

# Linguistic Linked Open Data (LLOD) Introduction and Overview

*Christian Chiarcos<sup>1</sup>, Philipp Cimiano<sup>2</sup>, Thierry Declerck<sup>3</sup> & John P. McCrae<sup>2</sup>*

<sup>1</sup> Goethe-Universität Frankfurt am Main, Germany  
chiarcos@uni-frankfurt.de

<sup>2</sup> Universität Bielefeld, Germany  
{cimiano|jmccrae}@cit-ec.uni-bielefeld.de

<sup>3</sup> Deutsches Forschungszentrum für Künstliche Intelligenz, Germany  
declerck@dfki.de

## Abstract

The explosion of information technology has led to a substantial growth in quantity, diversity and complexity of linguistic data accessible over the internet. 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, like, say, machine-readable lexicons, corpus data and terminology repositories. For these types of resources, domain-specific standards have been proposed, yet, issues of interoperability between different types of resources persist, commonly accepted strategies to distribute, access and integrate their information have yet to be established, and technologies and infrastructures to address both aspects are still under development.

The goal of the 2nd Workshop on Linked Data in Linguistics (LDL-2013) has been to bring together researchers from various fields of linguistics, natural language processing, and information technology to present and discuss principles, case studies, and best practices for representing, publishing and linking linguistic data collections, including corpora, dictionaries, lexical networks, translation memories, thesauri, etc., infrastructures developed on that basis, their use of existing standards, and the publication and distribution policies that were adopted.

## Background: Integrating Information from Different Sources

In recent years, the limited interoperability between linguistic resources has been recognized as a major obstacle for data use and re-use within and across discipline boundaries. After half a century of computational linguistics [8], quantitative typology [12], empirical, corpus-based study of language [10], and computational lexicography [16], researchers in computational linguistics, natural language processing (NLP) or information technology, as well as in Digital Humanities, are confronted with an immense wealth of linguistic resources, that are not only growing in number, but also in their heterogeneity.

Interoperability involves two aspects [14]:

**Structural ('syntactic') interoperability:** Resources use comparable formalisms to represent and to access data (formats, protocols, query languages, etc.),

so that they can be accessed in a uniform way and that their information can be integrated with each other.

**Conceptual (‘semantic’) interoperability:** Resources share a common vocabulary, so that linguistic information from one resource can be resolved against information from another resource, e.g., grammatical descriptions can be linked to a terminology repository.

With the rise of the Semantic Web, new representation formalisms and novel technologies have become available, and different communities are becoming increasingly aware of the potential of these developments with respect to the challenges posed by the heterogeneity and multitude of linguistic resources available today. Many of these approaches follow the **Linked (Open) Data paradigm** [1] that postulates four rules for the publication and representation of Web resources: (1) Referred entities should be designated by using URIs, (2) these URIs should be resolvable over HTTP, (3) data should be represented by means of W3C standards (such as RDF), (4) and a resource should include links to other resources. These rules facilitate information integration, and thus, interoperability, in that they require that entities can be addressed in a globally unambiguous way (1), that they can be accessed (2) and interpreted (3), and that entities that are associated on a conceptual level are also physically associated with each other (4).

In the definition of Linked Data, the **Resource Description Framework (RDF)** receives special attention. RDF was designed to provide metadata about resources that are available either offline (e.g., books in a library) or online (e.g., eBooks in a store). RDF provides a generic data model based on labeled directed graphs, which can be serialized in different formats. Information is expressed in terms of *triples* - consisting of a *property* (relation, i.e., a labeled edge) that connects a *subject* (a resource, i.e., a labeled node) with its *object* (another resource, or a literal, e.g., a string). RDF resources (nodes)<sup>1</sup> are represented by *Uniform Resource Identifiers (URIs)*. They are thus globally unambiguous in the web of data. This allows resources hosted at different locations to refer to each other, and thereby to create a network of data collections whose elements are densely interwoven.

Several data base implementations for RDF data are available, and these can be accessed using **SPARQL** [17], a standardized query language for RDF data. SPARQL uses a triple notation similar to RDF, only that properties and RDF resources can be replaced by variables. SPARQL is inspired by SQL, variables can be introduced in a separate **SELECT** block, and constraints on these variables are expressed in a **WHERE** block in a triple notation. SPARQL does not only support running queries against individual RDF data bases that are accessible over HTTP (so-called ‘SPARQL end points’), but also, it allows us to combine information from multiple repositories (federation). RDF can thus not only be used to *establish* a network, or cloud, of data collections, but also, to *query* this network directly.

