

LexBib / LexDo
May 2020
- KIM Workshop Lightning Talk -



David Lindemann david@lexbib.org

LexBib Zotero Collection

- Public Zotero Group
- Access: <http://lexbib.org/zotero>
- All metadata present, searchable
- API service for local citation managers
- Collaborative maintenance possible
- Group admins have access to full text attachments
- Public release planned at EURALEX 2020

The screenshot shows the Zotero web interface. On the left is a sidebar with a tree view of group libraries, including 'LexBib' and 'EURALEX'. The main area displays a list of items with columns for Title, Creator, and Date. The selected item is 'A Call for a Corpus-Based Sign Language Dictionary: An Overview of Croatian Sign Language Lexicography in the Early 21st Century' by Majetić and Bago (2018). Below the list, a detailed view of this item is shown, including its item type (Conference Paper), title, authors (Majetić, Klara and Bago, Petra), editors (Čibej, Jaka, Gorjanc, Vojko, Kosem, Iztok, Krek, Simon), date (2018), proceedings title, and conference name (XVIII EURALEX International Congress). An abstract is also visible on the right side of the detailed view.

Title	Creator	Date
A 3-Part, Fully Trilingual Tri-Directional Dictionary	Kernerman	1994
A Bilingual Corpus for Lexicographers	Citron and Widmann	2006
A Brief History from a Frustrated Seeker after Knowled...	Smith	1996
A Call for a Corpus-Based Sign Language Dictionary: A...	Majetić and Bago	2018
A Co-Occurrence Taxonomy from a General Language ...	Nazar and Renau	2012
A Combinatory Dictionary of English Discourse Connec...	Higashimori	1996
A Comparative Analysis of Definitions of Phrasal Verbs...	Perdek	2008
A Comparison Between COBUILD, LDOCE5 and CALD3:...	Chan	2012
A Computational Lexicographer's Workbench	Matsumoto et al.	1998
A Computational Multi-Layered Italian Lexicon for HLT ...	Ruimy	2006
A Computerized Model for Processing Lexical Combin...	L'Homme	1996
A conceptually organized bilingual dictionary as a fore...	Bratanić	1990
A Construction-based Approach to the Lexicalization o...	Fischer	1996

Item Type: Conference Paper

Title: A Call for a Corpus-Based Sign Language Dictionary: An Overview of Croatian Sign Language Lexicography in the Early 21st Century

Author: Majetić, Klara

Author: Bago, Petra

Editor: Čibej, Jaka

Editor: Gorjanc, Vojko

Editor: Kosem, Iztok

Editor: Krek, Simon

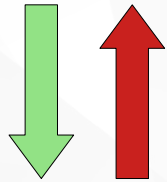
Date: 2018

Proceedings Title: Proceedings of the 18th EURALEX International Congress

Conference Name: XVIII EURALEX International Congress

Abstract: Many sign languages today are still not standardized nor accessible to a wider audience. Sign languages with high quality dictionaries are scarce. In this paper we give a brief description of sign languages and how they differ from spoken ones, in order to better understand the issues lexicographers might face when compiling dictionaries for these languages. We focus on Croatian Sign Language (HZJ), giving an overview of the current situation in HZJ lexicography following some criteria we find relevant for both online and printed sign language dictionaries. The criteria have been classified into twenty-five categories and applied to create a model for an online HZJ dictionary, briefly presented in this paper. By presenting an unsatisfactory status of HZJ lexicography, we are issuing an urgent call for the compilation of a HZJ corpus as a basis for a high quality dictionary that could benefit both the

