elexis european lexicographic infrastructure

D1.5

Best practices for lexicography – final report

Author(s): Carole Tiberius, Simon Krek, Michal Mechura, John McCrae, Toma Tasovac

Date: 31 July, 2022

H2020-INFRAIA-2016-2017

Grant Agreement No. 731015

ELEXIS - European Lexicographic Infrastructure

D1.5 Best practices for lexicography – final report

Deliverable Number:	D1.5
Dissemination Level:	Public
Delivery Date:	July 31, 2022
Version:	1
Author(s):	Carole Tiberius, Simon Krek, Michal Mechura, John McCrae, Toma Tasovac

Project Acronym:	ELEXIS
Project Full Title:	European Lexicographic Infrastructure
Grant Agreement No.:	731015

Deliverable/Document Information

Project Acronym:	ELEXIS
Project Full Title:	European Lexicographic Infrastructure
Grant Agreement No.:	731015

Document History

Version Date	Changes/Approval	Author(s)/Approved by
July 20, 2022	Initial draft	Carole Tiberius, Simon Krek, Michal Mechura, John McCrae, Toma Tasovac
July 25, 2022	Internal revision	Simon Krek
July 30, 2022	External evaluation	Jelena Kallas

Table of Contents

1. About this report

This report is the final report detailing best practices for lexicography (both multilingual and monolingual, as well as historical and contemporary lexicography). This deliverable is part of task 1.3. The aim of this task was to facilitate the creation of lexicographic resources in European institutions, by creating robust documentation, guidelines and collections of best practices in order to promote clearly defined workflows for producing, describing and annotating lexicographic resources (both synchronic and diachronic). The current report builds on the intermediate report D1.2 Best practices for lexicography - intermediate report and describes how ELEXIS has contributed to best practices in lexicography.

The report is structured as follows. Section 2 gives a summary of the surveys that have been carried out in ELEXIS to gain insight into current lexicographic practices in EUROPE. Section 3 discusses recent developments in the ELEXIS interoperability formats, TEI Lex-0 and Ontolex-Lemon, and introduces the ELEXIS data model, DMLex, which will be further developed after the end of the project. Section 4 describes the tools that have been developed within ELEXIS that will facilitate the creation and publishing of lexicographic resources in European institutions. In section 5, the relation to CLARIN is summarised and section 6 discusses the future plan for the Lexicographic Data Seal of Compliance. Section 7 concludes the deliverable.

2. ELEXIS surveys on lexicographic practices in EUROPE

Due to technological advances the field of lexicography has undergone radical changes over the past few decades, which has been noted by various authors. For instance, Leroyer and Køhler Simonsen (2020, p.184) write that „the digital revolution […] is leading to metamorphoses not only in dictionary making processes and dictionary forms, but also in dictionary use and in the general status of lexicography“. Michael Rundell observes that the field finds itself in a transitional phase and as yet there is little consensus on the way forward (Rundell 2015, p.310). Addressing these issues and paving the way for future lexicography, was precisely the goal of the Horizon 2020 ELEXIS project. In order to gain more insight into this fast changing field and into current lexicographic practices in Europe a number of surveys have been carried out in the ELEXIS project, three of which are particularly relevant in the context of this deliverable.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

- Survey on Lexicographic practices: A Survey of Lexicographers' Needs (45 countries (36 European and 9 outside Europe)) This survey was specifically targeted at individual lexicographers.

- Survey on Lexicographic practices: A Survey of Lexicographers' Needs for Lexicographic Partner Institutions (11 countries). This survey was targeted at the lexicographic partner institutions[1].

The results of these two surveys have been reported in D1.1 Lexicographic practices in Europe: a survey of user needs and (Kallas et al. 2019). Although these two surveys already gave us a lot of data, the institutional survey was only completed by 11 lexicographic partner institutions in the consortium, which is not really enough to generalise the findings. Therefore it was decided to extend the survey for the partner institutions to the observer institutions in 2020-2021. The survey for observer institutions constituted a revised and upgraded version of the survey which was originally conducted among the lexicographic partner institutions in 2018. It was the longest survey of the three. It contained 121 questions divided into 6 sections: (1) General information, (2) Types of lexicographic resources. Software and tools supporting the workflow, (3) Publication and access. Crowdsourcing and Gamification, (4) Retrodigitised dictionaries, (5) Data formats. Metadata. Availability, (6) Past and Future. Similar to the survey for the partner institutions, the intention was that one survey would be completed per institution and that it would be completed by a representative on behalf of the institution. The survey was opened from 13 July 2020 till 9 November 2021 to allow as many observer institutions as possible to complete it. Towards the end of this period personalised reminders were sent out. This had a positive effect as we achieved a response rate of 96%; 54 out of the 56 observers responded. Below, we briefly summarise some of the main results. For more information, the reader is referred to (Tiberius et al. 2022). For a detailed description of the earlier results the reader is referred to Deliverable 1.1.

