David Lindemann/Penny Labropoulou/Christiane Klaes

INTRODUCING LexMeta: A METADATA MODEL FOR LEXICAL RESOURCES

Abstract In this paper, we present **LexMeta**, a metadata model for the description of human-readable and computational lexical resources in catalogues. Our initial motivation is the extension of the **LexBib** knowledge graph with the addition of metadata for dictionaries, making it a catalogue *of* and *about* lexicographical works. The scope of the proposed model, however, is broader, aiming at the exchange of metadata with catalogues of Language Resources and Technologies and addressing a wider community of researchers besides lexicographers. For the definition of the LexMeta core classes and properties, we deploy widely used RDF vocabularies, mainly **Meta-Share**, a metadata model for Language Resources and Technologies, and **FRBR**, a model for bibliographic records.

Keywords Lexical resources metadata; linked data; Wikibase; semantic web

1. Introduction

In this paper we present LexMeta, a metadata model for the description of human-readable and computational lexical resources[1] in catalogues.

The goal is to develop a catalogue *of* and *about* lexicographical works to be integrated in the LexBib Wikibase Knowledge Graph of Lexicography and Dictionary Research, a research infrastructure targeting the lexicographic community. The LexBib project[2] (Lindemann/Kliche/Heid 2018; Kosem/Lindemann 2021) consists of various components among which LexBib Zotero[3] occupies a central place. This is a digital library of metalexicography research articles made available through the Zotero[4] platform, containing publicly available publication metadata, and a collection of full texts of articles available to the text processing objectives of the LexBib project.[5] It currently includes 10,000 metadata records for papers in several languages, out of which around 7,500 are included with their full texts. That bibliographical catalogue is represented as Linked Open Data (LOD) in LexBib Wikibase (Lindemann 2021) we present ongoing work concerning a workflow and software tool pipeline for collecting and curating bibliographical data of the domain of Lexicography and Dictionary Research, and data export in a custom JSON format as required by the Elexifinder application, a discovery portal for lexicographic literature. We present the employed software tools, which are all freely available and open source. A Wikibase instance has been chosen as central data repository. We also present requirements for bibliographical data to be suitable for import into Elexifinder; these include disambiguation of entities like natural persons and natural languages, and a processing of article full texts. Beyond the domain of Lexicography, the described workflow is applicable in general to single-domain small scale digital biblio-

[1] We use the terms "lexical resource" and "dictionary" interchangeably with a broad meaning, encompassing user dictionaries, general dictionaries, glossaries, thesauri, terminological lexica, etc.

[2] See https://lexbib.elex.is/wiki/Project:About.

[3] Accessible through https://lexbib.elex.is/wiki/LexBib_Zotero.

[4] Homepage at https://www.zotero.org/.

[5] For IPR reasons, we cannot make available physical copies of full text; nevertheless, where available, links to the locations where they can be accessed or downloaded from, are provided.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

graphies.","event":"SiKDD 21 Slovenian KDD Conference, October 4th, 2021","event-place":"Ljubljana","language":"en","publisher-place":"Ljubljana","title":"Zotero to Elexifinder: Collection, curation, and migration of bibliographical data","URL":"https://ailab.ijs.si/dunja/SiKDD2021/Papers/LindemannDavid.pdf","author":[{"family":"Lindemann","given":"David"}],"issued":{"date-parts":[["2021",10,4]]}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} . With the addition of metadata for dictionaries, the LexBib knowledge graph will cover lexicographical primary and secondary resources, along with other entity types related to both of these, such as persons, organisations, languages, places, events, and lexicographic terminology.

To increase the value and outreach of this catalogue, we foresee the import and export of metadata from and to other catalogues, especially those popular with our target audience. One such case is the CLARIN Virtual Language Observatory (VLO),[6] addressing researchers of the Social Sciences and Humanities disciplines. These catalogues serve different purposes and have, thus, adopted different approaches to the documentation of dictionaries: library catalogues of books mostly focus on bibliographical metadata, while catalogues of language resources, such as CLARIN, look at dictionaries (mainly those in digital form) as datasets and focus more on encoding information about their contents and accessing modes. Therefore, LexMeta seeks to bring together the metadata modelling approaches used in these two types of catalogues and cater for the description of lexical resources along both of these dimensions.

