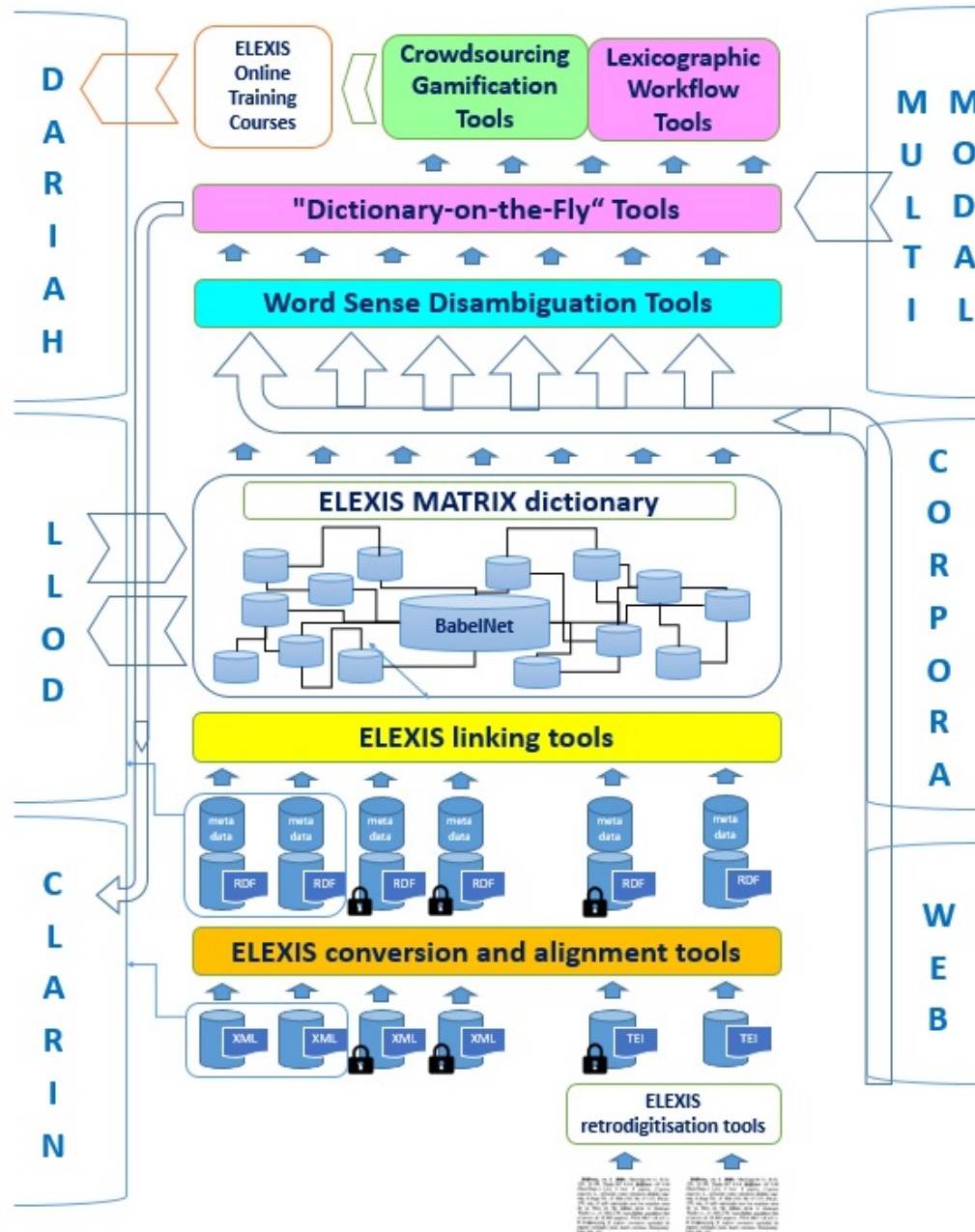




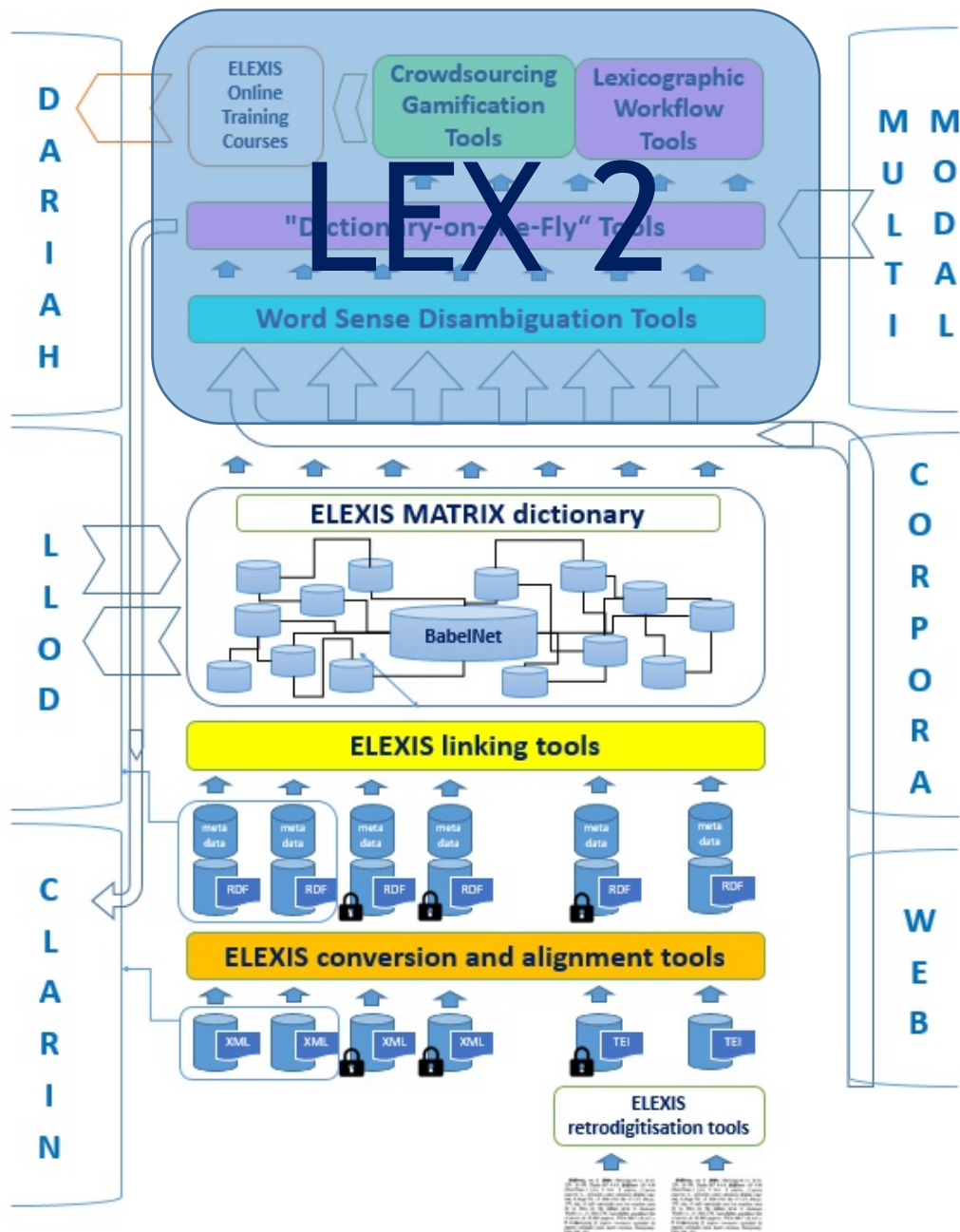
LEX 1 & LEX 3

Simon Krek (Jožef Stefan Institute, Ljubljana, Slovenia)