Considering the type of institution carrying out lexicographic work, we see that there were slightly more universities than public institutions among the observer institutions that completed the survey, whereas the lexicographic partner institutions are mostly public institutions or non-profit

[1] The ELEXIS lexicographic partner institutions are the Austrian Academy of Sciences, Institute for Bulgarian Language Prof Lyubomir Andreychin, Society for Danish Language and Literature, Institute of the Estonian Language, Trier University, Trier Center for Digital Humanities, Hungarian Academy of Sciences, Research Institute for Linguistics, K Dictionaries Ltd, Dutch Language Institute, Belgrade Center for Digital Humanities, Jožef Stefan Institute, and the Real Academia Española.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

organisations. To date there are no private / commercial companies among the observers. There is one commercial company among the ELEXIS lexicographic partners.

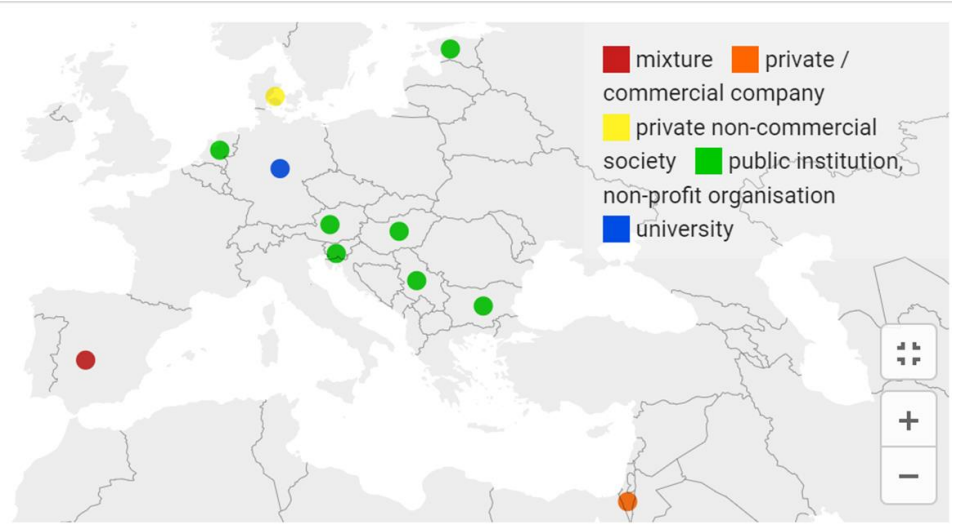


Figure 1. Type of partner institutions carrying out lexicographic work

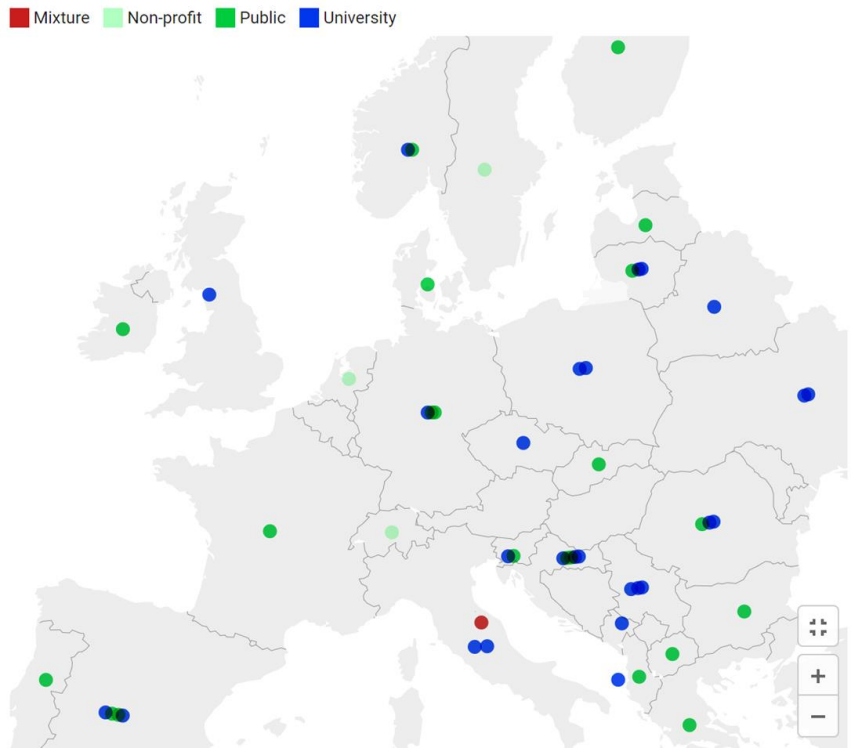


Figure 2. Type of observer institution carrying out lexicographic work

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

Both partner and observer institutions have a varied lexicographic expertise ranging from general and terminological dictionaries to specialised, learner's, historical, and dialect dictionaries.

Expertise on terminological dictionaries is, however, more represented among the observer institutions than among the lexicographic partner institutions.

Most institutions receive funding for their lexicographic work at a national level. This is in line with earlier results (Kallas et al. 2019a, p.57) which suggested that lexicographic work in Europe is mainly done in public institutions and non-profit organisations. It also corresponds with the findings of the European survey on dictionary use and culture (Kosem et al. 2019, p.96) where it was reported that in the majority of the countries participating in the survey, monolingual dictionaries are published solely or mainly by public institutions funded by the government. This observation is further confirmed by the answers on lexicographic expertise which show that monolingual projects are primarily carried out by public institutions, whereas bilingual and multilingual projects are mentioned more frequently by universities.

The main observations that we draw from the section on software and tools are that a lot of different systems seem to be available. A lot of different systems were mentioned, commercial, open-source and in-house systems. However, some observer institutions noted a lack of information regarding usability and effectiveness of these available commercial and open-source CQSs and DWSs. This is something we intend to take up in the context of the CLARIN ELEXIS Knowledge Centre for Lexicography. (For more on ELEXIS-KC see deliverable D6.5 - Final ELEXIS interoperability report including interaction with CLARIN/DARIAH services.)

