

5.2 GUIDELINES FOR PRODUCING ELEXIS TUTORIALS AND INSTRUCTION MANUALS

Author(s): Toma Tasovac, Rute Costa,
Fahad Khan, Iztok Kosem, John
McCrae, Monica Monachini, Ondřej
Matuška, Snežana Petrović,
Christophe Roche, Carole Tiberius,
Tanja Wissik

Date: 31.01.2020



H2020-INFRAIA-2016-2017

Grant Agreement No. 731015

ELEXIS - European Lexicographic Infrastructure

5.2 GUIDELINES FOR PRODUCING ELEXIS TUTORIALS AND INSTRUCTION MANUALS

Deliverable Number:	5.2
Dissemination Level:	Public
Delivery Date:	31.01.2020
Version:	1.0
Author(s):	Toma Tasovac Rute Costa Fahad Khan Iztok Kosem John McCrae Monica Monachini Ondřej Matuška Snežana Petrović Christophe Roche Carole Tiberius Tanja Wissik

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
---------	------	------------------	-----------------------

Table of Contents

Introduction.....	1
ELEXIS Curriculum.....	2
Scope and rationale	2
Target Audience.....	3
Learning Objectives, Outcomes, Workflows and Scenarios.....	3
Quality Assurance	4
Courses.....	5
Introduction to Dictionaries.....	5
Introduction to Dictionary Users.....	5
Introduction to Corpus-Based Lexicographic Practice.....	6
Capturing, Modeling and Transforming Lexical Data: An Introduction.....	6
LEX2: Mastering ELEXIS Corpus Tools for Lexicographic Purposes.....	6
Lexonomy: Mastering the ELEXIS Dictionary Writing System	7
Automating the Process of Dictionary Creation.....	7
CLARIN Tools and Resources for Lexicographic Work	7
Standards for Representing Lexical Data: An Overview.....	8
Modeling Dictionaries in TEI Lex-0	8
Modeling Dictionaries in OntoLex-Lemon	8
LEX3: Mastering ELEXIS Tools for Legacy Dictionaries.....	9
Lexicography in the Age of Open Data	9
Mastering LEX1: The Dictionary Matrix	9
Mastering oXygen XML Editor for Dictionary Nerds.....	10
Extracting Lexical Data: XPath for Dictionary Nerds	10
Transforming Lexical Data: XSLT for Dictionary Nerds.....	10
Trends in Computational Lexicography: A Pathfinder for Developers.....	11
Hosting platform: DARIAH-Campus.....	12
Rationale.....	12
Figures.....	14

Introduction

This report builds upon ELEXIS D5.1 *Skillset Report*, which provided a general assessment of the skills required for the active participation in the ELEXIS research infrastructure. Based on in-depth video interviews with 11 lexicographers from across Europe, both from within and beyond the ELEXIS Consortium, the *ELEXIS Skillset Report* highlighted the importance of skills development for lexicographers at two different levels:

- a) *generally speaking*, in a field, which is marked by fast-paced technological developments, continued skills development for lexicographers was seen as an essential contribution toward closing the gap between better and lesser resourced institutions, communities and languages; and
- b) *specifically*, in the context of ELEXIS as both a knowledge network and a service provider, training and education was seen as essential contribution toward attracting a diverse userbase and securing the social sustainability of the emerging lexicographic infrastructure.

It is beyond the scope of the current report to summarize *all* the findings of ELEXIS D5.1. It is however, especially important to highlight three high-level principles that have guided our work since the completion of D5.1:

1. to counter the lack of university curricula in lexicography or systemic, integrated training opportunities outside the university, ELEXIS should develop a coherent, integrated curriculum which will be based on a healthy balance of theory and practice;
2. in order not to leave anybody behind, the ELEXIS training materials should not assume any previous knowledge and should cover the very basics of lexicographic theory and practice. Or, as one informant put it: „Do not assume that anybody knows anything. Think how low you can go.”
3. ELEXIS should become a central hub or reference point for collecting and developing openly accessible training materials through a “specific website” or a “platform” about a range of topics related not only to specific ELEXIS tools and services but also, more broadly, related to the use of digital methods in lexicography in general.

The goal of the current report is, therefore, two-fold:

1. to present a tentative version of the ELEXIS Curriculum as a coherent, integrated set of training materials which contextualize ELEXIS tools and services in a broader pedagogical narrative, as recommended the *ELEXIS Skillset Report*; and
2. to describe DARIAH-CAMPUS, a discovery framework and hosting platform for learning resources as the platform of choice for hosting and sustaining ELEXIS training materials.

ELEXIS Curriculum

Scope and rationale

The ELEXIS Curriculum is an integrated set of training materials which contextualizes ELEXIS tools and services inside a broader, systematic pedagogic narrative. This means that the goal of the ELEXIS Curriculum is not simply to inform users about the functionalities of particular tools and services developed within the project, but to show how such tools and services are a) embedded in both lexicographic theory and practice; and b) representative of and contributing to the development of digital skills among lexicographers.