LexBib RDF



```
<owl:NamedIndividual rdf:about="http://lexbib.org/lexbib/item#ConfEuralex2018_789">
  <rdf:type rdf:resource="http://lexbib.org/lexdo/BibItem"/>
  <bibo:pages>789-798</bibo:pages>
  <dcterms:date>2018</dcterms:date>
  <dcterms:title>Lexicographie et terminologie au XIXe siècle: Vocabularu romano-francesu [Vocabulai
  <bibo:authorList>
    <rdf:Seq>
      <rdf:li rdf:resource="http://lexbib.org/agents/person#AldeaMaria"/>
    </rdf:Seq>
  </bibo:authorList>
  <lexdo:publicationLanguage rdf:resource="http://lexvo.org/id/iso639-3/fra"/>
  <lexdo:objectLanguage rdf:resource="http://lexvo.org/id/iso639-3/ron"/>
  <lexdo:objectLanguage rdf:resource="http://lexvo.org/id/iso639-3/fra"/>
  <lexdo:container rdf:resource="http://lexbib.org/lexbib/item#ConfEuralex2018_0"/>
  <lexdo:event rdf:resource="http://lexbib.org/events/ConfEuralex2018"/>
  <lexdo:abstract rdf:resource="http://lexbib.org/lexbib/item#ConfEuralex2018_789_abstract"/>
  <lexdo:fullTextUrl rdf:resource="https://euralex.org/publications/lexicographie-et-terminologie-au
  <lexdo:zoteroItemUri rdf:resource="http://zotero.org/groups/1892855/items/GED5VD5P"/>
  <lexdo:zoteroItemID>GED5VD5P</lexdo:zoteroItemID>
  <rdf:type rdf:resource="http://purl.org/ontology/bibo/Article"/>
  <dcterms:isPartOf>
    <bibo:Proceedings>
      <dcterms:publisher>
        <foaf:Organization>
          <address:localityName>Ljubljana</address:localityName>
          <foaf:name>Ljubljana University Press</foaf:name>
        </foaf:Organization>
      </dcterms:publisher>
      <bibo:isbn13>978-961-06-0096-1</bibo:isbn13>
      <dcterms:title>Proceedings of the 18th EURALEX International Congress</dcterms:title>
      <bibo:editorList>
        <rdf:Seq>
          <rdf:li rdf:resource="http://lexbib.org/agents/person#CibejJaka"/>
          <rdf:li rdf:resource="http://lexbib.org/agents/person#GorjancVojko"/>
          <rdf:li rdf:resource="http://lexbib.org/agents/person#KosemIztok"/>
          <rdf:li rdf:resource="http://lexbib.org/agents/person#KrekSimon"/>
        </rdf:Seq>
      </bibo:editorList>
    </bibo:Proceedings>
  </dcterms:isPartOf>
</owl:NamedIndividual>
```

Export:

- one-click-action, one-way
- Creation of two items:
 - lexdo:BibItem
 - zotexport:Item

LexDo: Towards Lexicography as Knowledge Graph

- ▼ Access: <http://lexbib.org/lexdo>
- ▼ Standard RDF vocabularies
 - ▼ BIBO, PROV, Dublin Core, FOAF, Ontolex
- ▼ Top-level concepts:
 - ▼ Agent
 - ▼ Person/Organization/Software
 - ▼ Resource
 - ▼ Bibliographical resource/Lexical resource/Software
 - ▼ Subject
 - ▼ Subjects Headings controlled vocabulary
 - ▼ Language
 - ▼ Event
 - ▼ Place

Ontology Specification Draft

Used software:
VocBench 3, Protégé,
OpenRefine, Widoco

LexDo - Domain Ontology of Lexicogra

Release 2020-03-14T16:50:24.513+01:00

This version:

<http://lexbib.org/lexdo/1.0.2/>

Authors:

[David Lindemann](#)

Contributors:

[Christiane Klaes](#)

[Laura Giacomini](#)

Publisher:

Institut Jožef Stefan / Universität Hildesheim / UPV/EHU Uni

Download serialization:

Format [JSON LD](#) Format [RDF/XML](#) Format [N Triples](#) Format [TTL](#)

License:

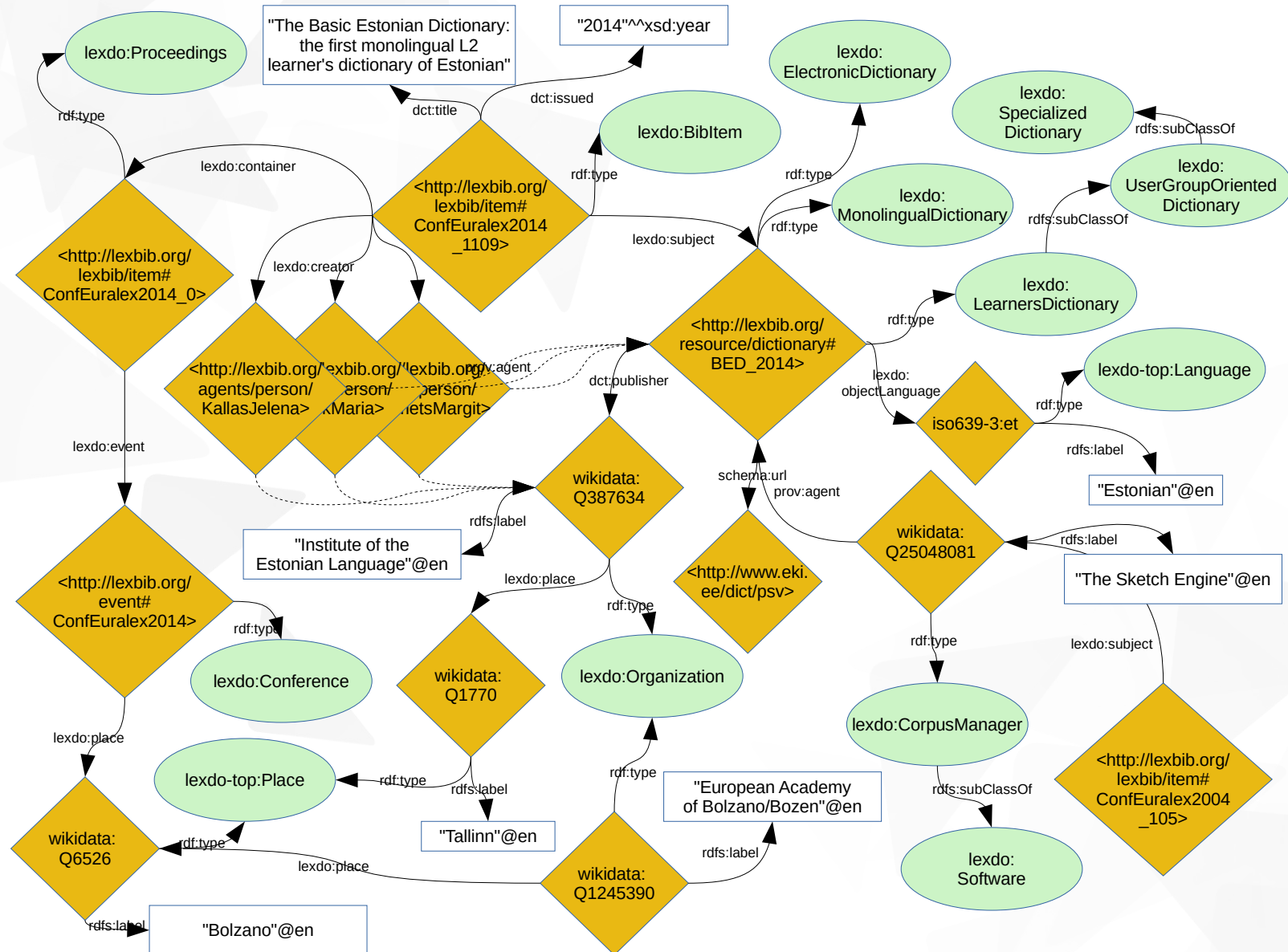
License <http://creativecommons.org/licenses/by/3.0/> [License](#)

Cite as:

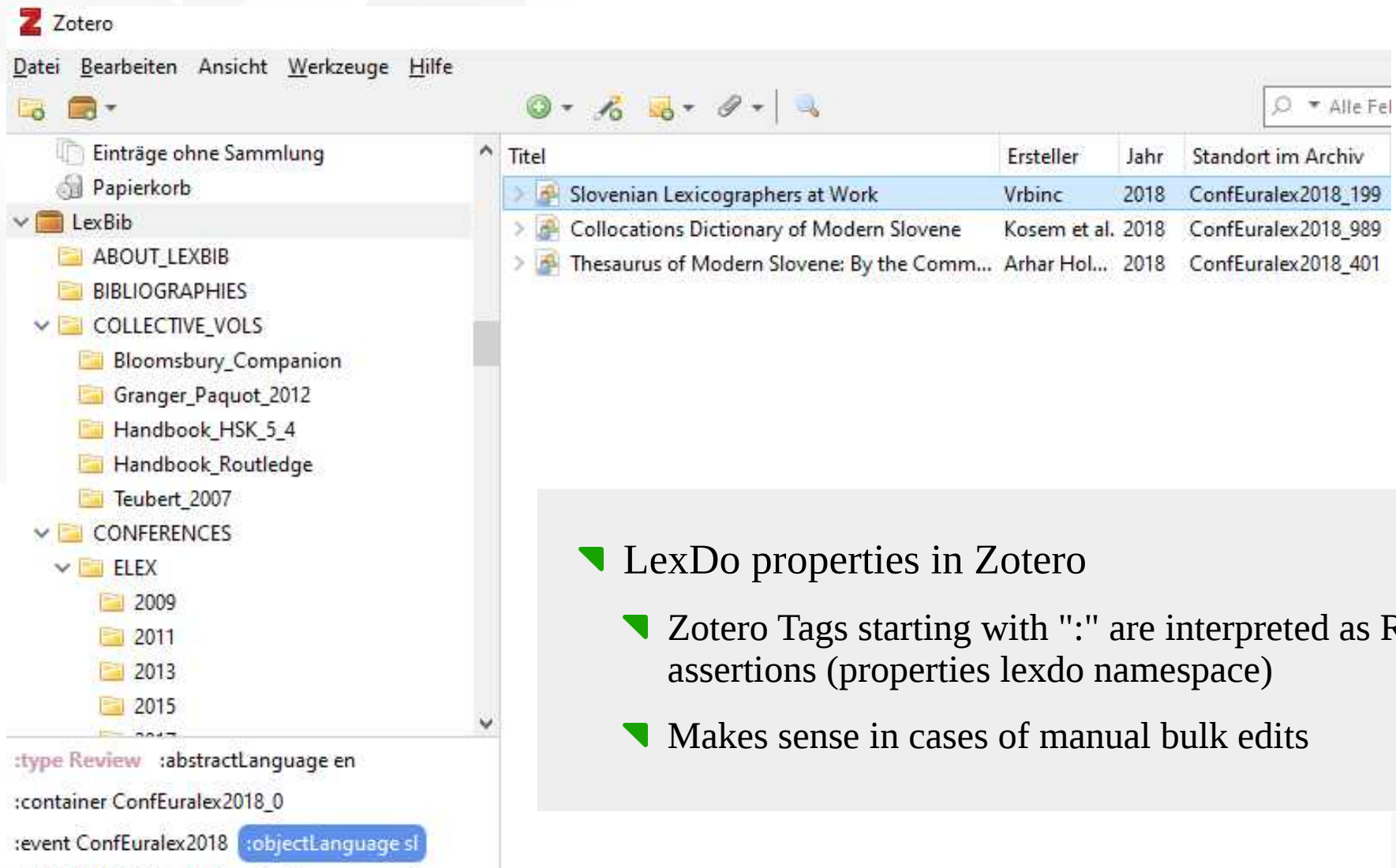
LexDo Domain Ontology of Lexicography and Dictionary Res
For full reference, see [About LexBib](#) at Zotero.

<http://lexbib.org/lexdo/1.0.2>

LexDo: Towards Lexicography as Knowledge Graph



LexDo properties in Zotero



The screenshot shows the Zotero application window. The left sidebar displays a hierarchical library structure under 'LexBib', including folders for 'COLLECTIVE_VOLS' and 'CONFERENCES'. The 'CONFERENCES' folder is expanded to show 'ELEX' subfolders for the years 2009, 2011, 2013, and 2015. The main pane shows a list of items with columns for 'Titel', 'Ersteller', 'Jahr', and 'Standort im Archiv'. The first item, 'Slovenian Lexicographers at Work', is selected. Below the list, the item's metadata is displayed, including 'Review' and 'ConfEuralex2018' tags.

Titel	Ersteller	Jahr	Standort im Archiv
Slovenian Lexicographers at Work	Vrbinc	2018	ConfEuralex2018_199
Collocations Dictionary of Modern Slovene	Kosem et al.	2018	ConfEuralex2018_989
Thesaurus of Modern Slovene: By the Comm...	Arhar Hol...	2018	ConfEuralex2018_401

Metadata for the selected item:

- :type Review
- :abstractLanguage en
- :container ConfEuralex2018_0
- :event ConfEuralex2018
- :objectLanguage sl

- ▼ LexDo properties in Zotero
 - ▼ Zotero Tags starting with ":" are interpreted as RDF assertions (properties lexdo namespace)
 - ▼ Makes sense in cases of manual bulk edits

LexDo: Lexicalizations of Subject concepts

- Subjects are labelled with concept representations stemming from reference resources
- Three test resources
 - Wiegand WLWF lemmata, multilingual
 - Müller-Spitzer index, English
 - Kipfer Glossary of Lexicographic Terms
- Modeled in SKOS-XL
 - A straightforward way to include additional information about the label, e.g. where to find reference in Wiegand / Müller-Spitzer / Kipfer

```
macrostructure wlwf 3682ni

---

IRI: http://lexbib.org/terms/WlwfTerms#Macrostructure_Wlwf_3682

---

belongs to  
lexical conceptc  
has facts  
in scheme op wlwf lemmata  
is concept of op macrostructure  
created dp "2020-02-14T00:36:11.810+01:00"^^date time
```

```
microstructure ms 058ni

---

IRI: http://lexbib.org/terms/MsTerms#Microstructure_Ms_058

---

belongs to  
lexical conceptc  
has facts  
in scheme op ms term index  
is concept of op microstructure  
created dp "2020-02-13T20:20:31.637+01:00"^^date time
```

