

European Lexicographic Infrastructure (ELEXIS)

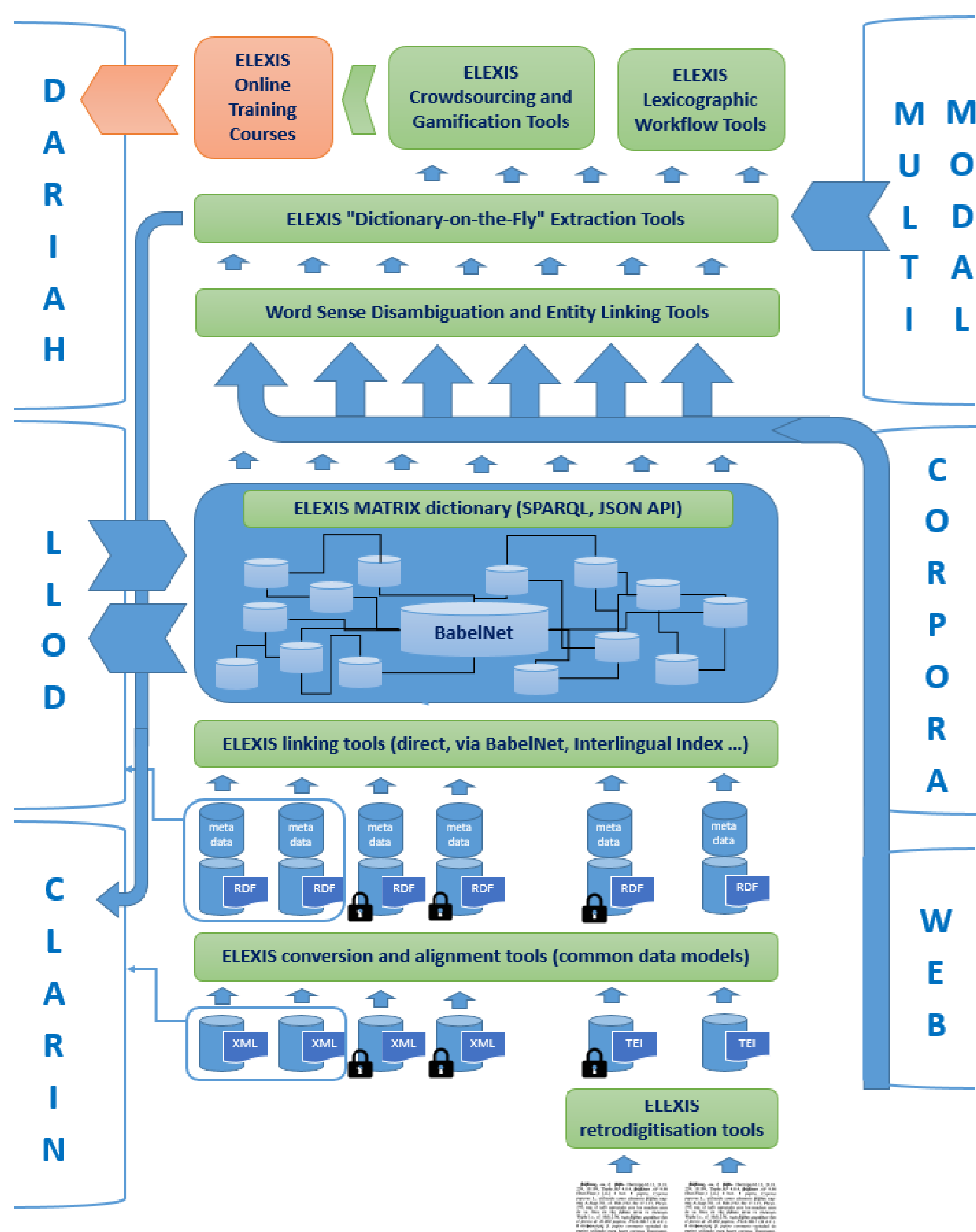
Simon Krek¹, Iztok Kosem¹, John P. McCrae², Roberto Navigli⁵,
Bolette S. Pedersen⁶, Carole Tiberius⁴, Tanja Wissik³

Jožef Stefan Institute, Slovenia,¹ Insight Centre for Data Analytics, NUIG, Ireland,² Austrian Academy of Sciences, Austria,³
Dutch Language Institute, Netherlands,⁴ Sapienza University of Rome, Italy,⁵ University of Copenhagen, Denmark⁶

ELEXIS infrastructure

ELEXIS “virtual access” infrastructure – providing online access to data, tools and services – will consist of three sub-infrastructures: **LEX1**, **LEX2**, **LEX3**.

www.elex.is



LEX1: is dedicated to automatic **segmentation** and **structuring** of content for dictionaries that are currently produced in digital environments but are typically encoded in their own custom data format. **Conversion** and **alignment** tools provide users of the infrastructure with the possibility to harmonise and convert their lexicographic resources to a uniform data format that allows their integration in Linked Open Data. To provide conceptual interoperability, services enabling **linking** of lexicographic resources will be developed and made available in the linking tools segment of the platform. This will provide the possibility to link lexical entries, senses and fundamental concepts in different lexical resources.

LEX2: Based on the contribution of lexicographic data, a new infrastructure will be developed that includes **word sense disambiguation** and **entity linking** tools dedicated to semantic processing of corpus data. These tools will have an important impact on disambiguation and corpus analysis, and will open up the possibility to create lexicographic data from corpora in a fully automated process. This is included in the »dictionary-on-the-fly« segment of the platform. The service will be able to produce a proto-dictionary with sense distribution, extracted definitions, collocations, multi-word expressions, (good dictionary) examples, translation equivalents and data in other modalities.

LEX3: The third set of services is dedicated to **retrodigitised dictionaries** in the part of the platform that includes (1) tools for automatic segmentation and structuring of content, and (2) an online publication tool for retrodigitised dictionaries.

TRANS-NATIONAL ACCESS: During the lifetime of the project ELEXIS will organise calls enabling researchers (a) to work with data at host institutions; (b) to gain knowledge and expertise in close contact with lexicographers and experts in NLP and artificial intelligence.

ABOUT: ELEXIS is a new EU infrastructure project dedicated to lexicography. The project is part of the Horizon 2020 programme, with a duration of four years (2018-2022). The result of the project will be an infrastructure which will (1) enable efficient access to high quality lexicographic data, and (2) bridge the gap between more advanced and lesser-resourced scholarly communities working on lexicographic resources.

OBSERVERS: Observer status will be granted to research institutions or other legal entities that are involved in lexicography or are active in the fields connected with the goals defined in ELEXIS infrastructure project – natural language processing, artificial intelligence etc.



Institutions wishing to become ELEXIS observers can check the web site and fill in the request form for observer status.

ELEXIS organisation