---

<sup>1</sup>The term ‘resource’ is ambiguous: *Linguistic* resources are structured collections of data which can be represented, for example, in RDF. In RDF, however, ‘resource’ is the conventional name of a node in the graph, because, historically, these nodes were meant to represent objects that are described by metadata. We use the terms ‘node’ or ‘concept’ whenever *RDF* resources are meant in ambiguous cases.

RDF has been applied for various purposes beyond its original field of application. In particular, it evolved into a generic format for knowledge representation. RDF was originally conceived as the building block of the **Semantic Web** and was then readily adopted by disciplines as different as biomedicine and bibliography. Due to its application across discipline boundaries, RDF is maintained by a large and active community of users and developers, and it comes with a rich infrastructure of APIs, tools, databases, query languages, and multiple sub-languages that have been developed to define data structures that are more specialized than the graphs represented by RDF. These sub-languages can be used to create *reserved vocabularies* and *structural constraints* for RDF data. For example, the Web Ontology Language (OWL) defines the datatypes necessary for the representation of ontologies as an extension of RDF, i.e., *classes* (concepts), *instances* (individuals) and *properties* (relations).

The concept of Linked Data is closely coupled with the idea of **openness** (otherwise, the linking is only partially reproducible), and in 2010, the original definition of Linked Open Data has been extended with a 5 star rating system for data on the Web.<sup>2</sup> The first star is achieved by publishing data on the Web (in any format) under an open license, and the second, third and fourth star require machine-readable data, a non-proprietary format, and using standards like RDF, respectively. The fifth star is achieved by linking the data to other people's data to provide context. If (linguistic) resources are published in accordance with these rules, it is possible to follow links between existing resources to find other, related data and exploit network effects.

## Linked Data: Benefits

Publishing Linked Data allows resources to be globally and uniquely identified such that they can be retrieved through standard Web protocols. Moreover, resources can be easily linked to one another in a uniform fashion and thus become structurally interoperable. Linking to central terminology repositories facilitates conceptual interoperability. Beyond this, Chiarcos et al. [7] identified the following main benefits of Linked Linguistic Data: (a) linking through URIs, (b) federation, (c) dynamic linking between resources, and (d) the availability of a rich ecosystem of formats and technologies.

### Linking through URIs

Linked Data requires that every resource is identified by a Uniform Resource Identifier (URI) that figures both as a global identifier and as a Web address – i.e., a description of the resource is available if you request it from its URI on the Web. However, RDF allows for a standard description of such resources on the Web and hence for automatic processing of these resources. It is not necessarily the case that the data must be solely available as RDF, as the HTTP protocol supports *content negotiation*: as one example, the RDF data under [http://de.dbpedia.org/data/Linked\\_Open\\_Data.rdf](http://de.dbpedia.org/data/Linked_Open_Data.rdf) can be rendered in human-readable HTML, see [http://de.dbpedia.org/page/Linked\\_Open\\_Data](http://de.dbpedia.org/page/Linked_Open_Data).

---

<sup>2</sup><http://www.w3.org/DesignIssues/LinkedData.html>, paragraph 'Is your Linked Open Data 5 Star?'

## Information Integration at Query Runtime (Federation)

As resources can be uniquely identified and easily referenced from any other resource on the Web through URIs, the connections between these resources can be navigated even during query runtime. In effect, this allows the creation of a linked web of data similar to the effect of hyperlinks in the HTML Web. Moreover, it is possible to use existing Semantic Web methods such as Semantic PingBack [18] to be informed of new incoming links to your resource. Along with HTTP-accessible repositories and resolvable URIs, it is possible to combine information from physically separated repositories in a single query at runtime. Information from different resources in the cloud can then be integrated freely.