<http://lexbib.org/lexdo/1.0.2>

BibItem subject indexation

- ▼ Represented as owl:ObjectProperty "lexdo:subject" (subproperty of dct:subject)
 - ▼ points to subject member of lexdo-top:Subject
 - ▼ subject concepts are part of a skos:ConceptScheme
 - ▼ lexdo RWO can also be subjects (???)
- ▼ Subject concept has skosxl label(s), that are matched in
 - ▼ BibItem title
 - ▼ BibItem abstract
 - ▼ BibItem (lemmatized) full text content
- ▼ Later: Term extraction in full texts
 - ▼ enables subject term weight ranking
 - ▼ enables discovery of new concepts and synonym lexicalizations

[urn:isbn:978-3-11-034116-4](http://www.lexbib.org/lexdo/1.0.2)ⁿⁱ

IRI: urn:isbn:978-3-11-034116-4

First Testitem: Müller-Spitzer (ed.) 2014 - Using Online Dictionaries

belongs to

[bib item](#)^c

[Book](#)^c

has facts

[creator](#)^{op} [muller spitzer carolin](#)

[publicationLanguage](#)^{op} [English](#)

[subject](#)^{op} [dictionary use](#)

<http://lexbib.org/lexdo/1.0.2>

LOD-ifying languages

▼ Languages in LexBib / LexDo

- ▼ Publication language
- ▼ Object language

```
lexdo:objectLanguage <http://lexvo.org/id/iso639-3/deu> ;  
lexdo:publicationLanguage <http://lexvo.org/id/iso639-3/eng> ;
```

▼ LEXVO Ontology

- ▼ URI built with ISO-639-3 codes
- ▼ Multilingual labels for languages
- ▼ Wikidata (inference via ISO-639-3 code)
 - ▼ e.g., where spoken ("country")
 - ▼ typological features (e.g. "agglutinative")
 - ▼ ...

subclass of	Vasconic languages
	▼ 0 references
	Southern European languages
	▶ 1 reference

part of	regional languages of France
	▶ 1 reference

native label	Euskara (Basque)
	▶ 1 reference

country	Spain
	▶ 1 reference
	France
	▶ 1 reference

LOD-ifying creators (authors, editors)

- ▼ Zotero author fields manually curated
 - ▼ Two persons with same name not allowed
 - ▼ Person name variants allowed, but we should try to avoid them
 - ▼ lexdo:Person individual created by Zotero export script
 - ▼ Example:
<http://lexbib.org/agents/person#KrekSimon>
- ▼ Refinement (conversion into linked data)
 - ▼ OpenRefine: VIAF, ORCID, Wikidata
 - ▼ Inference of more properties
 - ▼ at the moment, only from Wikidata
 - ▼ e.g. employer, other IDS
 - ▼ Zotero RDF output → CSV → OpenRefine → lexbib.org/agents/persons

The screenshot shows the OpenRefine interface with a reconciliation dialog box open. The dialog box is titled "Add columns from reconciled column [full name]". It has two main sections: "Add Property" and "Preview".

Add Property: A text input field contains "ORCID id". Below it, a list of "Suggested Properties" includes "cause of death", "child", and "country of citizenship".