In the following sections, we present the background and main features of LexMeta, as well as its application in the LexBib catalogue. More specifically, section 2 presents the methodology for its development and gives an overview of the main models and deployed resources. section 3 describes the model itself illustrated with examples. section 4 introduces the current status of the LexBib catalogue of dictionaries and, finally, section 5 concludes with future plans.

2. Background

2.1 Requirements and methodology

The LexMeta model aims to cater for the description of lexical resources included in catalogues of libraries and repositories. It must satisfy the requirements and needs of the respective catalogue users but also have a broader outlook, considering recent developments and initiatives in the metadata and data-related areas, the most prominent being the formulation of the FAIR principles[7] (Wilkinson et al. 2016).

More specifically, in terms of content, the model must cover not only bibliographical metadata (e. g., title, author(s), publication date), but also information on the contents and accessibility of the resource, relations between versions of the same resource, and provenance metadata. It should also support easy discovery of the catalogue entries by both human users and machines, and thus exploit existing standards and best practices, especially those used by the involved communities. Extensibility and flexibility are important desiderata given the evolving data landscape. Interoperability with other schemas plays a crucial role in its design in order to facilitate exchange of metadata between catalogues.

6 See https://vlo.clarin.eu.

7 See https://www.go-fair.org/fair-principles/.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

For the design of the model, we first made an inventory of the metadata information that should be included in it based on the requirements of the envisaged use case. We also conducted a survey through which we identified a set of models and vocabularies that are popular in the targeted domains and explored their adoption for our needs as outlined in the next subsection.

2.2 Overview of models

For our survey we have investigated models and vocabularies used in the bibliographical and lexicographical domains and the domain of datasets. These are presented below with a short description of the features that are of interest for our model.

FRBR (Functional Requirements for Bibliographic Resources) is a conceptual model for describing bibliographic metadata (IFLA Study Group on the Functional Requirements for Bibliographic Records 1998)"language":"en","publisher-place":"Munich","title":"Functional Requirements for Bibliographic Records: Final Report.","URL":"http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records","author":[{"family":"IFLA Study Group on the Functional Requirements for Bibliographic Records","given":""}], "issued":{"date-parts":[["1998"]]}}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} . It is an international standard implemented in numerous local applications. FRBR distinguishes between the concepts of *Work* (e. g., an abstract notion of a lexicographical creation), *Expression* (the realisation of a single work, such as a certain version or edition), and *Manifestation* (the distribution of a single realisation, e. g., on paper, or as a digital dataset) as core classes.

BIBO (The Bibliographic Ontology) was developed in the Semantic Web community, to provide a generic RDF vocabulary for describing bibliographic resources and citation relations. Building on widely used vocabularies such as Dublin Core,[8] BIBO provides specific classes and properties to classify and describe documents in a Linked Data environment. BIBO properties may relate to all FRBR core concepts.

The Meta-Share ontology **(MS-OWL** or **MS)**[9] (Gavrilidou et al. 2012; McCrae et al. 2015) caters for language resources, including data resources (structured or unstructured datasets, lexica, language models, etc.) and technologies used for language processing (taggers, parsers, machine translation applications, etc.). It builds around three key concepts: *resource type*, *media type* and *distribution*, which give rise to the core classes of the model. Focusing on lexical resources, the class ms:LexicalConceptualResource (subclass of ms:LanguageResource) covers resources such as term glossaries, dictionaries, semantic lexica, ontologies, etc., organised on the basis of lexical or conceptual units (lexical items, terms, concepts, phrases, etc.) along with supplementary information (e. g., grammatical, semantic, statistical information, etc.). The class ms:DatasetDistribution represents the accessible form of a resource, e. g., a spreadsheet or plain text file with the contents of a lexicon, or an online dictionary accessible through a user interface.[10] Properties are assigned to the most relevant class. Descriptive and administrative metadata, such as those used for identification purposes (title, description, etc.), recording provenance (creation, publication dates, creators, providers,

[8] See https://www.dublincore.org/specifications/dublin-core/dcmi-terms/.

[9] See http://w3id.org/meta-share/meta-share.