## Dynamic Import

If cross-references between linguistic resources are represented by resolvable URIs instead of system-defined ID references or static copies of parts from another resource, it is not only possible to resolve them at runtime, but also to have access to the most recent version of a resource. For community-maintained terminology repositories like the ISO TC37/SC4 Data Category Registry [20, 19, ISocat], for example, new categories, definitions or examples can be introduced occasionally, and this information is available immediately to anyone whose resources refer to ISocat URIs.

## Ecosystem

RDF as a data exchange framework is maintained by an interdisciplinary, large and active community, and it comes with a developed infrastructure that provides APIs, database implementations, technical support and validators for various RDF-based languages, e.g., reasoners for OWL. For developers of linguistic resources, this ecosystem can provide technological support or off-the-shelf implementations for common problems, e.g., the development of a database that is capable of supporting flexible, graph-based data structures as necessary for multi-layer corpora [15].

Beyond this, another advantage warrants a mention: The distributed approach of the Linked Data paradigm facilitates the distributed development of a web of resources and collaboration between researchers that provide and use this data and that employ a shared set of technologies. One consequence is the emergence of interdisciplinary efforts to create large and interconnected sets of resources in linguistics and beyond. LDL-2013 aims to provide a forum to discuss and to facilitate such on-going developments.

## LLOD: Building the Cloud

Recent years have seen not only a number of approaches to provide linguistic data as Linked Data, but also the emergence of larger initiatives that aim at interconnecting these resources, culminating in the creation of a Linguistic Linked Open Data (LLOD) cloud, i.e., a Linked Open Data (sub-)cloud of linguistic resources.

LDL-2013 is organized in the context of two recent community efforts, the Open Linguistics Working Group (OWLG), and the W3C Ontology-Lexica Community Group (OntoLex). The Open Linguistics Working Group has spearheaded the creation of new data and the republishing of existing linguistic resources as part of the emerging Linguistic Linked Open Data (LLOD) cloud. Similarly, the W3C Ontology-Lexica Community Group is seeking to develop standard models for representing and publishing (ontology-) lexica and other lexical resources as RDF.

## The LLOD Cloud

Aside from benefits arising from the actual *linking* of linguistic resources, various linguistic resources from various fields have been provided in RDF and related standards in the last decade.

In particular, this is the case for **lexical resources** (Fig. 1, LEXICON), e.g., WordNet [11], which represent a cornerstone of the Semantic Web and which are firmly integrated in the Linked Open Data (LOD) cloud. Other types of linguistic resources with less relevance for AI and Knowledge Representation, however, have been absent from the LOD cloud.

The Linked Data paradigm also facilitates the management of information about language (Fig. 1, LANGUAGE\_DESCRIPTION), i.e., linguistic terminology and linguistic databases. **Terminology repositories** serve an important role to establish conceptual interoperability between language resources. If resource-specific annotations or abbreviations are expanded into references to repositories of linguistic terminology and/or metadata categories, linguistic annotations, grammatical features and metadata specifications become more easily comparable. Important repositories developed by different communities include GOLD [9] and ISOcat [20, 19], yet, only recently these terminology repositories were put in relation with each other using Linked Data principles and with linguistic resources, e.g., within the OLiA architecture [5]. **Linguistic databases** are a particularly heterogeneous group of linguistic resources; they contain complex and manifold types of information, e.g., feature structures that represent typologically relevant phenomena, along with examples for their illustration and annotations (glosses) and translations applied to these examples (structurally comparable to corpus data), or word lists (structurally comparable to lexical-semantic resources). RDF as a generic representation formalism is thus particularly appealing for this class of resources.

Finally, for **linguistic corpora** (Fig. 1, CORPORA), the potential of the Linked Data paradigm for modeling, processing and querying of corpora is immense, and RDF conversions of semantically annotated corpora have been proposed early [3]. RDF provides a graph-based data model as required for the interoperable representation of arbitrary kinds of annotation [2, 15], and this flexibility makes it a promising candidate for a general means of representation for corpora with complex and heterogeneous annotations. RDF does not only establish interoperability between annotations within a corpus, but also between corpora and other linguistic resources [4]. In comparison to other types of linguistic resources, corpora are currently underrepresented in the LLOD cloud, but the development of schemes for corpora and/or NLP annotations represents an active line of research [6, 13] also addressed in the workshop.

