



ELEXIS Standardisation

Simon Krek (Slovenia)

Jožef Stefan Institute (Artificial Intelligence Laboratory)



Standardisation: from the Grant Agreement

- ELEXIS aims to develop an infrastructure which will:
- Objective 3
 - develop strategies, tools and **standards** for extracting, **structuring** and linking of lexicographic resources
- Objective 4
 - enable **access to standards**, methods, lexicographic data and tools for scientific communities, industries and other stakeholders

Key Performance Indicators (KPI)

OBJECTIVE 2

Establish common standards and solutions for the development of lexicographic resources

Target outcome

Publication of
standardisation
documentation for
lexicographic resources

Indicator

Developed standard
included in ISO or OASIS
(end of project)

Target

≥ 1 new standard

Data is not available in standardised data models

- Therefore, although a huge number of lexicographic resources exist for various languages, as results of long-term projects in which a huge amount of work was dedicated to their compilation, they are typically encoded in **incompatible data structures**, and the **data** from these resources **is lost for** extensive, interoperable and generally accessible **computer use**.

ELEXIS platform architecture

- The existence of **common data models** and **standards** that are **produced bottom-up** from within the lexicographic community fostered by ELEXIS is a necessary condition for successful development of this segment of the platform.
- Standards will be **developed** and **tested during the project** on the data provided by the lexicographic partners and implemented in the newly developed service.

Partners with expertise in standardisation

- Standardisation partners are already active in standardisation efforts in case of existing standards, such as Text Encoding Initiative (Jožef Stefan Institute, Belgrade Centre for Digital Humanities, Dutch Language Institute)
- Lexical Mark-up Framework (Institute for Computational Linguistics «A. Zampolli»)
- or Ontolex-Lemon (National University of Ireland, Galway Austrian Academy of Sciences)
- Some of the partners are also directly involved in standardisation bodies, such as International Standards Organisation (TC37) (Universidade NOVA de Lisboa)
- or OASIS Open (Lexical Computing)

New standard: important factors

- Openness? Lexical markup framework (LMF) **580 CHF**
 - ISO 24613-1:2019 Part 1: **Core model** - CHF88, p13
 - ISO 24613-2:2020 Part 2: **Machine-readable dictionary (MRD) model** - CHF118, p21
 - ISO 24613-3:2021 Part 3: **Etymological extension** - CHF118, p22
 - ISO 24613-4:2021 Part 4: **TEI serialization** - CHF118, pages : p22
 - ISO 24613-5:2022 Part 5: **Lexical base exchange (LBX) serialization** - CHF138, p32
 - ISO/WD 24613-6 Part 6: **Syntax and Semantics**
- Suitability? Text Encoding Initiative (TEI) -> excellent for retro-digitised dictionaries, less so for born-digital
- Comprehensibility? Ontolex-Lemon -> hard to understand/handle for lexicographers

LEXIDMA (OASIS)

- Technical Committee: OASIS Lexicographic Infrastructure Data Model and API (LEXIDMA)
 - Developing an open standards based framework for internationally interoperable lexicographic work.
 - The LEXIDMA TC successfully held its first meeting on **16th December 2019** 1600 UTC. Mr. Tomaz Erjavec of the Josef Stefan Institute (IJS) was elected chair and Dr. David Filip of Trinity College Dublin (ADAPT) was elected Secretary. Seven members attended.
 - Participation in the OASIS Lexicographic Infrastructure Data Model and API (LEXIDMA) TC is open to all interested parties. Contact join@oasis-open.org for more information.
 - https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=lexidma



Coffee!