[10] MS includes an additional class for media parts not presented here because the LexMeta model is currently restricted to textual resources.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

etc.), are assigned to the class ms:LanguageResource, while more technical features and classification elements are attached to the appropriate subclasses. Thus, properties for ms:LexicalConceptualResource encode the subtype (e. g., computational lexicon, ontology, dictionary, etc.), and the contents of the resource (unit of description, types of accompanying linguistic and extralinguistic information, etc.). The ms:DatasetDistribution class provides information on how to access the resource (i. e., how and where it can be accessed), technical features of the physical files (such as size, format, character encoding) and licensing terms and conditions.

DCAT[11] (Data Catalog Vocabulary) is an RDF vocabulary for representing data catalogues. For our purposes, we have looked into two of its core classes and their properties.[12] dcat:Dataset represents "a collection of data, published or curated by a single agent or identifiable community; the notion of dataset is broad and inclusive, covering data in many forms, including numbers, text, pixels, imagery, sound and other multi-media, and potentially other types". A dictionary or any other lexical resource can safely be considered a dataset in DCAT terms. dcat:Distribution represents an accessible form of a dataset such as a downloadable file. The design of Meta-Share has been influenced by DCAT; thus, ms:LexicalConceptualResource and ms:DatasetDistribution are represented as subclasses of dcat:Dataset and dcat:Distribution respectively. Further alignments between them are currently under development.

The **LexVoc** Vocabulary of Lexicographic Terms[13] is part of the LexBib Wikibase graph. It is a structured controlled list of terms related to lexicographical and metalexicographical concepts. It has been developed by re-using and extending term lists from various authoritative sources and organising them in semantic domains with several goals in mind (Kosem/Lindemann 2021, section 3); among other applications, LexVoc terms are used for the content-describing indexation of LexBib bibliographical items, and can be used for the classification of dictionaries along various parameters. LexVoc is implemented using the SKOS model.[14]

2.3 Technical implementation considerations

With regard to the implementation of the model, we have decided to follow the Linked Data paradigm.[15] To this end, we have considered Semantic Web technologies (e. g., RDF, OWL,

[11] See https://www.w3.org/TR/vocab-dcat-3/

[12] The current version (v3), published in January 2022, as a working draft, is based around seven core classes. One of these, namely dcat:DatasetSeries, was introduced in this version. It also has a potential interest for the model and we are currently investigating its usefulness. This class stands for "a dataset that represents a collection of datasets published separately but sharing common characteristics that group them together".

[13] See http://lexbib.elex.is/wiki/LexVoc.

[14] The SKOS standard can be used for the representation of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web. For more information, see https://www.w3.org/TR/2009/REC-skos-reference-20090818/.

[15] For an introduction to Linked Data, see https://www.w3.org/standards/semanticweb/data.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

SKOS) and the model behind Wikidata,[16] an open knowledge graph based on the Wikibase software.[17]

LexBib Wikibase is an instance of Wikibase, an open source software solution. Wikimedia Germany,[18] a non-profit organisation, is in charge of providing Wikibase as a service to a broader community, by an endeavour called Wikibase Cloud.[19] The goal is to enable an ecosystem of federated Wikibases, with Wikidata as the central hub; federation is possible through shared persistent identifiers and an interoperable querying standard, SPARQL, which allows for accessing different Wikibases at the same time. A Wikibase may provide additional data describing entities represented on Wikidata.

Advantages of Wikibase compared to other Linked Open Data (LOD) database infrastructures are described in Lindemann (2021) we present ongoing work concerning a workflow and software tool pipeline for collecting and curating bibliographical data of the domain of Lexicography and Dictionary Research, and data export in a custom JSON format as required by the Elexifinder application, a discovery portal for lexicographic literature. We present the employed software tools, which are all freely available and open source. A Wikibase instance has been chosen as central data repository. We also present requirements for bibliographical data to be suitable for import into Elexifinder; these include disambiguation of entities like natural persons and natural languages, and a processing of article full texts. Beyond the domain of Lexicography, the described workflow is applicable in general to single-domain small scale digital bibliographies.","event":"SiKDD 21 Slovenian KDD Conference, October 4th, 2021","event-place":"Ljubljana","language":"en","publisher-place":"Ljubljana","title":"Zotero to Elexifinder: Collection, curation, and migration of bibliographical data","URL":"https://ailab.ijs.si/dunja/SiKDD2021/Papers/LindemannDavid.pdf","author":[{"family":"Lindemann","given":"David"}],"issued":{"date-parts":[["2021",10,4]]},"suppress-author":true}],"schema":"https://github.com/citation-style-language/schema/raw/master/csl-citation.json"} . Furthermore, Wikibase as an infrastructure supports FAIR data and metadata; (meta)data in Wikibase are

- findable by machines through unique and persistent identifiers,
- accessible using standardised protocols (in particular, SPARQL),
- interoperable through the use of broadly used vocabularies that follow the same FAIR principles, and allow cross-references to other datasets, and
- reusable through appropriate licensing.