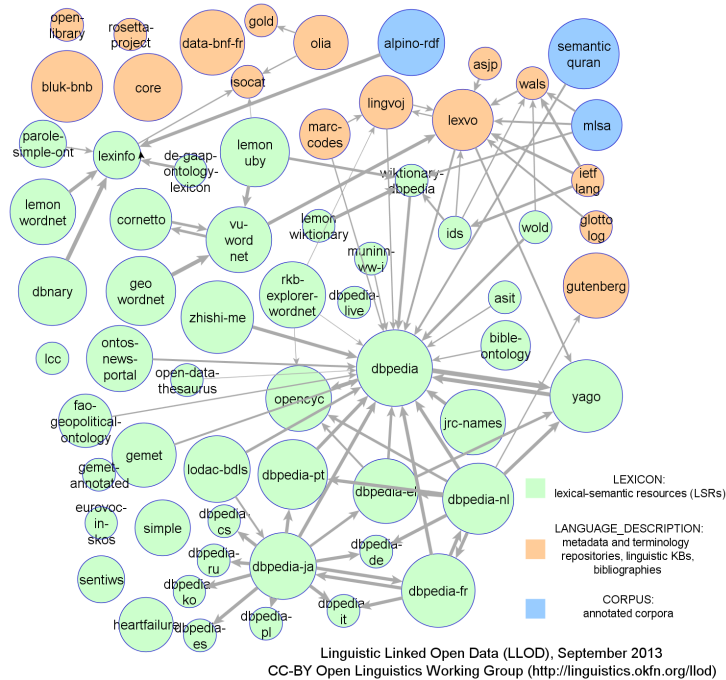


Figure 1: Linguistic Linked Open Data cloud as of September 2013.

Only recently, the efforts to apply RDF to linguistic resources of different types have begun to converge towards an actual Linked Open Data (sub-) cloud of linguistic resources, the Linguistic Linked Open Data (LLOD) cloud.

## Community Efforts

The LLOD cloud is a result of a coordinated effort of the **Open Linguistics Working Group (OWLG)**,<sup>3</sup> a network open to anyone interested in linguistic resources and/or the publication of these under an open license. The OWLG is a working group of the Open Knowledge Foundation (OKFN),<sup>4</sup> a community-based non-profit organization promoting open knowledge (i.e., data and content that is free to use, re-use and to be distributed without restriction).

Since its formation in 2010, the Open Linguistics Working Group has grown steadily. One of our primary goals is to attain openness in linguistics through:

1. Promoting the idea of open linguistic resources,
2. Developing the means for the representation of open data, and
3. Encouraging the exchange of ideas across different disciplines.

<sup>3</sup><http://linguistics.okfn.org>

<sup>4</sup><http://okfn.org/>

The OWLG represents an open forum for interested individuals to address these and related issues. At the time of writing, the group consists of about 100 people from 20 different countries. Our group is relatively small, but continuously growing and sufficiently heterogeneous. It includes people from library science, typology, historical linguistics, cognitive science, computational linguistics, and information technology; the ground for fruitful interdisciplinary discussions has been laid out. One concrete result emerging out of collaborations between a large number of OWLG members is the LLOD cloud as already sketched above.

The emergence of the LLOD cloud out of a set of isolated resources was accompanied and facilitated by a series of **workshops and publications** organized under the umbrella of the OWLG, including the Open Linguistics track at the Open Knowledge Conference (OKCon-2010, July 2010, Berlin, Germany), the First Workshop on Linked Data in Linguistics (LDL-2012, March 2012, Frankfurt am Main, Germany), the Workshop on Multilingual Linked Open Data for Enterprises (MLODE-2012, September 2012, Leipzig, Germany), the Linked Data for Linguistic Typology track at ALT-2012 (September 2013, Leipzig, Germany). Plans to create a LLOD cloud were first publicly announced at LDL-2012, and subsequently, a first instance of the LLOD materialized as a result of the MLODE-2012 workshop, its accompanying hackathon and the data postproceedings that will appear as a special issue of the Semantic Web Journal (SWJ). The Second Workshop on Linked Data in Linguistics (LDL-2013) continues this series of workshops. In order to further contribute to the integration of the field, it is organized as a joint event of the OWLG and the W3C Ontology-Lexica Community Group.