- See you in half an hour



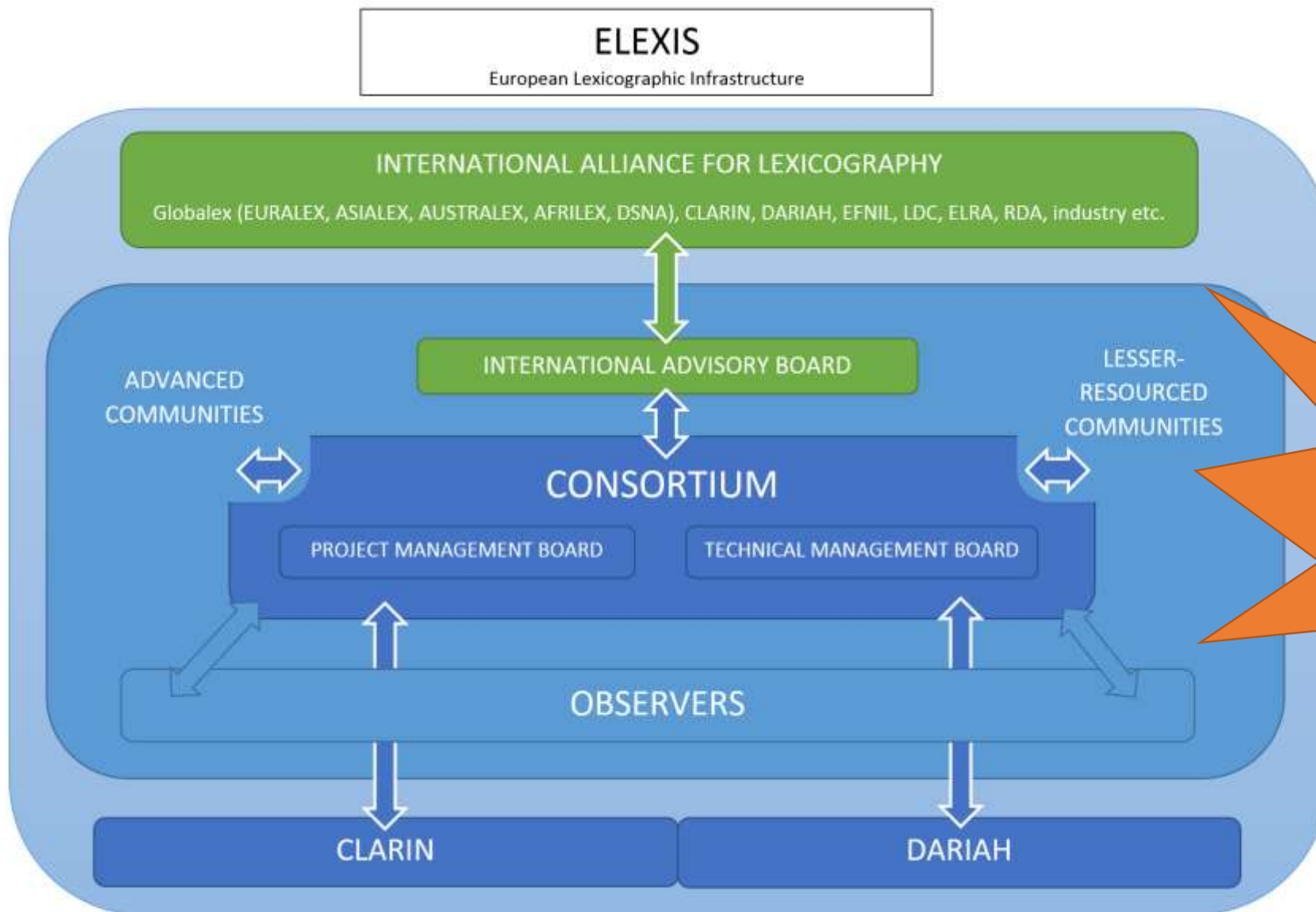


ELEXIS Sustainability

Simon Krek (Slovenia)

Jožef Stefan Institute (Artificial Intelligence Laboratory)





What will happen?

ELEXIS-KC (CLARIN Knowledge Centre for Lexicography)

- ELEXIS partners and observers mailing list: **13 April 2022**
- Invitation to join the application:
 - Please write to me if you are already sure you would like to participate in the initial Knowledge Centre application
 - If you are not sure if participation in the new Knowledge Center is interesting or viable for you, you can always join later.
- Application was sent to CLARIN Knowledge Infrastructure Committee
- 14 institutions, 46 experts
- **MORE ABOUT ELEXIS-KC TOMORROW!**

CLARIN Knowledge Centres (K-centres)

- are a cornerstone of the CLARIN Knowledge Infrastructure (KI), one of the main components ensuring a continuous transfer of knowledge between all players involved in the construction, operation and use of the infrastructure.
- The mission of the CLARIN KI is to ensure that the available knowledge and expertise does not exist as a fragmented collection of unconnected bits and pieces, but is made accessible in an organised way to both the CLARIN community and the social sciences and humanities research community more widely.

Administrative data

- Full name: ELEXIS - CLARIN Knowledge Centre for Lexicography
- Short name: ELEXIS-KC
- Host institution: Jožef Stefan Institute
- Place and country: Ljubljana, Slovenia
- Name of main contact(s): Simon Krek

What the K-centre intends to offer and to whom

- ELEXIS-KC offers the results of the ELEXIS (European Lexicographic Infrastructure) project, which has produced open access data, tools and services for lexicographers.
- These include: (1) **data**: semantically annotated corpora, extracted data from corpora (dictionary-on-the-fly), and lexicographic data that is being added to the CLARIN.SI repository, (2) **online services and tools**: Lexonomy, Elexifier, NAISC and BabelNet linker, gaming apps Cross the Word and Game of Words, and Dictionary Matrix.
- ELEXIS-KC maintains these services and provides help about their use.