3. LexMeta Model

3.1 LexMeta presentation

Through the analysis of the descriptive requirements for our model and the survey of the models and vocabularies, we have established alignments between them and identified conceptual gaps, for which we have introduced new elements in a unified model.

16 See https://www.mediawiki.org/wiki/Wikibase/DataModel.

17 See https://wikiba.se/.

18 See http://wikimedia.de.

19 See http://wikibase.cloud.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

The LexMeta model is built around three main classes,[20] which follow the FRBR and relevant MS conceptual distinctions:

- the *Lexicographic Work* (lwb:Q41)[21] corresponds to the abstract notion of a lexicographical creation and is defined as subclass of frbr:Work;[22]
- *Lexical/Conceptual Resource* (*LCR*, lwb:Q4) represents the realisation of a single work, such as a certain version or edition of a lexicographic work, and corresponds to frbr:Expression and ms:LexicalConceptualResource;
- *LCR Distribution* (lwb:Q24) is the physical form in which a lexicographical work is realized (e. g., as a printed book or as a digital file), and is aligned to frbr:Manifestation and ms:DatasetDistribution.

This distinction allows us to group and link different publications (e. g., print publications, reprints, and digital versions) with the same content as well as to describe them more consistently by attaching their properties at the appropriate level.

Lexicographic Work groups the various editions and versions (*expressions*/*LCRs*) of the same *work*. Content-describing metadata are common across *manifestations* (*distributions*) of the same *expression* (*LCR*) and are assigned to the *LCR* level. Publication metadata and technical features are attached at the *distribution* level.

More specifically, properties for a *Lexicographic Work* include identification metadata (title, identifier) and the *has realisation* property (lwb:Q118, frbr:realization) that links it to the *LCR* objects.

Properties attached to the *LCR* class relate to identification, administrative and provenance metadata (e. g., title,[23] author, holder of Intellectual Property Rights, etc.) that are common across all its Distributions. The property *has distribution* (lwb:P55) is used to link the *LCR* to one or more *LCR Distributions* while specific properties (taken from MS) are used to relate different *LCRs* to each other, e. g. *replaces LCR* (lwb:P135, ms:replaces). To encode the language(s) of the contents, four distinct properties are included: *source* and *target language* (for multilingual resources), *object language* and *metalanguage*. Properties describing LCR structure and type include the following:

- *lemma type* (lwb:P151), describing types of headwords included in a dictionary (e. g., single-word or multi-word units, abbreviations, neologisms, etc.),
- *linguality type* (lwb:P115), indicating whether the LCR describes one, two or more languages,
- *dictionary scope type* (lwb:P90), pointing to dictionary typology terms, such as "learner dictionary", "dialect dictionary", "etymological dictionary",

[20] Hereafter, we use the terms *class* and *property*, as in RDF vocabularies, to represent the objects we wish to describe and their features respectively.

[21] The namespace prefix "lwb" (short for "LexBib Wikibase" resolves to http://lexbib.elex.is/entity/).

[22] The MS ontology has no similar class; for the connection between versions of the same resource, it relies solely on properties that link them together (e. g. ms:isContinuedby, ms:isPartOf, etc.).

[23] Title (and other identification data) is a property that can be used for all the classes. This is deliberate to allow for cases where, for instance, distributions have different titles from that of the LCR and between them (e. g. "Paperback edition of Dictionary X", "Dictionary X: the online version", etc.).

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

- *dictionary function type* (lwb:P120), pointing to basic terms describing communicative and cognitive dictionary functions, e. g. "text translation",
- *dictionary access type* (lwb:P121), with two values, "onomasiological" and "semasiological dictionary",
- *microstructure feature* (lwb:P127), pointing to terms describing microstructural data presentation features as well as linguistic features of the presented content,
- *dictionary text part* (lwb:P152), indicating parts present in the dictionary text, such as front and back matters, and types of entries.