The **Ontology-Lexica Community (OntoLex) Group**<sup>5</sup> was founded in September 2011 as a W3C Community and Business Group. It aims to produce specifications for a lexicon-ontology model that can be used to provide rich linguistic grounding for domain ontologies. Rich linguistic grounding includes the representation of morphological, syntactic properties of lexical entries as well as the syntax-semantics interface, i.e., the meaning of these lexical entries with respect to the ontology in question. An important issue herein will be to clarify how extant lexical and language resources can be leveraged and reused for this purpose. As a byproduct of this work on specifying a lexicon-ontology model, it is hoped that such a model can become the basis for a web of lexical linked data: a network of lexical and terminological resources that are linked according to the Linked Data Principles forming a large network of lexico-syntactic knowledge.

The OntoLex W3C Community Group has been working for more than a year on realizing a proposal for a standard ontology lexicon model, currently discussed under the the designation *lemon*. As the core specification of the model is almost complete, the group started to develop of additional modules for specific tasks and use cases, and some of these are presented at LDL-2013.

## LDL-2013: The 2nd Workshop on Linked Data in Linguistics

The goal of the 2nd Workshop on Linked Data in Linguistics (LDL-2013) has been to bring together researchers from various fields of linguistics, NLP,

---

<sup>5</sup><http://www.w3.org/community/ontolex>

and information technology to present and discuss principles, case studies, and best practices for representing, publishing and linking linguistic data collections, including corpora, dictionaries, lexical networks, translation memories, thesauri, etc., infrastructures developed on that basis, their use of existing standards, and the publication and distribution policies that were adopted.

For the 2nd edition of the workshop on Linked Data in Linguistics, we invited 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. Recent research in this direction has led to the emergence of a Linked Open Data cloud of linguistic resources, the Linguistic Linked Open Data (LLOD) cloud, where Linked Data principles have been applied to language resources, allowing them to be published and linked in a principled way. Although not restricted to lexical resources, these play a particularly prominent role in this context. The topics of interest mentioned in the call for papers were the following ones:

1. Use cases for creation, maintenance and publication 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
4. Applications of such data, other ontologies or linked data from any sub-discipline of linguistics
5. Descriptions of data sets, ideally following Linked Data principles
6. Legal and social aspects of Linguistic Linked Open Data

In response to our call for papers we received 17 submissions which were all reviewed by at least two members of our program committee. On the basis of these reviews, we decided to accept 8 papers as full papers and 2 as short papers, giving an overall acceptance rate of around 50%.

LDL-2013 was collocated with the 6th International Conference on Generative Approaches to the Lexicon (GL2013): Generative Lexicon and Distributional Semantics, and hence, **lexical-semantic resources** represent a particularly important group of resources at the current edition of the workshop.

The contributions by Koide and Takeda and Bartolini et al. describe the conversion of the Japanese and Italian WordNet and related resources as well as their linking to (L)LOD resources such as the DBpedia.

Buitelaar et al. describe the specification and use of a model for the interoperable representation of language resources for sentiment analysis. The model is based directly on *lemon*, and in the EuroSentiment project it will be used to represent language resources for sentiment analysis such as WordNet Affect in an interoperable way.

Similarly, Moran and Brümmer employ *lemon* for the modeling of dictionary and wordlist data made available by a project on quantitative historical linguistics. Using Linked Data principles, more than fifty disparate lexicons and dictionaries were combined into a single dataset, which then provides researchers



with a translation graph allowing users to query across the underlying lexicons and dictionaries to extract semantically-aligned wordlists.

An extension of *lemon* is developed by Fiorelli et al. who present LIME (Linguistic Metadata), a new vocabulary aiming at completing *lemon* with specifications for linguistic metadata. In many usage scenarios currently developed as extensions of *lemon* (e.g. ontology alignment, localization etc...), the discovery and exploitation of linguistically grounded datasets may benefit from reassuming information about their linguistic expressivity. While the VoID vocabulary covers the need for general metadata about linked datasets, specifically linguistic information demands a dedicated extension.