Of the two systems, CQSs are used commonly by observer institutions as well as by partner institutions and both seem satisfied with the system they use. However, for DWSs, the situation at the observer institutions is clearly different from that of the partner institutions. Less than half use a DWS and of those that do not, 14 feel that they need a DWS urgently. Reasons mentioned by the observer institutions for not using a DWS are financial difficulties in purchasing lexicographic software or tools, but also the absence of knowledge and technical skills. This suggests that open-source tools such as Lexonomy, which was further developed in the context of ELEXIS, are much needed.

Furthermore, some observer institutions mention the availability of documentation and training materials as preliminary requirements for adopting a particular solution. Training and education are

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

part of the ELEXIS agenda and the ELEXIS curriculum[2] on Dariah Campus provides courses on mastering the ELEXIS tools.

Looking at the results from the section on retrodigitisation, we see that interest in retrodigitisation of printed dictionaries is observed in the whole lexicographic community – among the ELEXIS partners and observers. In both surveys similar procedures and software tools were mentioned for the different phases of retrodigitisation (image capture, text capture, data encoding and data enrichment). This is reassuring and suggests that there are already some best practices in place for the retrodigitisation workflow.

Although a shift can be observed from non-structured data to structured data, there are still quite a few institutions (46% of the observer institutions and 36% of the lexicographic partner institutions) using non-structured data format (e.g. in Microsoft Word) for at least some of their projects. This is frequently mentioned as a major hurdle for technological advances. ELEXIS aims to overcome this obstacle with the development of a general open standards based framework for internationally interoperable lexicographic work within OASIS (see section 3.3)

Considering the obstacles that were mentioned, one of the biggest concerns seems to be funding. The need for funding is voiced in all the ELEXIS surveys and in all parts of Europe, from Ukraine to Iceland, from Portugal to Sweden, although it seems even more urgent in Eastern Europe where the phrase 'lack of funding' tends to be used, whereas in Western Europe the respondents speak of 'difficulties' obtaining funding. In addition, concerns are expressed about the low status of lexicographic work, which forms a constant worry for many institutions.

In line with the results from the lexicographic partner institutions and the individual lexicographers, some observer institutions expressed their concern about the low quality and reliability of (semi-)automatically built resources while high quality lexicographic data is still kept closed under restrictive licences (both, public institutions and private publishing houses). Within ELEXIS serious efforts have been made to address licensing issues (Boelhouwer et al. 2020) and a number of flexible and diverse licensing options have been identified to encourage contribution of data (or parts of it) to the

[2] https://campus.dariah.eu/curriculum/the-elexis-curriculum

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

Dictionary Matrix. The survey results show that the process of making lexical resources more openly available has already started in the lexicographic community.