At *LCR Distribution* we attach publication metadata (e. g., publication date, publisher, ISBN), as found in a library catalogue, compatible with how publication metadata are represented in LexBib for metalexicographical publications.[24] We also attach properties describing how they can be accessed, i. e., the form of access or distribution type (e. g., "dictionary book publication" or "dictionary app.") and the URL where they can be accessed or downloaded. Where possible, we have opted for SKOS controlled vocabularies instead of free text to increase consistency and standardisation. Re-use of existing vocabularies, such as LexVoc, is preferred. In some cases, we have imported and enriched the LexMeta vocabularies with terms from other vocabularies. For example, the vocabulary containing terms that describe dictionary microstructure features is an extension of the MS vocabulary of content types, which is used in the range of the property ms:linguisticInformation.

3.2 Implementation

For the implementation of the model, we have decided to use two conventions: (1), Following the Wikibase data model, as an ontology of Wikibase entities, since the catalogue will be integrated in the LexBib Wikibase, and (2), as an OWL ontology,[25] a widely used formal knowledge representation language for the description of digital data.

In a more detailed documentation,[26] we specify the LexMeta core classes as implemented in the LexBib Wikibase, and the LexMeta properties attached to each of these classes, their datatype, and, for properties that take values from controlled vocabularies or classes of items, the respective set of values. We also include the alignments to the classes and properties of the vocabularies presented in section 2.2.

In the LexBib Wikibase implementation, LexMeta classes and properties are represented using URIs from the LexBib Wikibase's own namespace, and following Wikibase naming conventions, i. e. numeral identifiers preceded by the letter Q for items (i. e. classes and instances), and the letter P for properties. The LexMeta ontology is currently under construction. Where possible, we opt for re-using classes and properties from other vocabularies (mainly Meta-Share, BIBO, DCterms, etc.) instead of creating new ones.

The alignment between the two forms is foreseen at both sides. At the LexMeta OWL side, the OWL equivalence semantic relations can be used for linking to the LexBib Wikibase entities. In the LexBib Wikibase, this is already represented with a property of type "exter-

[24] This allows, at the same time, the creation of bibliographic items for LCR distributions on LexBib Zotero in a straightforward way.

[25] See https://www.w3.org/OWL/.

[26] Accessible at http://lexbib.elex.is/wiki/LexMeta.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

nal identifier" (lwb:P42), which links to the identifier of the equivalent LexMeta OWL entities.[27] Hence, in a data export of the LexBib entries, the LexBib identifier can be translated to its LexMeta OWL equivalent.[28]

4. Population of LexBib

The LexBib catalogue is already populated with metadata of dictionaries; that was initially done for a set of example items which were manually created and annotated with properties from our model.