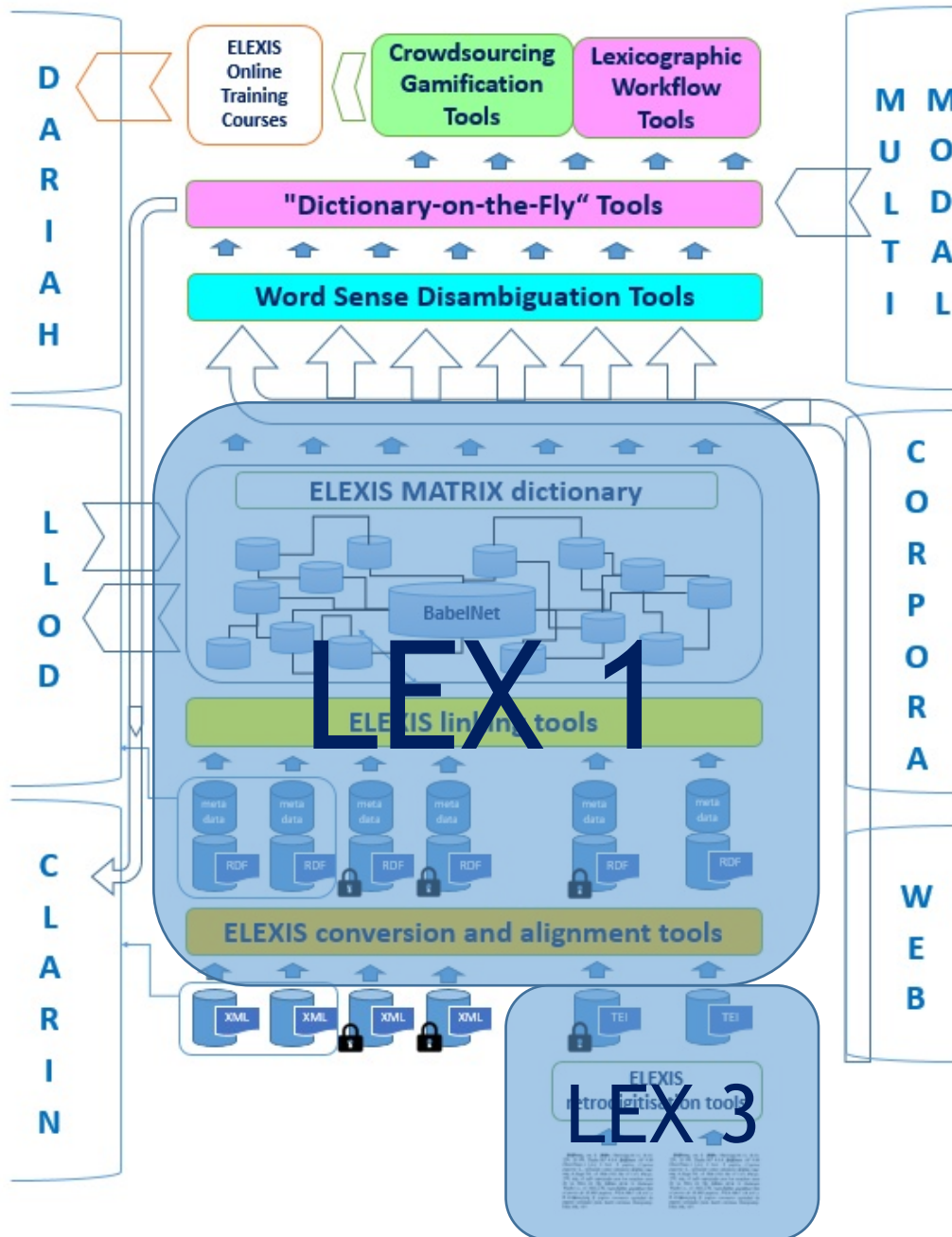




VIRTUAL ACCESS



VIRTUAL ACCESS



VIRTUAL ACCESS

LEX 1 – segmentation, conversion tools

- The **LEX1 infrastructure** will be set up during the project and will consist of several sets of tools:
- (1) tools for automatic **segmentation and structuring** of lexicographic content for dictionaries that are currently produced in digital environments and are encoded in their own custom data format.
- (2) **conversion and alignment tools** that will provide users of the infrastructure with the possibility to harmonise and convert their lexicographic resources to a uniform data format that allows their seamless integration in Linked Open Data.

LEX 1 – linking tools

- **(3) linking tools** that will provide conceptual interoperability, enabling linking of ELEXIS lexicographic resources.
- This will provide the possibility to **link lexical entries, senses** and fundamental **concepts** in different lexical resources, using a semi-automatic approach.
- **BabelNet**, as an existing multilingual resource to provide cross-lingual linking, will be exploited for this purpose.

LEX 1 – matrix dictionary = dictionary matrix

- (4) **ELEXIS matrix dictionary**: extensive linking of existing lexicographic resources by pivoting through BabelNet will enable the creation of what we call ELEXIS matrix dictionary -
- a universal **repository of linked senses**, meaning descriptions, etymological data, collocations, phraseology, translation equivalents, examples of usage and all other types of lexical information found in all types of existing lexicographic resources, monolingual, multilingual, modern, historical etc., available through a RESTful web service as **part of LEX1**.
- Freshly renamed: **DICTIONARY MATRIX - a matrix of dictionaries**

LEX 3 infrastructure

- LEX3 infrastructure will be dedicated to retrodigitised dictionaries and will include
- (1) tools for automatic **segmentation and structuring** of content in retro-digitised dictionaries, and
- (2) **Legacy Dictionary Viewer**: a generic, modular dictionary publication tool for retrodigitised dictionaries which offers interfaces for the analysis and profiling of the underlying lexical data.
 - The viewer will allow users – lexicographers and digital humanists alike – to publish user-friendly, online editions or retrodigitised dictionaries without facing extensive technical or financial obstacles.

Segmentation and structuring

LEX 1 scenario: digital, unstructured

LEX 3 scenario: retro-digitised

dictatorial

197

diffuse

dictatorial /dɪktətəˈrɪəl/. ADJ Kdor je **dictatorial**, svojo oblast uporablja preveč nasilno. ♦ diktatorski, tiranski; ...*dictatorial regimes*. ...*their dictatorial attitude*.

dictatorship /dɪkteɪtəʃɪp/, dictatorships. **1** UNCOUNT N **Dictatorship** je vladavina diktatorja. ♦ diktatura; *Democracy soon gave way to dictatorship*. **2** COUNT N **A dictatorship** je država, ki ji vlada diktator. ♦ diktatorska država

diction /dɪkʃən/. UNCOUNT N **One's diction** je jasnost, s katero nekdo govori ali poje. ♦ način izražanja, izgovorjava, dikcija

dictionary /dɪkʃənəˈrɪ/, dictionaries. COUNT N **A dictionary** je knjiga, v kateri je seznam besed nekega jezika po abecednem redu, poleg njih pa so navedeni pomeni ali ustrezne besede v drugem jeziku. ♦ slovar; ...*an English-French dictionary*.