Most partners and observers make their dictionaries available online for free and more and more institutions attach a Creative Common licence to their dictionaries. However, access to the source data for reuse by others is still more restricted. This suggests that more promotion and raising awareness is needed to open up lexicographic data.

A change in the role of lexicographers, as well as a shift in skills, can also be observed. These days, lexicographers are commonly involved in project management, data management, fundraising, teaching, and public relations. Also, there is a shift in the role of lexicographic institutions, as they become more of a data provider and less of a dictionary publisher. One of the ELEXIS goals is precisely to enable reuse of lexicographic data in other fields.

The results from all surveys have already provided valuable input for various tasks within the ELEXIS project, and will continue to inspire future developments within the infrastructure. On the basis of the combined results, a lexicographic practice map of Europe can be devised, which is something we will explore in the context of a planned journal publication.

3. Standards in Lexicography

The current report builds on the intermediate report D1.2 which discussed data formats and standards used in lexicography and gave an introduction to the two ELEXIS interoperability formats, TEI Lex-0 and Ontolex-Lemon. Here we will briefly summarise the main developments in TEI Lex-0 and Ontolex-Lemon since January 2020, before discussing the ongoing work on DMLex, the data model which originates from the ELEXIS project.

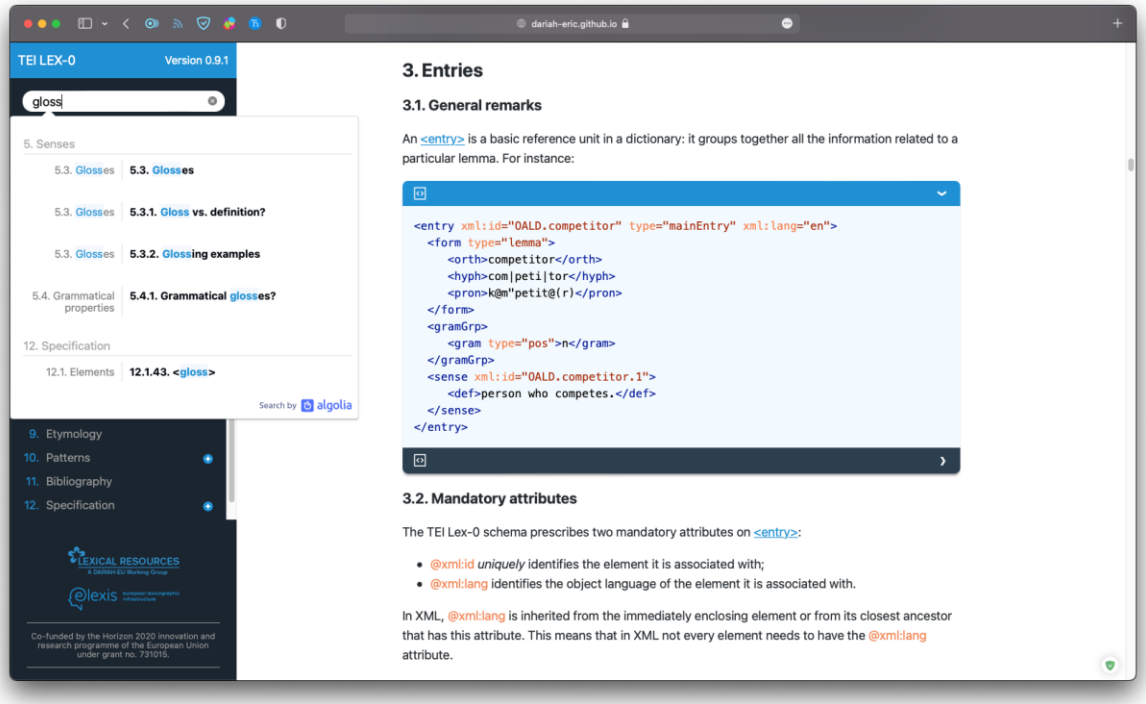
3.1. TEI Lex-0

TEI Lex-0 is both a technical specification and a set of community-based recommendations for encoding machine-readable dictionaries. It is rooted in the Guidelines of the Text Encoding Initiative (TEI) and delivered as a customization of the TEI schema. TEI Lex-0 aims at establishing a baseline

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

encoding and a target format to facilitate the interoperability of heterogeneously encoded lexical resources. TEI Lex-0 has a special focus on retrodigitised dictionaries.[3]

TEI Lex-0 is hosted by the DARIAH Working Group "Lexical Resources" in a GitHub repository: https://github.com/DARIAH-ERIC/lexicalresources. The TEI Lex-0 format is actively and openly discussed using the GitHub ticketing system, while the TEI Lex-0 Guidelines have – in large part thanks to the ELEXIS project – received a complete overhaul and are now searchable and available for online consultation in a user-friendly format, with multiple examples from real-life dictionaries and extensive notes on the reasoning behind certain encoding practices.



