Academia Sinica, Taiwan)  
Alexis Dimitriadis (Universiteit Utrecht, The Netherlands)  
Jeff Good (University at Buffalo)  
Asunción Gómez Pérez (Universidad Politécnica de Madrid, Spain)  
Jorge Gracia (Universidad Politécnica de Madrid, Spain)  
Walther v. Hahn (University of Hamburg, Germany)  
Eva Hajicova (Charles University Prague, Czech Republic)  
Harald Hammarström (Radboud Universiteit Nijmegen, The Netherlands)  
Yoshihiko Hayashi (Osaka University, Japan)  
Sebastian Hellmann (Universität Leipzig, Germany)  
Dominic Jones (Trinity College Dublin, Ireland)  
Lutz Maicher (Universität Leipzig, Germany)  
Pablo Mendes (Open Knowledge Foundation Deutschland, Germany)  
Steven Moran (Universität Zürich, Switzerland/Ludwig Maximilian University, Germany)  
Sebastian Nordhoff (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany)  
Antonio Pareja-Lora (Universidad Politécnica de Madrid, Spain)  
Maciej Piasecki (Wroclaw University of Technology, Poland)  
Adam Przepiorkowski (IPAN, Polish Academy of Sciences)  
Laurent Romary (INRIA, France)  
Felix Sasaki (Deutsches Forschungszentrum für Künstliche Intelligenz, Germany)  
Andrea Schalley (Griffith University, Australia)  
Marco Tadic (University of Zagreb, Croatia)  
Marieke van Erp (VU University Amsterdam, The Netherlands)  
Daniel Vila (Universidad Politécnica de Madrid, Spain)  
Menzo Windhouwer (Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands)

## References

- [1] Chiarcos, C., J. McCrae, P. Cimiano, C. Fellbaum (2013), Towards open data for linguistics: Lexical Linked Data. In: Oltramari et al. (eds.) *New Trends of Research in Ontologies and Lexical Resources*. Springer, Heidelberg.
- [2] Chiarcos, C., S. Nordhoff, S. Hellmann (2012, eds.), *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata*, Springer, Heidelberg.
- [3] Oltramari, A., P. Vossen, L. Qin, L., E. Hovy (2013, eds.), *New Trends of Research in Ontologies and Lexical Resources*, Springer, Heidelberg.