did /dɪd/. **Did** je preteklik glagola **TO DO**. ♦ storiti

didactic /dɪˈdæktɪk/. ADJ Kar je **didactic**, naj bi nas nekaj naučilo; knjižna raba. ♦ poučen

differ from the early ones in many ways. **2** VB Kadar ljudje **differ about something**, se o nečem ne strinjajo med seboj. ♦ ne strinjati se; *We differ about moral standards*.

difference /dɪfəˈrəns/, differences. **1** COUNT N or UNCOUNT N **The difference between things** je način, po katerem se stvari med seboj razlikujejo. ♦ razlika, razloček; *There is an essential difference between computers and humans. Look at their difference in size*. **2** SING N **The difference between two amounts** je količina, za katero je en znesek manjši ali večji od drugega. ♦ razlika **3** COUNT N Kadar ljudje **have their differences**, se z nečim ne strinjajo. ♦ nestrinjanje, spor **4** PHRASE Kadar nekaj **makes a difference**, spremeni neki položaj. ♦ biti pomembno; *It makes no difference whether he is a citizen or not*.

different /dɪfəˈrənt/. **1** ADJ Če je ena stvar **different from another**, je v določeni lastnosti drugačna od druge stvari. ♦ razlikovati se; *The meeting was different from any that had gone before. His message is very different to theirs*.

ABERDEEN [*aberdin'*], v. d'Ecosse, ch.-l. de comté; port sur la mer du Nord; 170.000 h. Université.

ABERDEEN (G. H. Gordon, *comte d'*), homme d'Etat anglais, né à Edimbourg. Premier ministre en 1852, il conclut avec la France une alliance contre la Russie (1784-1860).

ABER-VRACH, fl. côtier du Finistère (Atlantique); 34 kil. Station marémotrice d'essai.

ABGAR, nom de huit rois d'Edesse, en Mésopotamie (132 av. J.-C.-216 apr.).

ABIA, roi de Juda, fils de Roboam, vainqueur de Jéroboam, roi d'Israël (957-955 av. J.-C.).

ABIDJAN, ch.-l. de la Côte-d'Ivoire (A.-O. F.), sur une vaste lagune navigable; 15.000 h.

ABIMÉLECH [*lèk*], fils de Gédéon. Il devint Juge d'Israël, après avoir fait égorger ses frères; il établit son pouvoir sur Sichem et fut tué au siège de Thèbes, en Palestine (vers 1100 av. J.-C.).

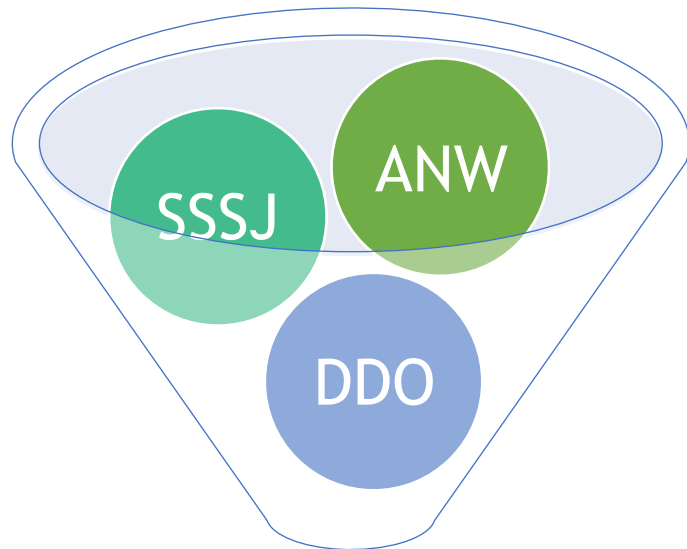
ABIRON, lévite qui fut englouti dans la terre avec Coré et Dathan, tous trois révoltés contre Moïse et Aaron (*Bible*).

GROBID-dictionaries

- GROBID-Dictionaries is a GROBID sub-module, implementing a java machine learning library, for **structuring digitized lexical resources** and entry-based documents with encyclopedic or bibliographic content.
- It allows the parsing, extraction and structuring of text information in such resources.
- Presentation: Mohamed Khemakhem
- On GitHub: <https://github.com/MedKhem/grobid-dictionaries>
- ELEXIS installation: <http://grobid.elex.is/>