The ELEXIS-supported work on TEI Lex-0 has been well-received and widely appreciated by the community. As a result, the DARIAH WG Lexical Resources was awarded the 2020 Rahtz Prize for TEI Ingenuity[4]. This prize is awarded by the Text Encoding Initiative (TEI) Consortium to an individual or team judged to have made a significant contribution to the consortium's mission of developing and

[3] Of the lexicographic resources that were collected for ELEXIS (see D1.6), 8 were encoded in TEI Lex-0. All of those were retrodigitised dictionaries.

[4] https://www.dariah.eu/2020/11/20/dariah-working-group-on-lexical-resources-wins-innovation-prize/

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

maintaining a set of high-quality guidelines for the encoding of humanities texts, and supporting their use by a wide community of projects, institutions, and individuals.

> Members of the DARIAH Working Group Lexical Resources have made a valuable contribution to the Dictionaries Chapter of the TEI Guidelines. Their efforts and their expertise have been formidable and highly appreciated by the TEI Community for many years. — *Martina Scholger, Chair of the TEI Technical Council*

TEI Lex-0 and best practices in lexical data modelling using TEI have been introduced to more than 90 young scholars from across Europe at a number of face-to-face training events including the Lexical Data Masterclass in 2018, as well as the Lisbon Summer School in Linguistics in 2018 and 2019. The European Digital Humanities Masterclass 2020, which was going to include a TEI Lex-0 strand, had to be cancelled due to the Corona pandemic.

3.2. Ontolex-Lemon

Ontolex-Lemon is the de facto standard for representing lexical information as RDF. The model has been developed by the Ontolex Community Group (CG) of the World Wide Web Consortium to act as a model for the representation of lexical information in ontologies. The OntoLex-Lemon model was based on the Monnet-lemon model, which was designed as a standard for the extensions of ontologies with lexicographic information and was intended as a standard for the semantic web. However, the OntoLex CG recognised a wider use case of transforming traditional lexicographic resources to modern web standards. As such, the OntoLex-Lemon model plays a key role in mediating between lexicographic resources and state-of-the-art computational techniques, in particular natural language processing.

Within Ontolex-Lemon, a special module was developed to better deal with traditional dictionaries, the lexicography module. This module was published as a community recommendation on the 17th September 2019 and provides extensions that allow traditional lexicographic resources to be compatible with OntoLex's design that is intended to support NLP applications. In addition, the community group, with the assistance of ELEXIS project members, are developing two new modules. Firstly, the Frequency, Attestation and Corpus information (FrAC) module provides extensions that allow lexical resources in OntoLex to be connected to corpora. In this way, it is envisioned that these lexical resources will be more useful for a wide range of NLP applications and in particular support is

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

provided in this model for representations from modern NLP such as word embeddings. The second module being developed is the morphology module, which is providing both descriptive and generative representations of morphology within the lexicon. In this way, this module aims to cover traditional use cases of morphological lexicons, by recording the decomposition of forms into their morphs, while also enabling applications to generate forms at will, enabling NLP applications for morphologically-rich languages. The continued developments of these models are being enabled by regular public tele-conferences that are also being supported by the NexusLinguarum COST action.

3.3. DMLex - a new standardised data model for Lexicography

3.3.1 Why DMLex

As confirmed by the survey results, the lexicographic landscape in Europe is still rather heterogeneous. It is characterised by stand-alone lexicographic resources that are typically encoded in their own custom data format, i.e. proprietary XML, (customised) TEI, HTML, JSON-LD or are stored in a relational database.

In order to ensure semantic interoperability between these diverse dictionary structures, ELEXIS needed a common data model. Such a model was necessary to a) streamline the integration of lexicographic data into the ELEXIS infrastructure, b) to allow reliable linking of the data in the dictionary matrix, and c) to form a basic template for the creation of new lexicographic resources, such that they can automatically benefit from the tools and services provided by the ELEXIS infrastructure. This common data model is DMLex.