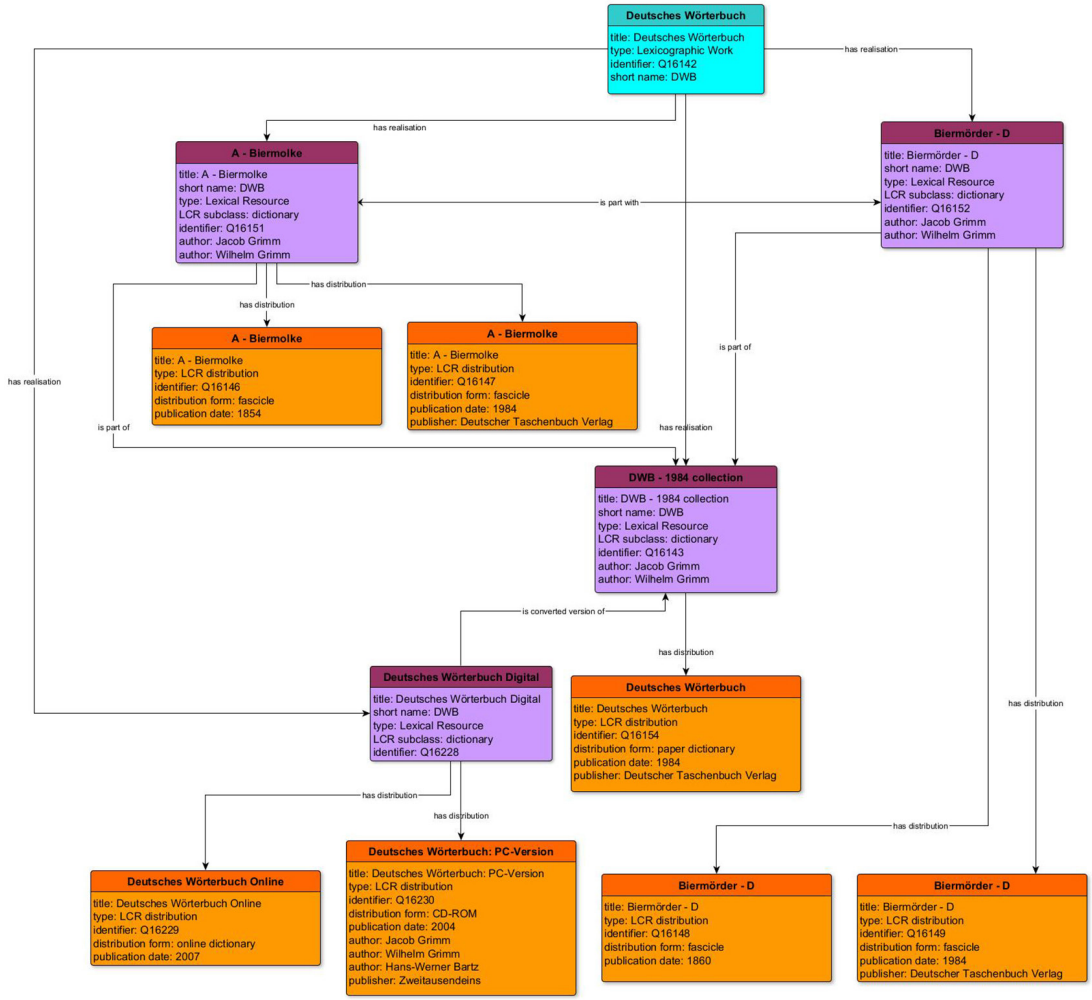

Fig. 1: Relations between instances of the LexMeta core classes

As an example that illustrates the complexity of relations our model allows between entities of the three core classes (see Fig. 1), *Deutsches Wörterbuch*, a lexicographical endeavour started in the mid 1850ies by the Grimm brothers, can be represented as one *Lexicographic*

[27] LexMeta OWL namespace is http://w3id.org/meta-share/lexmeta/.

[28] The Wikibase data model has also a particular data structure vis-a-vis the RDF model. Thus, the mapping process includes a step for the conversion of this structure, see http://lexbib.elex.is/wiki/LexMeta_OWL.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

work[29] with several expressions as *LCR*. The work was initially released as a set of fascicles, with different contributors and content features, and each had a different distribution: the book publications date from 1854 (fascicle 1) to 1954 (fascicle 32). After 1984, the fascicles are reprinted, with the same contents as the original ones, and are thus considered distributions of the original LCR. At the same time, they were issued as a complete collection, with a different size (i. e., the original LCR is linked to this with a part-of relation). This collection distributed in print in 1984 was later converted to a digital resource, which is distributed as an offline electronic dictionary, and also made accessible through a web portal.[30]

Other metadata entries already accessible at LexBib Wikibase stem from various source catalogues, such as OBELEX-dict,[31] Glottolog,[32] Worldcat,[33] and Wikidata. Properties that relate items describing dictionaries to items describing metalexicographical publications are part of LexMeta, namely *is reviewed in* (lwb:P26) and *cites* (lwb:P147, bibo:cites), which enables setting review and citation relations in the graph.

5. Conclusions and outlook

In this paper, we have presented the LexMeta model for lexical resources and its use in the population of the LexBib knowledge graph with metadata of dictionaries.

We are currently in discussions with scholars from the lexicographical and linguistic linked data communities in the framework of the ELEXIS[34] and NexusLinguarum[35] projects respectively and expect valuable feedback from them that will be used for the improvement of the model and its documentation. We are also collaborating on proposals aiming at a (community-driven) curation of lexicographical primary and secondary resources metadata on the LexBib wikibase, including assertions regarding review and citation relations.

Among our future plans is the enrichment of the LexBib catalogue with (mass) imports of metadata from other catalogues and, if and where needed, alignment of LexMeta with models used for these catalogues as well as with more general widespread metadata models for data resources.