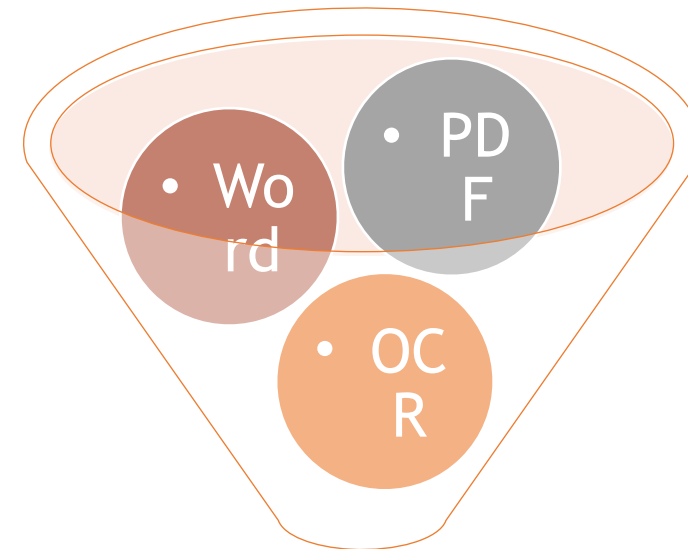
Conversion and alignment

ELEXIS (XML schema) conversion tools



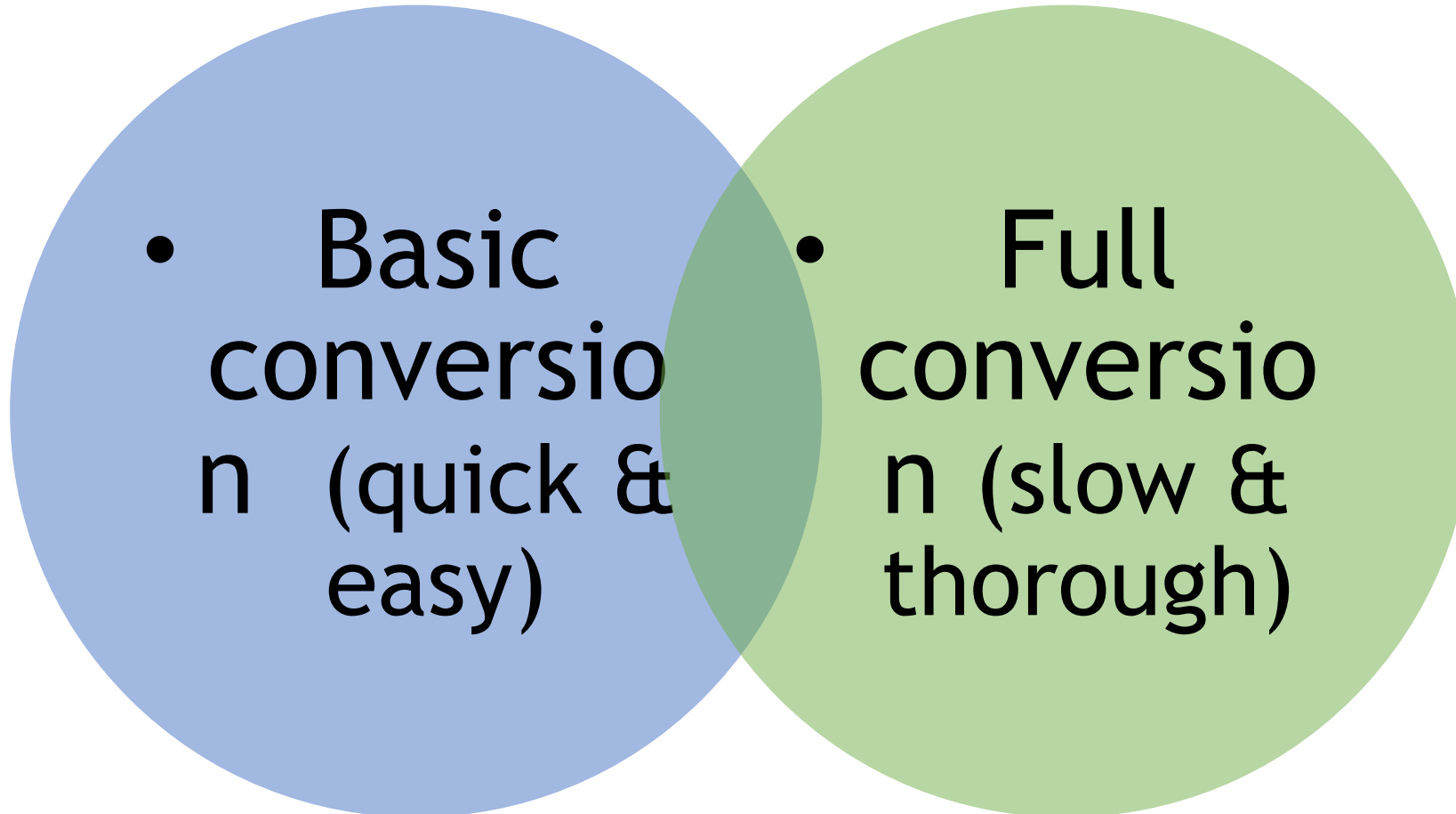
TEI-
Lex0

GROBID-dictionaries



• TEI-
Lex0

Conversion and alignment – TEI-Lex0



Conversion tools - process

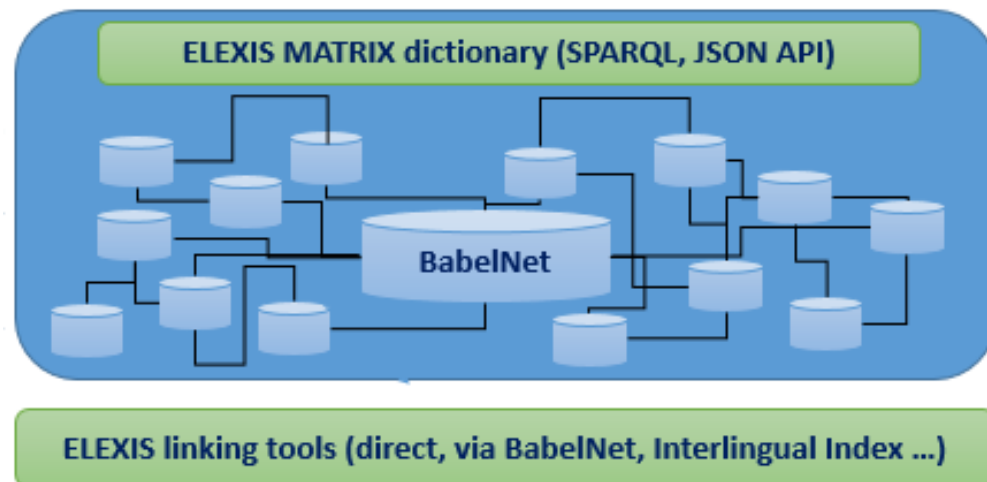
- Quick and easy
 - Upload your dictionary in custom XML format -> parse (or provide schema)
 - **Identify basic dictionary elements defined by ELEXIS data model**
 - Convert your dictionary to TEI-Lex0 and download
- Slow and thorough
 - Upload your dictionary in custom XML format -> parse (or provide schema)
 - **Align your schema and TEI-Lex0**
 - Convert your dictionary to TEI-Lex0 and download

Conversion and alignment tools

- **(E)LEXIFIER**

- *Lexifier - the dominant language of a pidgin or creole that serves as the basis for most of its vocabulary (Wiktionary)*
- The tool identifies only basic dictionary elements (entry, headword, sense, definition, example etc.) and produces a file in valid TEI-Lex0 format with original dictionary content.
- XSLT tools for “complete” conversion

ELEXIS DICTIONARY MATRIX



fancy² noun
 Menu

fancy plural *fancies*

1 **LIKING/WISH** [singular] especially British English

a) a feeling, especially one that is not particularly strong or urgent, that you like someone or want to have something

take a fancy to somebody/something (=decide that you like someone

or want to have so:

Mr Hill took a

Wanting to go

not last long).

Because of its

the fancy takes y

b) take/catch

you like it or want:

Did you see an

2 tickle somebody

somebody:

The idea of playing a joke on her tickled his fancy.

3 **IDEA** [countable] old-fashioned an idea or opinion that is not based on fact:

Oh, that was just a fancy of his.

4 **IMAGINATION** [uncountable] literary imagination or something that you imagine → flight of fancy at FLIGHT (6)

Noun

- **S: (n) illusion, fantasy, phantasy, fancy** (something many people believe that is false) "they have the illusion that I am very wealthy"
- **S: (n) fancy** (imagination or fantasy; held by Coleridge to be more casual and superficial than true imagination) "never had the wildest flights of fancy imagined such magnificence"
- **S: (n) fondness, fancy, partiality** (a predisposition to like something) "he had a fondness for whiskey"; "she had dismissed him quite brutally, relegating him to the status of a passing fancy, or less"

BabelNet – five concepts FANCY-n



- Noun
- Verb
- Adjective

fancy ENGLISH TRANSLATE INTO...

All Concepts Named Entities 8 concepts

Noun

fancy, fantasy, illusion
 Something many people believe that is false
 ID: 00033616n | Concept

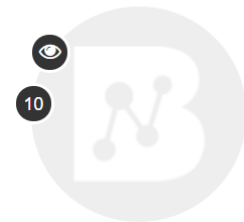
fancy
 A kind of imagination that was held by Coleridge to be more casual and superficial than true imagination
 ID: 00033617n | Concept

fancy, fondness, partiality
 A predisposition to like something
 ID: 00033618n | Concept

Animal fancy, Fancy (animal breeding)
 Animal fancy is a hobby involving the appreciation, promotion, or breeding of pet or domestic animals.
 ID: 02436504n | Concept

Fancy (surname)
 Fancy is the surname of:
 ID: 17409395n | Concept

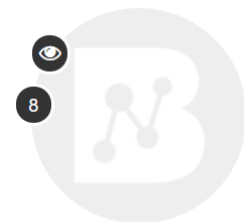
Noun



fancy, fantasy, illusion

Something many people believe that is false

ID: 00033616n | Concept



fancy

A kind of imagination that was held by Coleridge to be more casual and superficial than true imagination

