



LREC 2022 Workshop  
Language Resources and Evaluation Conference  
20-25 June 2022

**Proceedings of the 8th Workshop on Linked Data in  
Linguistics  
(LDL-2022)**

**PROCEEDINGS**

Editors:  
Thierry Declerck, John P. McCrae, Elena Montiel, Christian  
Chiarcos, Maxim Ionov

# **Proceedings of the LREC 2022 workshop on on Linked Data in Linguistics: Revisiting a Decade of Linguistic Linked Open Data (LDL 2022)**

Edited by:

Thierry Declerck, John P. McCrae, Elena Montiel, Christian Chiarcos, Maxim Ionov

The 10th anniversary edition of the LDL can count on the support of the COST Action CA18209 “NexusLinguarum: European Network for Web-centered Linguistic Data Science”, as well as two Horizon 2020 projects. Firstly, the Prêt-à-LLOD project (grant agreement no. 825182), which is making linguistic linked open data ready-to-use, and, secondly, the ELEXIS project on building a lexicographic infrastructure (grant agreement No 731015).

**ISBN: 979-10-95546-93-1**

**EAN: 9791095546931**

**For more information:**

European Language Resources Association (ELRA)

9 rue des Cordelières

75013, Paris

France

<http://www.elra.info>

Email: [lrec@elda.org](mailto:lrec@elda.org)



© European Language Resources Association (ELRA)

These workshop proceedings are licensed under a Creative Commons  
Attribution-NonCommercial 4.0 International License

## Preface

This volume documents the Proceedings of the 8th Workshop on Workshop on Linked Data in Linguistics, held on Friday the 24th of June as part of the LREC 2022 conference (International Conference on Language Resources and Evaluation).

Since its inception, the workshop series on Linked Data in Linguistics (LDL) established itself as the main venue for discussing how Linked Open Data (LOD) and semantic web technologies can be used for processing, analysing, publishing, and managing linguistic data. This includes the fields of natural language processing (NLP), language resources (LRs), lexicography and digital humanities (DH), and has been leading to the development of linguistic data science as a new area of study. The LDL workshop series has contributed greatly to the development of the Linguistic Linked Open Data (LLOD) cloud and the development of best practices for publishing and accessing language resources and providing language technology services on the web. Most notably, this includes community standards such as the NLP Interchange Format (NIF), the OntoLex-Lemon model of the W3C Community Group Ontology-Lexica, and numerous domain-specific adaptations and extensions that these models have had an influence on.

In addition, there are an increasing number of national, European, and international research projects that build on LLOD technology. These will contribute to its further development and will help ensure the success of this workshop and a high attendance rate. The 10th anniversary edition of the LDL can count on the support of the COST action “NexusLinguarum: European Network for Web-centered Linguistic Data Science”, as well as two Horizon 2020 projects. Firstly, the Prêt-à-LLOD project, which is making linguistic linked open data ready-to-use, and, secondly, the ELEXIS project on building a lexicographic infrastructure.



## **Organizers**

Thierry Declerck – DFKI GmbH, Saarland Informatics Campus

John P. McCrae – National University of Ireland Galway

Elena Montiel – Universidad Politécnica de Madrid

Christian Chiarcos – Goethe University Frankfurt

Maxim Ionov – Goethe University Frankfurt

## **Program Committee:**

Sina Ahmadi (NUI Galway, Ireland)

Paul Buitelaar (Insight, Ireland)

Sara Carvalho (University of Aveiro, Portugal)

Nicoletta Calzolari (ILC-CNR, Italy)

Milan Dojchinovski (Czech Technical University in Prague, Czech Republic)

Agata Filipowska (Poznan University of Economics, Poland)

Francesca Frontini (ILC-CNR, Italy)

Jeff Good (University at Buffalo, USA)

Yoshihiko Hayashi (Waseda University, Tokyo, Japan)

Eero Hyvönen (Aalto University, Finland)

Fahad Khan (ILC-CNR, Italy)

Chaya Liebeskind (Jerusalem College of Technology, Israel)

Gerard de Melo (HPI/Universität Potsdam, Germany)

Steve Moran (University of Neuchâtel, Switzerland)

Verginica Mititelu (Research Institute for Artificial Intelligence of the Romanian Academy, Romania)

Roberto Navigli (“La Sapienza” Università di Roma, Italy)

Sebastian Nordhoff (Language Science Press, Berlin, Germany)

Petya Osenova (IICT-BAS, Bulgaria)