3.3.2 The design goals of DMLex

DMLex is a high-level data model for lexicographic data. It was designed with the following goals in mind:

- To provide interoperability between existing dictionary standards (in particular TEI Lex0 and Ontolex-Lemon) and various other, ad-hoc dictionary schemas. DMLex defines data types to which data types from other standards and schemas can be mapped.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

- To support a shift in lexicography from unstructured to structured data. In DMLex, a dictionary is understood not as a "document" but as a "dataset" consisting of objects (such as entries, headwords, senses, example sentences) which are connected to one another through relations.
- To simplify the data structures used in lexicography. DMLex is relatively sparse in that it defines only a small number of data types based on the common core vocabulary that was defined within ELEXIS (Tiberius et al. 2021). Some data types which are used in lexicography, such as subentry and subense, do not exist in DMLex. Instead, DMLex remodels these as relations between more basic data types such as entry and sense. The result is a refreshingly non-complex, IT-friendly data model.
- To provide guidance to IT professionals who want to represent a dictionary computationally, for example in a dictionary writing system or in other software that processes dictionaries.

3.3.3 An overview of DMLex

The data model defined in DMLex is abstract and independent of any specific implementation. In addition to that, DMLex provides four recommended implementations: one in XML, one in JSON, one as a relational database, and one as an RDF triplestore.

The DMLex standard consists of an obligatory **Core** followed by several optional **Modules**. The Core defines the most basic data types which every lexicographic resource is required to implement: entry, sense and so on, as well as relations between them. These core data types can be used to construct the basic entry-and-sense skeleton of any (monolingual) dictionary. The core also defines a mechanism for creating lists of controlled values (such as part-of-speech labels and usage labels) and mapping them to external ontologies, making it possible to represent the meaning of these labels in a machine-understandable way.

The DMLex **Crosslingual Module** extends the Core with data types needed for the representation of bilingual and multilingual lexicographic resources.

The DMLex **Linking Module** extends the Core with data types needed for representing various kinds of relations between senses and entries including morphosyntactic relations (spelling variants, gender pairs), semantic relations (synonymy, antonymy) and presentational relations (subsensing,

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

subentrying). The module also defines a mechanism for representing the meaning of these relations in a machine-understandable way.

The DMLex **Inline Markup Module** extends the Core with data types needed to represent various kinds of inline markup in headwords, definitions, example sentences and translations. It can be used to make the internal structure of multi-word headwords explicit, to mark up the occurrence of a headword in an example sentence or a definition, and also to mark up collocates and semantic roles in example sentences – again in a way which is machine-understandable.

3.3.4 Dissemination and impact of DMLex

The DMLex standard is being developed by the Lexicographic Infrastructure Data Model and API Technical Committee within the OASIS organisation. Once approved it will be published as an *OASIS Standard*. Additionally, the main authors are going to explain the reasoning and motivation of DMLex in a forthcoming peer-reviewed publication.

In summary, DMLex is a high-level, abstract formalism for the computational representation of dictionaries. It exists because, thanks in no small part to the ELEXIS project, lexicography is taking a turn towards being a deeply digitised, computational discipline. On the one hand, DMLex reflects the fact that this transformation is happening. On the other hand, it enables this transformation to happen.

4. ELEXIS tools to facilitate the creation and publishing of lexicographic data

Within ELEXIS a set of tools have been developed that facilitate the creation and publication of lexicographic data. The figure below represents the graphical guide to the ELEXIS dictionary tools in the LEX1 part of the infrastructure. In this part of the infrastructure, data can be transformed (using ELEXIFIER), created, edited, enriched and linked (using Lexonomy), and published (using Lexonomy or Publex).