ID: 00033617n | Concept



fancy, fondness, partiality

A predisposition to like something

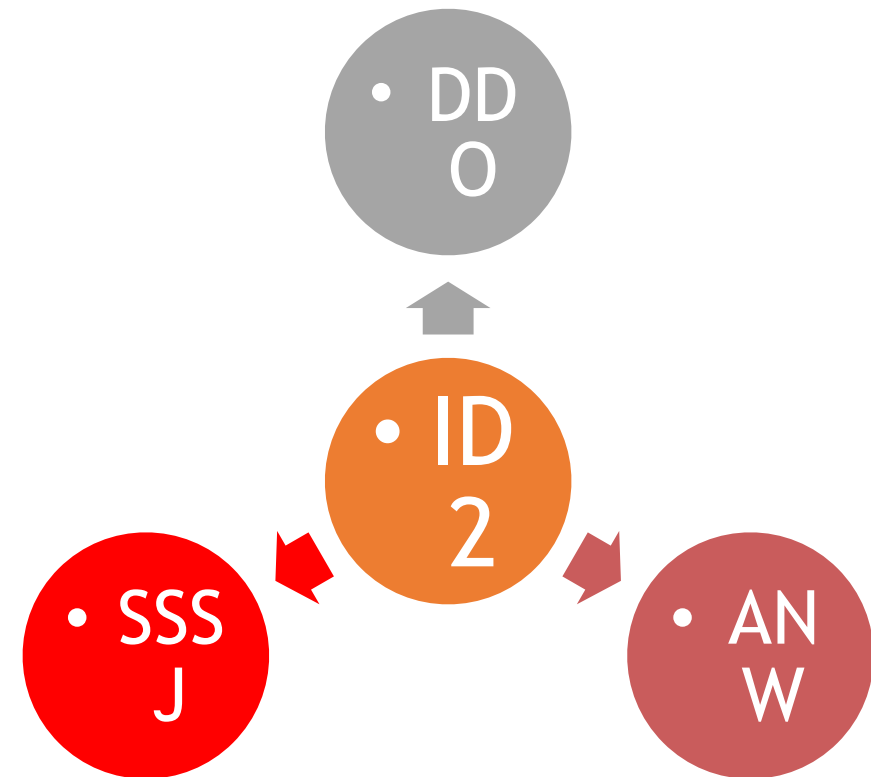
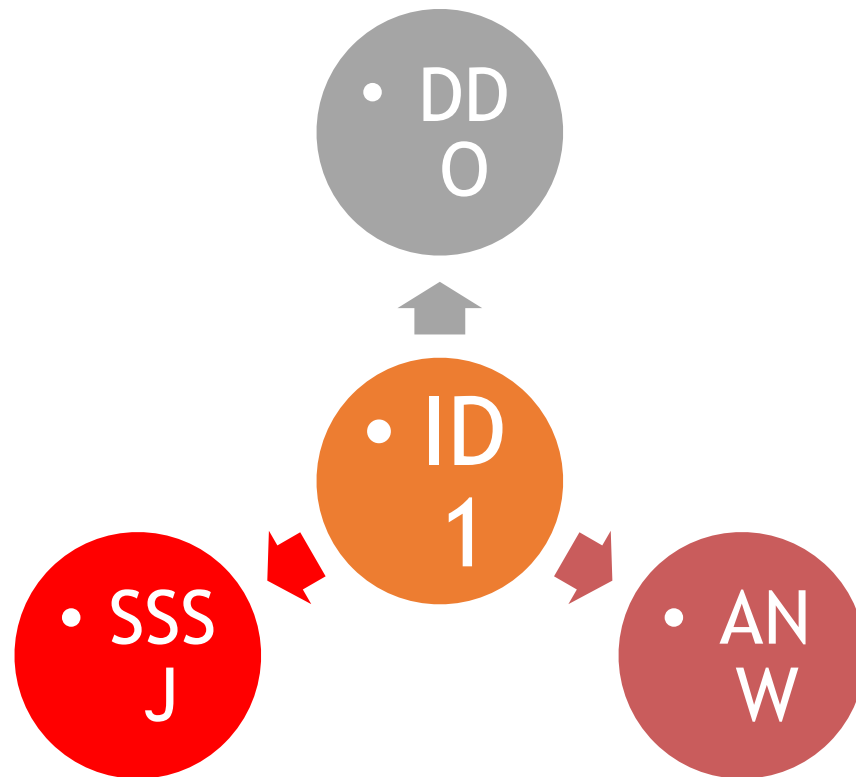
ID: 00033618n | Concept



Repository of senses – “matrix dictionary”