In addition, the LexBib catalogue is planned to be made available through a CLARIN Knowledge Centre[36] (under construction) dedicated to Lexicography. In this case, the metadata of dictionaries will be exposed for harvesting by the CLARIN VLO, which is based on the OAI-PMH protocol and the use of metadata profiles that are compatible with the Component MetaData Infrastructure (CMDI) framework[37] (Broeder et al. 2012; International Organization for Standardization 2020). The conversion of the metadata into a CMDI-compatible profile can benefit from the fact that the Meta-Share schema is already included

[29] The URI of *Deutsches Wörterbuch*, an entity of class Work, is http://lexbib.elex.is/entity/Q16142.

[30] Graphical representations of these relations are available at https://lexbib.elex.is/wiki/Dictionaries.

[31] See https://www.owid.de/obelex/dict/en.

[32] See https://glottolog.org/langdoc, selecting “Doctype dictionary”.

[33] Accessible at https://www.worldcat.org/.

[34] Homepage at https://elex.is/.

[35] Homepage at https://nexuslinguarum.eu/.

[36] See https://www.clarin.eu/content/knowledge-centres.

[37] See https://www.clarin.eu/content/component-metadata.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

among them and can therefore be based on the re-use of the Meta-Share entities in the LexMeta Model.

References

Broeder, D. et al. (2012): CMDI: a component metadata infrastructure. In: Proceedings of the workshop describing language resources with metadata: towards flexibility and interoperability in the documentation of language resources. LREC 2012, May 22, 2012, Istanbul, Turkey. Paris, pp. 1–4. https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/10867.

Gavrilidou, M. et al. (2012): The META-SHARE Metadata Schema for the description of language resources. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12). Istanbul, Turkey, pp. 1090–1097. http://www.lrec-conf.org/proceedings/lrec2012/pdf/998_Paper.pdf.

IFLA Study Group on the Functional Requirements for Bibliographic Records (1998): Functional requirements for bibliographic records: final report. Munich. http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records.

International Organization for Standardization (2020): ISO 24622-1:2015: Language resource management – Component Metadata Infrastructure (CMDI) – Part 1: The Component Metadata Model. ISO. https://www.iso.org/standard/37336.html.

Kosem, I./Lindemann, D. (2021): New developments in Elexifinder, a discovery portal for lexicographic literature. In: Gavriilidou, Z./Mitits, L./Kiosses, S. (eds.): Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress, 7–11 September 2021, Alexandroupolis, Vol. 2. Alexandroupolis, pp. 759–766. https://euralex2020.gr/proceedings-volume-2/.

Lindemann, D. (2021): Zotero to Elexifinder: collection, curation, and migration of bibliographical data. In: SiKDD 21 Slovenian KDD Conference, October 4th, 2021. Ljubljana. https://ailab.ijs.si/dunja/SiKDD2021/Papers/LindemannDavid.pdf.

Lindemann, D./Kliche, F./Heid, U. (2018): LexBib: a corpus and bibliography of netalexicographcal publications. In: Proceedings of EURALEX 2018. Ljubljana, pp. 699–712. http://euralex.org/publications/lexbib-a-corpus-and-bibliography-of-metalexicographical-publications/.

McCrae, J. P. et al. (2015): One ontology to bind them all: the META-SHARE OWL Ontology for the Interoperability of Linguistic Datasets on the Web. In: Gandon, F. et al. (eds.): The Semantic Web: ESWC 2015 Satellite Events, pp. 271–282. https://doi.org/10.1007/978-3-319-25639-9_42.

Wilkinson, M. D. et al. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: Scientific Data 3, p. 160018. http://doi.org/10.1038/sdata.2016.18.

Contact information

David Lindemann
UPV/EHU University of the Basque Country (Spain)
david.lindemann@ehu.eus

Penny Labropoulou
Institute for Language and Speech Processing (ILSP)/Athena R. C. (Greece)
penny@athenarc.gr

Christiane Klaes
TU Braunschweig (Germany)
University Library
c.klaes@tu-braunschweig.de

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Acknowledgements

The research presented in this article has received funding and support from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 731015 (ELEXIS, https://elex.is), the COST Action NexusLinguarum European network for Web centered linguistic data science (CA18209), supported by COST (European Cooperation in Science and Technology, https://www.cost.eu/), from Monumenta Linguae Vasconum 6 research project (PID2020-118445GB-I00), funded by the Government of Spain, and from the research group IT1344-19, funded by the Basque Government.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.