Audience(s) targeted

- lexicographers
- language teachers
- linguists (working in the field of semantics, lexicology, lexicography, terminology)
- computational linguists (working on topics related to semantic technologies)
- corpus linguists
- digital humanists (working on historical lexicography)
- translators
- Language Technology and Natural Language Processing companies

Types of services offered

- Lexicographic tools
- Lexicographic data
- Assistance with the use of lexicographic data and tools
- Online training courses on lexicography
- Best practice documents on lexicography
- Consultancy on lexicography
- Standardisation related to lexicography

Quality, sustainability and organisation (1)

- ELEXIS-KC is supported by the following institutions:
- Jožef Stefan Institute (SLOVENIA, CLARIN.SI)
 - URL(s): <https://www.ijs.si/>, <http://www.clarin.si/info/about/>
 - expertise: Existing ELEXIS infrastructure, Helpdesk
- Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH), Austrian Academy of Sciences (AUSTRIA, CLARIAH-AT)
 - URL(s): <https://www.oeaw.ac.at/acdh/acdh-ch-home>
 - expertise: Existing ELEXIS infrastructure, lexicography, corpus linguistics, language resources and text technologies

Quality, sustainability and organisation (2)

- University of Copenhagen (DENMARK, CLARIN-DK)
 - URL(s): Centre for Language Technology: <https://cst.ku.dk/english/> Danish
CLARIN: <https://clarin.dk/clarindk/forside.jsp>
 - expertise: Computational lexicography, wordnets, lexical-semantic resources for NLP, practical lexical projects for NLP
- Danish Society for Language and Literature (DENMARK)
 - URL(s): <https://dsl.dk/>, <https://ordnet.dk/>
 - expertise: lexicography (historical and contemporary) for the Danish language, computational lexicography and the creation of formal semantic lexicons for Danish (WordNet, FrameNet etc.)

Quality, sustainability and organisation (3)

- Institute of Mathematics and Computer Science, University of Latvia (LATVIA, CLARIN-LV)
 - URL(s): clarin.lv, ailab.lv, tezaurs.lv
 - expertise: language resources and technology for Latvian, creation of WordNet-like and FrameNet-like lexical resources, computational lexicography, semantic parsing
- Vytautas Magnus University (LITHUANIA, CLARIN-LT)
 - URL(s): <http://clarin-lt.lt/?lang=en>
 - expertise: Lithuanian language technology, corpus linguistics, lexicography, terminology

Quality, sustainability and organisation (4)

- Dutch Language Institute (NETHERLANDS, CLARIAH-NL/CLARIAH-VL)
 - URL(s): <https://ivdnt.org/>; <https://portal.clarin.ivdnt.org/>;
<https://kdutch.ivdnt.org>
 - expertise: lexicography (historical and contemporary) for the Dutch language, terminology, corpus linguistics, language resources
- Institute for the Estonian language (ESTONIA)
 - URL(s): <https://portaal.eki.ee/>
 - expertise: lexicography, corpus lexicography, (multilingual) lexicographic resources for Estonian, corpus linguistics

Quality, sustainability and organisation (5)

- Institut za hrvatski jezik i jezikoslovlje (CROATIA)
 - URL(s): <http://ihjj.hr/>
 - expertise: lexicography, historical lexicography, etymology, semantics, terminology, general linguistics
- NOVA CLUNL (PORTUGAL)
 - URL(s): www.clunl.fcsh.unl.pt
 - expertise: lexicography, terminology, standards, computational lexical semantics

Quality, sustainability and organisation (6)

- The National Library of Norway (NORWAY)
 - URL(s): <https://www.nb.no/sprakbanken/en/resource-catalogue/>,
<https://www.nb.no/dh-lab/>
 - expertise: computational lexical semantics, Natural Language Processing, digital resources for written and spoken modern Norwegian and for medieval Norwegian Latin
- Università Cattolica del Sacro Cuore (ITALY)
 - URL(s): <https://centridiricerca.unicatt.it/circse-home?rdeLocaleAttr=en>
 - expertise: Digital Lexical and Textual Resources for Latin; Linguistic Linked Open Data; Natural Language Processing; valency and subcategorization

Quality, sustainability and organisation (7)

- University of Trier/Trier Center for Digital Humanities
 - URL(s): <https://tcdh.uni-trier.de/en>, www.woerterbuchnetz.de
 - expertise: (retro)digitization, XML-markup, interlinking and online publication of dictionaries, federated content search, digital humanities
- Institute of Philology of Borys Grinchenko Kyiv University (UKRAINE)
 - URL(s): <https://partner.kubg.edu.ua/2012-02-09-16-42-36/institute-of-philology/65-institute-of-philology/557-institute-chairs.html#department-of-romance-philology-and-comparative-typological-linguistics>
 - expertise: digital neography, digital terminography, linguistic innovation modelling, corpus linguistics, digital humanities, digital learning