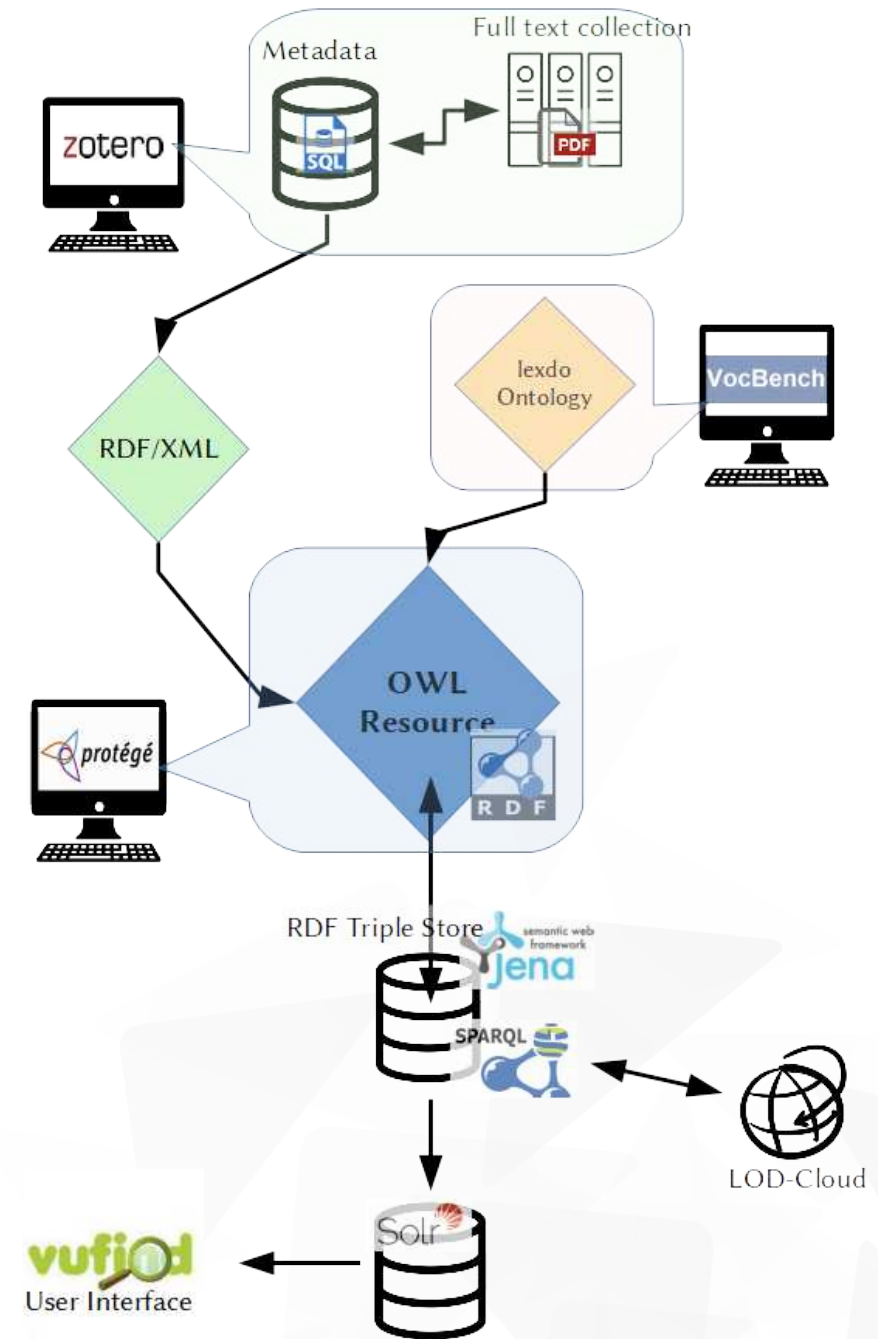
Preview: A table showing the results of adding the property. The table has three columns: "[full name]", "ORCID id", and "employer".

[full name]	ORCID id	employer
<not reconciled>		
<not reconciled>		
Simon Krek	0000-0001-8965-6863	University of Ljubljana
		University of Ljubljana
<not reconciled>		

Buttons for "remove" and "configure" are visible under each property in the preview table. A "Reset" button is located in the top right corner of the dialog box.

RDF data exhibit / Access GUI

- ▼ Mid-term scenario:
LexBib RDF triples →
Solr index for **VuFind** user frontend
- ▼ as in LexBib project proposal
(Lindemann et al. 2018)
- ▼ Quick interim solution???
- ▼ Data source triple store
- ▼ Listing of properties for BibItem, Creator,
etc. individuals
- ▼ Resources pointed to in the list as active
links
- ▼ Something like **SKOSMOS**, but for
LexBib (OWL and SKOS)



LexBib / LexDo links and publications

- ▼ Homepage: <http://lexbib.org>
- ▼ Papers: See <https://www.zotero.org/groups/1892855/lexbib/collections/E65TS5DR>
- ▼ ResearchGate: <https://www.researchgate.net/project/LexBib-Corpus-and-Bibliography>
- ▼ Github: https://github.com/dlindem/zotero_lexbib_rdf
- ▼ Kontakt: team@lexbib.org | david@lexbib.org | david.lindemann@ehu.eus
- ▼ Elexis project: <https://elex.is/>
- ▼ Elexifinder: <https://elex.is/tools-and-services/elexifinder/>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.