Finally, Bonial et al. describe SemLink, a comprehensive resource for NLP that maps and unifies several high quality lexical resources: PropBank, VerbNet, FrameNet, and OntoNotes sense groupings. Each of these resources was created for different purposes, and therefore each carries unique strengths and limitations. SemLink allows users to leverage the strengths of each resource and provides the groundwork for incorporating these lexical resources effectively. Although SemLink is not immediately based on the application of the Linked Data paradigm, it represents an important contribution to the LLOD cloud, as it provides links between classical resources for word-level semantics (e.g., WordNet) long established in the (L)LOD cloud, and frame-semantic resources. In this function, an earlier instantiation of SemLink represents a fundamental component of the lemonUby data set shown in Fig. 1.

An approach to model of **language description** data as Linked Data is presented by Littauer et al. who feed spreadsheet data about a group of endangered languages and where they are spoken in West Africa into an RDF triple store. They use RDF tools to organize and visualize these data on a world map, accessible through a web browser. The functionality they develop allows researchers to see where these languages are spoken and to query the language data, thereby providing a powerful tool for linguists studying the genealogical relatedness of the Dogon languages.

A different type of information about language is addressed by Hayashi who describes the modeling of psycholinguistic semantic feature norms. Semantic feature norms, originally utilized in the field of psycholinguistics as a tool for studying human semantic representation and computation, have recently attracted some NLP/IR researchers who wish to improve their task performances. Currently available semantic feature norms are, however, rarely well structured, making them difficult to integrate with existing resources of various types. This paper provides a case study, it extracts a tentative set of semantic feature norms that are psycholinguistically considerable, and draws a technical map to formalize them by observing the Linked Data paradigm.

LDL-2013 features three contributions addressing **corpora** that we identified above as being underrepresented in the LLOD cloud: Menke et al. describe a framework for releasing multimodal corpora as Linked Data, and experiences in releasing a multimodal corpus based on an online chat game on that basis. Heuss presents an experiment in translating excerpts of a natural language story into a formal RDF structure, so that it is accessible by machines on a word or concept level. Finally, Pareja-Lora et al. describe the first steps taken to transform a set of linguistic resources from the Data Transcription and Analysis Tool's (DTA) metadata and data into an open and interoperable language resource.

## Acknowledgements

We would like to express our gratitude to the organizers of the GL2013 for hosting our workshop and support with respect to local organization. Further, we thank the OWLG and its members for active contributions to the LLOD cloud, to the workshop and beyond. In particular, we have to thank the contributors and the program committee for their invaluable work and engagement.

## Bibliography

- [1] T. Berners-Lee. Design issues: Linked data. URL <http://www.w3.org/DesignIssues/LinkedData.html> (July 31, 2012), 2006.
- [2] S. Bird and M. Liberman. A formal framework for linguistic annotation. *Speech Communication*, 33(1-2):23–60, 2001.
- [3] A. Burchardt, S. Padó, D. Spohr, A. Frank, and U. Heid. Formalising multi-layer corpora in OWL/DL – lexicon modelling, querying and consistency control. In *3rd International Joint Conference on NLP (IJCNLP 2008)*, Hyderabad, India, 2008.
- [4] C. Chiarcos. Interoperability of corpora and annotations. In C. Chiarcos, S. Nordhoff, and S. Hellmann, editors, *Linked Data in Linguistics*, pages 161–179. Springer, Heidelberg, 2012.
- [5] C. Chiarcos. Ontologies of linguistic annotation: Survey and perspectives. In *8th International Conference on Language Resources and Evaluation (LREC-2012)*, pages 303–310, Istanbul, Turkey, May 2012.
- [6] C. Chiarcos. POWLA: Modeling linguistic corpora in OWL/DL. In *9th Extended Semantic Web Conference (ESWC-2012)*, pages 225–239, Heraklion, Crete, May 2012.
- [7] C. Chiarcos, J. McCrae, P. Cimiano, and C. Fellbaum. Towards open data for linguistics: Linguistic linked data. In A. Oltramari, Lu-Qin, P. Vossen, and E. Hovy, editors, *New Trends of Research in Ontologies and Lexical Resources*. Springer, Heidelberg, to appear.
- [8] L. Dostert. The Georgetown-IBM experiment. In W. Locke and A. Booth, editors, *Machine Translation of Languages*, pages 124–135. John Wiley & Sons, New York, 1955.
- [9] S. Farrar and T. Langendoen. Markup and the GOLD ontology. In *EMELD Workshop on Digitizing and Annotating Text and Field Recordings*. Michigan State University, July 2003.
- [10] W. N. Francis and H. Kucera. Brown Corpus manual. Technical report, Brown University, Providence, Rhode Island, 1964. revised edition 1979.