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

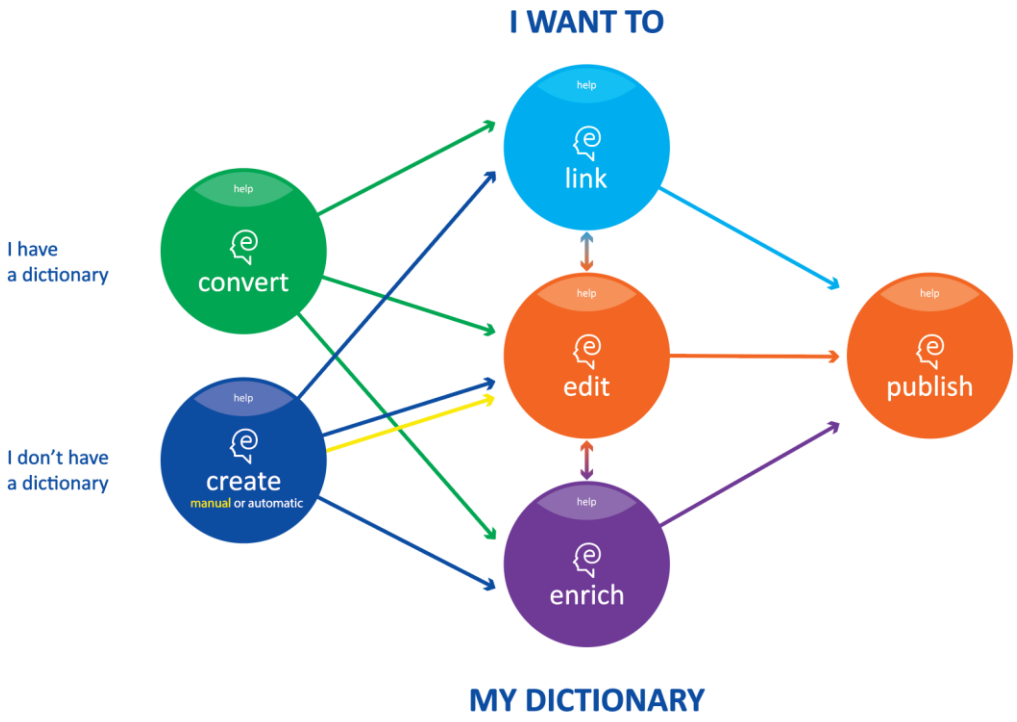


Figure 3: ELEXIFIER is part of the LEX1 infrastructure of ELEXIS and allows users to convert dictionaries into a standard format, and subsequently to link, edit, enrich and publish them

Below a short description of each of the tools is given.

4.1. Elexifier

Elexifier plays an important role in ELEXIS, specifically in terms of feeding data to the infrastructure. It allows lexicographers with limited computer programming skills to convert their legacy dictionaries into a standardised common format and upload them to the ELEXIS infrastructure. Elexifier is a cloud-based dictionary conversion service which takes as input an XML or PDF dictionary and produces a TEI Lex-0 compliant XML file. For more information on Elexifier, see D1.4 ELEXIS Conversion Tools and D1.3 Tools for the automatic segmentation and identification of lexicographic content. For more information on how to use Elexifier, see the course in the ELEXIS curriculum LEX3: Transforming Legacy Dictionaries using Elexifier.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

4.2. Lexonomy

Lexonomy is a cloud-based lexicographic tool and dictionary writing system. It can be used for creating dictionaries from scratch or for further editing dictionaries that have been transformed by Elexifier. Lexonomy also allows the user to publish the dictionary online. For more information on how to use Lexonomy, see the course in the ELEXIS curriculum Lexonomy: Mastering the ELEXIS Dictionary Writing System.

4.3. Publex

Publex is a software tool for publishing dictionary data which is annotated in XML. It allows the user to define the display of the dictionary individually. Publex is accessed and operated via a web browser, and the data is stored and published on the ELEXIS server. For more information on how to use Lexonomy, see the course in the ELEXIS curriculum LEX3: Publishing Legacy Dictionaries with Publex.

5. CLARIN

According to plans present already in the project proposal and Grant Agreement, the usage of existing CLARIN and DARIAH services was envisaged wherever possible. One of the cases is the use of the LINDAT repository as a part of the CLARIN.SI infrastructure, which is also aggregated in the CLARIN (ERIC) Virtual Language Observatory: https://vlo.clarin.eu/?1&q=elexis

After the end of the ELEXIS project, the tools and services that have been developed within ELEXIS will be hosted by CLARIN.SI infrastructure, and (some of the) activities started in the project will continue within the proposed ELEXIS Knowledge Centre for Lexicography as a part of CLARIN ERIC knowledge infrastructure (for more on ELEXIS-KC see deliverable D6.5 - Final ELEXIS interoperability report including interaction with CLARIN/DARIAH services).

Within the CLARIN.SI infrastructure, a special metadata collection was created to ensure consistency in the metadata that is associated with data that has been contributed to ELEXIS. See: https://www.clarin.si/repository/xmlui/handle/11356/1479. All the tools in the ELEXIS infrastructure pull the metadata from this collection.