Important takeaways

- New tools and services that were developed in ELEXIS will be maintained by the CLARIN.SI B-Centre in the foreseeable future (possibly in combination with other CLARIN centres in ELEXIS-KC)
- The initial set of 14 institutions in ELEXIS-KC can be expanded in the future if there will be interest
- The establishment, organisation and maintenance of ELEXIS-KC will follow regular CLARIN ERIC rules, regulations and practices
- We will try to find further funding through regular CLARIN (or ESFRI) channels

Other options

- A **European Cooperative Society (SCE)** is an optional legal form of a cooperative. It aims to facilitate cooperatives' cross-border and trans-national activities.
 - Example: <https://read.transkribus.eu/coop/>
 - Transkribus is a comprehensive platform for the digitisation, AI-powered text recognition, transcription and searching of historical documents
- Luxembourgish **association sans but lucratif**
 - Model: EFNIL (European Federation of National Institutions for Language).
- Offer from the **European Science Foundation (ESF)**
 - The European Science Foundation (ESF) is a non-governmental, internationally-oriented, non-profit association established in France in 1974. ESF is committed to promoting the highest quality science in Europe to drive progress in research and innovation.

ELEXIS Association

- JSI prepared a light-weight „partnership agreement“
- It defines the basics: the aims, governance, administration
- Interesting for those who would like to:
 - be actively involved in further development of (lexicographic) data, tools and services (therefore not only providing knowledge, as in the case of ELEXIS-KC)
 - search for funding options
 - general collaboration

Article 1 - Preliminary observations

- The Parties to this Agreement initially note that by concluding this Agreement, the Parties enter into a contractual partnership called EUROPEAN LEXICOGRAPHIC INFRASTRUCTURE Association or ELEXIS Association.

Article 2 - Objectives and activities

- The objective of the ELEXIS Association is organisation and coordination of activities related to lexicography, and activities related to natural language processing tasks on the topic of semantics, insofar they are of interest to lexicography.
 - To this end, the ELEXIS Association seeks to engage in:
 - the development of lexicographic tools and (web) services,
 - the exchange and/or linking of lexicographic data,
 - the development of lexicographic standards facilitating the interoperability of lexicographic data,
 - scientific research on lexicography and semantics, including the organisation of conferences, the engagement in research projects and similar,
 - the exchange of expertise including the organisation of training, research visits and similar,
 - all other activities which contribute to its aims.

Article 3 - Services and activities

- The detailed scope of services and/or activities, in line with the objectives in Article 2, including possible financial arrangements and allocation of ownership and terms of exercising, protection and exploitation of results, must be agreed between the Parties of this Agreement and specified in a separate written document.

Article 4 - Board of Representatives

- The Parties shall establish the Board of Representatives, within thirty (30) days after the conclusion of this Agreement.
- The Board of Representatives shall be composed of one (1) authorised representative of each Party.
- Parties appoint their representatives in accordance with and in the manner determined by their internal affairs acts.

The Board of Representatives (1)

- The Board of Representatives advises and decides on important matters relating to this Agreement, ELEXIS Association and its activities.
- It decides on activities, fees and other contributions, establishes the text of the Agreement, accepts or rejects new candidate Members, establishes the action lines and performs other activities.
- The Board of Representatives shall be chaired by the President of the Board.
- The Board of Representatives selects the President among its members by a majority of votes, for a two-year period.

The Board of Representatives (2)

- The President selects a Deputy among the members of the Board of Representatives who takes the duties of the President in her/his absence.
- It is called by the President to convene the Board of Representatives at least once a year.
- All Members must receive at least four (4) weeks' advance written notice of Board of Representatives session.
- Board of Representatives decisions are based on a simple majority of the votes of the representatives. The session is quorate if the majority of representatives the Board of Representatives are present.

Article 5 - Membership

- An institution wishing to become a Party to this Agreement and a Member of the ELEXIS Association must apply for membership by sending an application letter to the Coordinator.
- The admission of the candidate institution is decided by the Board of Representatives by a majority vote, not later than three (3) months after the receipt of application.

Article 6 - Coordinator

- The Coordinator shall have the following functions:
 - keeping the address list of Members and other contact persons updated and available;
 - administration and provision of the President of the Board of Representatives and follow-up of Board's decisions;
 - transmission of any relevant documents and information connected with this Agreement;
 - any other duties that are agreed on between the Board of Representatives and the Coordinator.

Article 7 - Administrative fee

- The Coordinator shall be entitled to receive an annual administrative fee from each Party of this Agreement for performing the activities as set in Article 6.
- The annual administrative fee is **XXX** EUR.
- The Coordinator will issue a separate invoice to each Party of this Agreement.

Article 8

- Nothing in this Agreement shall constitute or be deemed to constitute any formal business organisation or legal entity between the Parties.
- Each Party shall act as independent contractor and not as the agent of any of the other Parties.

Results





Coffee!

- See you in half an hour