CONCEPT 1 - “FANCY-n”: something many people believe is false

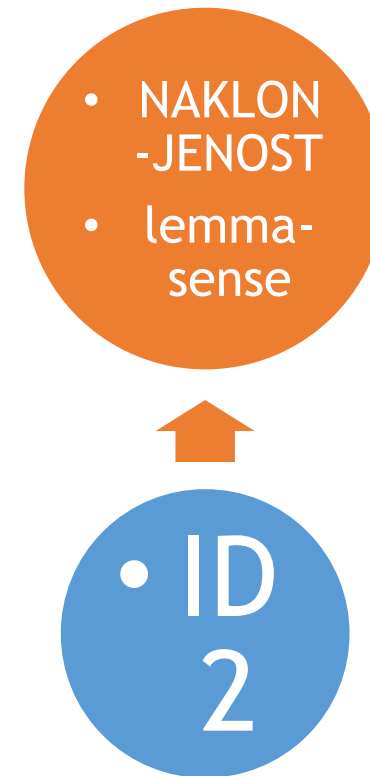
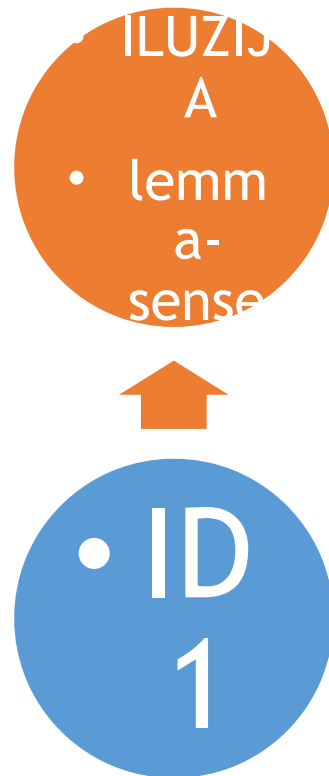
CONCEPT 2 “FANCY-n” CONCEPT 2: a predisposition to like something



Repository of senses - SloWNet

CONCEPT 1 - “FANCY-n”: something many people believe is false

CONCEPT 2 “FANCY-n” CONCEPT 2:
a predisposition to like something



EDM visibility

European Dictionary Portal (EDP)

Home Find a dictionary About English ▼

dictiO:naryportal.eu

Find a dictionary

Albanian (1)

Basque (2)

Breton
General dictionaries (1) · Historical dictionaries (2)

British Sign Language (1)

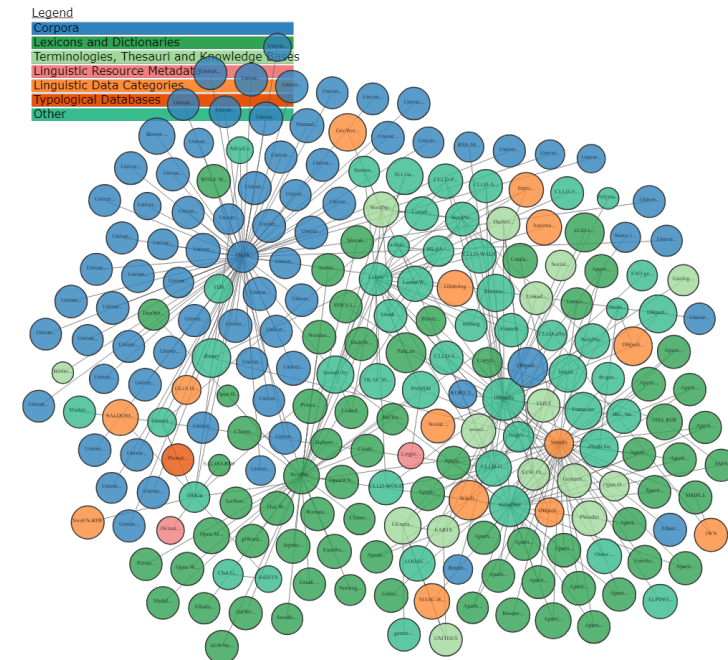
Bulgarian
General dictionaries (1) · Etymological dictionaries (1) · Historical dictionaries (1) · Dictionaries on special topics (1)

Filter

OBJECT LANGUAGE
(any language) ▼
The language the dictionary describes; the language the dictionary is about.

METALANGUAGE
(any language) ▼
The language in which the dictionary gives definitions and explanations. If same as object language, then the dictionary is monolingual. If different, then the dictionary is bilingual.

Linguistic Linked Open Data cloud (LLOD)



The Linguistic Linked Open Data Cloud from lod-cloud.net



Presentations

- GROBID-Dictionaries
- ELEXIS data model
- Linking & dictionary matrix
- Word Sense Disambiguation
- **Mohamed Khemakhem**
 - Inria-ALMAAnaCH
- **Carole Tiberius**
 - Dutch Language Institute
- **John McCrae**
 - National University of Ireland
- **Roberto Navigli**
 - Sapienza University of Rome



Timeline – LEX1 / 3

- **T1.4 Data Structuring for existing resources: M1-M24**
 - GROBID-dictionaries, ELEXIFIER
 - D1.3 Tools for the automatic segmentation and identification of lexicographic content (**M24**) (software)
 - D1.4 ELEXIS conversion tools (**M36**) (software)
- **T2.2 Semi-automatic linking of lexical resources: 12M-48M**
 - just started → end of project
 - D2.2 Interoperable interface for Lemon and TEI resources (**M24**) (software)
 - D2.4 Cross-lingual Lexical Resource Linking Web Service (**M42**) (software)