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

6. Lexicographic Data Seal of Compliance

Within ELEXIS a proposal has been put forward for a Lexicographic Data Seal of Compliance (LexSeal), a community-based certificate of compliance with best scholarly practices to be awarded to individual lexicographic datasets in recognition of their creators' self-assessed and well-documented adherence to the principles of trustworthiness, interoperability, stewardship, citability, reciprocity and openness. This proposal has been described in D6.4 Lexicographic Data Seal of Compliance.

In order to achieve the widest possible community uptake, the LexSeal proposal will be submitted to CLARIN and DARIAH for consideration once the new proposed ELEXIS Association has been established. We will propose that the future Lex Seal Governing Board is made up of two members nominated by the DARIAH Board of Directors, two members nominated by the CLARIN Board of Directors and two members nominated by the ELEXIS Association.

7. Conclusion

With the end of the ELEXIS project, tools and services will be hosted by CLARIN.SI infrastructure, and (some of the) activities started in the project will continue within the proposed ELEXIS Knowledge Centre for Lexicography as a part of CLARIN ERIC knowledge infrastructure

Other activities will continue as a part of the new proposed ELEXIS Association which will be established after the project. For more on ELEXIS-KC and the ELEXIS Association see deliverable D6.5 - Final ELEXIS interoperability report including interaction with CLARIN/DARIAH services.

One of the important activities is the finalisation and official release of the DMLex data model in OASIS.

8. References

Boelhouwer, B., Kosem, I., Nimb, S., Jakubíček, M, Tiberius, C., Krek, S., Rosenmeier, M. (2020): ELEXIS deliverable 6.2 Recommendations on Legal and IPR Issues for Lexicography. https://elex.is/wp-content/uploads/2020/02/ELEXIS_D6_2_Reccommendations_on_Legal_and_IPR_Issues_for_Lexicography.pdf (last access: 25-07-2022).

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.

Kallas, J., Koeva, S., Kosem, I., Langemets, M., Tiberius, C. (2019a): ELEXIS deliverable 1.1 Lexicographic Practices in Europe: A Survey of User Needs. https://elex.is/wp-content/uploads/2020/06/Revised-ELEXIS_D1.1_Lexicographic_Practices_in_Europe_A_Survey_of_User_Needs.pdf (last access: 25-07-2022).

Kallas, J., Koeva, S., Langemets, M., Tiberius, C., Kosem, I. (2019b): Lexicographic Practices in Europe: Results of the ELEXIS Survey on User Needs. https://doi.org/10.5281/zenodo.3726841 (last access: 25-07-2022).

Kosem, I., Lew, R., Müller-Spitzer, C., Ribeiro Silveira, M., Wolfer, S. et al. (2019): The image of the monolingual dictionary across Europe. Results of the European survey of dictionary use and culture. In: International Journal of Lexicography, 32(1), pp. 92-114. https://doi.org/10.1093/ijl/ecy022 (last access: 25-07-2022).

Leroyer, P. and Køhler Simonsen, H. (2020): Reconceptualizing Lexicography: The Broad Understanding. In: Gavriilidou, Z., Mitsiaki, M., Fliatouras, A. (eds.): Proceedings of XIX EURALEX Congress: Lexicography for Inclusion, Vol. I, Democritus University of Thrace, pp. 183-192. https://euralex2020.gr/wp-content/uploads/2020/11/EURALEX2020_ProceedingsBook-p183-192.pdf (last access: 25-07-2022).

Rundell, M. (2015): From print to digital: Implications for dictionary policy and lexicographic conventions. In: Lexikos, 25(1), pp. 301-322. https://www.ajol.info/index.php/lex/article/view/126250 (last access: 25-07-2022).

Tiberius, C., Krek, S., Depuydt, K., Gantar, P., Kallas, J., Kosem, I., Rundell, M. (2021) Towards the ELEXIS data model: defining a common vocabulary for lexicographic resources. In: Kosem, I., Cukr, M., Jakubíček, M., Kallas, J., Krek, S. & Tiberius, C. (eds.) 2021. Electronic lexicography in the 21st century. Proceedings of the eLex 2021 conference. 5–7 July 2021, virtual. Brno: Lexical Computing CZ, s.r.o.

Tiberius, C., Kallas, J., Koeva, S., Langemets, M., Kosem, I. (2022). AN INSIGHT INTO LEXICOGRAPHIC PRACTICES IN EUROPE: Results of the extended ELEXIS Survey on User Needs. https://doi.org/10.5281/zenodo.6861747

This project received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Union.