Ana Ostroški Anić (Institute of Croatian Language and Linguistics, Croatia)

Antonio Pareja-Lora (Universidad de Alcalá de Henares)

Felix Sasaki (SAP, Germany)

Andrea Schalley (Karlstad University, Sweden)

Gilles Sérasset (University Grenoble Alpes, France)

Milena Slavcheva (IICT-BAS, Bulgaria)

Ranka Stankovic (University of Belgrade, Serbia)

Armando Stellato (University of Rome, Tor Vergata, Italy)

Andrius Utka (Vytautas Magnus University, Lithuania)



## Table of Contents

<i>The Annohub Web Portal</i>	
Frank Abromeit .....	1
<i>From ELTeC Text Collection Metadata and Named Entities to Linked-data (and Back)</i>	
Milica Ikonić Nešić, Ranka Stanković, Christof Schöch and Mihailo Skoric .....	7
<i>IMTVault: Extracting and Enriching Low-resource Language Interlinear Glossed Text from Grammatical Descriptions and Typological Survey Articles</i>	
Sebastian Nordhoff and Thomas Krämer .....	17
<i>Linking the LASLA Corpus in the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin</i>	
Margherita Fantoli, Marco Passarotti, Francesco Mambrini, Giovanni Moretti and Paolo Ruffolo	26
<i>Use Case: Romanian Language Resources in the LOD Paradigm</i>	
Verginica Barbu Mititelu, Elena Irimia, Vasile Pais, Andrei-Marius Avram and Maria Mitrofan .	35
<i>Fuzzy Lemon: Making Lexical Semantic Relations More Juicy</i>	
Fernando Bobillo, Julia Bosque-Gil, Jorge Gracia and Marta Lanau-Coronas .....	45
<i>A Cheap and Dirty Cross-Lingual Linking Service in the Cloud</i>	
Christian Chiarcos and Gilles Sérasset .....	52
<i>Spicy Salmon: Converting between 50+ Annotation Formats with Fintan, Pepper, Salt and Powla</i>	
Christian Fäth and Christian Chiarcos .....	61
<i>A Survey of Guidelines and Best Practices for the Generation, Interlinking, Publication, and Validation of Linguistic Linked Data</i>	
Fahad Khan, Christian Chiarcos, Thierry Declerck, Maria Pia Di Buono, Milan Dojchinovski, Jorge Gracia, Giedre Valunaite Oleskeviciene and Daniela Gifu .....	69
<i>Computational Morphology with OntoLex-Morph</i>	
Christian Chiarcos, Katerina Gkirtzou, Fahad Khan, Penny Labropoulou, Marco Passarotti and Matteo Pellegrini .....	78

# Conference Program

Friday June 24, 2022

*The Annohub Web Portal*

Frank Abromeit

*From ELTeC Text Collection Metadata and Named Entities to Linked-data (and Back)*

Milica Ikonić Nešić, Ranka Stanković, Christof Schöch and Mihailo Skoric

*IMTVault: Extracting and Enriching Low-resource Language Interlinear Glossed Text from Grammatical Descriptions and Typological Survey Articles*

Sebastian Nordhoff and Thomas Krämer

*Linking the LASLA Corpus in the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin*

Margherita Fantoli, Marco Passarotti, Francesco Mambrini, Giovanni Moretti and Paolo Ruffolo

*Use Case: Romanian Language Resources in the LOD Paradigm*

Verginica Barbu Mititelu, Elena Irimia, Vasile Pais, Andrei-Marius Avram and Maria Mitrofan

*Fuzzy Lemon: Making Lexical Semantic Relations More Juicy*

Fernando Bobillo, Julia Bosque-Gil, Jorge Gracia and Marta Lanau-Coronas

*A Cheap and Dirty Cross-Lingual Linking Service in the Cloud*

Christian Chiarcos and Gilles Sérasset

*Spicy Salmon: Converting between 50+ Annotation Formats with Fintan, Pepper, Salt and Powla*

Christian Fäth and Christian Chiarcos

*A Survey of Guidelines and Best Practices for the Generation, Interlinking, Publication, and Validation of Linguistic Linked Data*

Fahad Khan, Christian Chiarcos, Thierry Declerck, Maria Pia Di Buono, Milan Dojchinovski, Jorge Gracia, Giedre Valunaite Oleskeviciene and Daniela Gifu

*Computational Morphology with OntoLex-Morph*

Christian Chiarcos, Katerina Gkirtzou, Fahad Khan, Penny Labropoulou, Marco Passarotti and Matteo Pellegrini