- [11] A. Gangemi, R. Navigli, and P. Velardi. The OntoWordNet project: Extension and axiomatization of conceptual relations in WordNet. In R. Meersman and Z. Tari, editors, *Proceedings of On the Move to Meaningful Internet Systems (OTM2003)*, pages 820–838, Catania, Italy, November 2003.
- [12] J. Greenberg. A quantitative approach to the morphological typology of languages. *International Journal of American Linguistics*, 26:178–194, 1960.
- [13] S. Hellmann, J. Lehmann, and S. Auer. Linked-data aware URI schemes for referencing text fragments. In *18th International Conference on Knowledge Engineering and Knowledge Management (EKAW-2012)*, Galway, Ireland, 2012.
- [14] N. Ide and J. Pustejovsky. What does interoperability mean, anyway? Toward an operational definition of interoperability. In *Second International Conference on Global Interoperability for Language Resources (ICGL 2010)*, Hong Kong, China, 2010.
- [15] N. Ide and K. Suderman. GRAF: A graph-based format for linguistic annotations. In *1st Linguistic Annotation Workshop (LAW 2007)*, pages 1–8, Prague, Czech Republic, 2007.
- [16] W. Morris, editor. *The American Heritage Dictionary of the English Language*. Houghton Mifflin, New York, 1969.
- [17] E. Prud’Hommeaux and A. Seaborne. SPARQL query language for RDF. *W3C working draft*, 4(January), 2008.
- [18] S. Tramp, P. Frischmuth, N. Arndt, T. Ermilov, and S. Auer. Weaving a distributed, semantic social network for mobile users. In *8th Extended Semantic Web Conference (ESWC-2011)*, pages 200–214, Heraklion, Crete, 2011.
- [19] M. Windhouwer and S.E. Wright. Linking to linguistic data categories in ISocat. In C. Chiarcos, S. Nordhoff, and S. Hellmann, editors, *Linked Data in Linguistics*, pages 99–107. Springer, Heidelberg, 2012.
- [20] S.E. Wright. A global data category registry for interoperable language resources. In *Proceedings of the Fourth Language Resources and Evaluation Conference (LREC 2004)*, pages 123–126, Lisboa, Portugal, May 2004.



## Organizing Committee

Christian Chiarcos	Johan Wolfgang Goethe Universität Frankfurt
Philipp Cimiano	Universität Bielefeld
Thierry Declerck	Deutsches Forschungszentrum für Künstliche Intelligenz
John Philip McCrae	Universität Bielefeld

## Program Committee

Guadalupe Aguado	Universidad Politécnica de Madrid, Spain
Maria Blume	Cornell University, USA
Peter Bouda	Interdisciplinary Centre for Social and Language Documentation, Portugal
Steve Cassidy	Macquarie University, Australia
Damir Cavar	Eastern Michigan University, USA
Michael Cysouw	Ludwig-Maximilian-Universität München, Germany
Ernesto William De Luca	University of Applied Sciences Potsdam, Germany
Gerard de Melo	University of California at Berkeley, USA
Dongpo Deng	Institute of Information Sciences, Academia Sinica, Taiwan
Alexis Dimitriadis	Universiteit Utrecht, The Netherlands
Judith Eckle-Kohler	Technische Universität Darmstadt, Germany
Jeff Good	University at Buffalo, USA
Jorge Gracia	Universidad Politécnica de Madrid, Spain
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, German
Elena Monsiel-Ponsoda	Universidad Politécnica de Madrid, Spain
Steven Moran	Universität Zürich, Switzerland/Ludwig Maximilian University, Germany
Sebastian Nordhoff	Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
Antonio Pareja-Lora	Universidad Complutense Madrid, Spain
Felix Sasaki	Deutsches Forschungszentrum für Künstliche Intelligenz, Germany
Andrea Schalley	Griffith University, Australia
Marieke van Erp	VU University Amsterdam, The Netherlands
Menzo Windhouwer	Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
Alena Witzlack-Makarevich	Universität Zürich, Switzerland