While there is no doubt that future users of the ELEXIS infrastructure will need to know what the features of particular ELEXIS tools are and how to use them to accomplish particular tasks in the digital environment, the challenges facing ELEXIS – and lexicography in general – are not exclusively technical. Lexicographers who come to ELEXIS will have different backgrounds, not only linguistically and culturally, but also in terms of how much they and their institutions are exposed to computational practices. For a lexicographer working on a flagship national dictionary in Microsoft Word using paper slips with excerpted citations from a hundred years ago without having recourse to linguistic corpora, the path towards an active participation in and meaningful contribution to ELEXIS will be very different than the one taken by his or her more computationally aware colleagues.

If ELEXIS is to fulfill its stated mission of integrating, extending and harmonizing national and regional efforts in the field of lexicography throughout Europe, it will need to cater equally to both types of users: those with little or no experience in e-Lexicography, and those who have intermediate and advanced skills. Both types of users are equally important for lexicography as a scholarly practice with a rich tradition and an important social and cultural role. Some lexicographic communities are yet to make the leap into the digital, but that does not make their existing skills, knowledge and ongoing projects any less indispensable in their respective linguistic communities. Only by attracting and maintaining the involvement of a diverse userbase will ELEXIS manage to build and sustain a truly European infrastructure.

This is why the collection of ELEXIS training materials is not meant to be synonymous with a software documentation library.

The ELEXIS Curriculum will, with resources that are available to the Consortium and members of the Work Package 5 *Training and Education*, provide both a gentle introduction to and broad coverage of topics relevant to lexicographic practice in the computer age, paying special attention to:

- maintaining a healthy balance between theory and practice; and
- establishing explicit links between the more theoretical topics and the description of concrete tools and services.

By embedding the description of ELEXIS tools and services within a broader pedagogical narrative, WP5 will contribute to the wider adoption of those tools. In addition, by creating learning resources that are methodologically anchored and not exclusively tool-centric, WP5 will encourage critical engagement with the same tools. As one informer in ELEXIS D5.1 said: „The tools of today are not

necessarily the tools of tomorrow. We should prepare students not only how to learn tools but also to be able to switch to different tools.”

Target Audience

While most of the courses featured on the ELEXIS Curriculum are understandably geared toward lexicographers and students of lexicography, participants in the workshop *The Future of Academic Lexicography: Linguistic Knowledge Codification in the Era of Big Data and AI*, held from Nov 4-8, 2019 at the Lorentz Center in Leiden, Netherlands, identified an additional target group that should be covered by the ELEXIS Curriculum: namely, programmers and developers who join lexicographic projects but have little lexicographic background and need to be quickly brought up to speed, both in terms of their understanding of the fundamental lexicographic concepts as well as their familiarity with recent trends in computational tools and methods.

It would have been beyond the scope of WP5 to develop advanced tutorials in various programming languages, database management, or web technologies. We focus on lexicographers, their methods and tools. But to address the explicit needs of the programmers working for and with lexicographers, we will direct them towards our introductory courses (see *Introduction to Dictionaries* and *Introduction to Dictionary Use*) while at the same time creating a specific educational resource that will serve as a technical overview and guide to existing technical literature on topics that they should become familiar with (see *Trends in Computational Lexicography: A Pathfinder for Developers*).

Learning Objectives, Outcomes, Workflows and Scenarios

The courses offered as part of the ELEXIS curriculum shall have clearly defined learning objectives and outcomes, which will be explicitly defined at the beginning of each course, together with any possible prerequisites.

In addition to being classified according to three different levels of expertise (introductory, intermediate, advanced), the courses will be built in a modular fashion to support mutual linking as well as different activity workflows and learning scenarios.

For instance, a user who accesses ELEXIS through the LEX1 platform will be presented with 6 different buttons representing possible *activity workflows* within the ELEXIS infrastructure proper: using ELEXIS tools and services to *convert, link, publish, create, edit* and *enrich* one's dictionary. Each of these workflows will link to the relevant information about the specific ELEXIS tools and services that can be used for the given purpose as well as the relevant portions of the ELEXIS Curriculum that elaborate upon the said tools and services.

The curriculum itself will however also provide *learning scenarios* that go beyond the use of specific tools and services. For instance, absolute beginners who want become experts in TEI for the purpose of creating advanced digital editions of historical dictionaries would be encouraged to consider a number of different courses:

- at the introductory level: *Introduction to Dictionaries* (in order to gain basic knowledge about lexicographic terminology); and *Capturing, Modeling and Transforming Lexical Data: An Introduction* (in order to gain basic knowledge about retrodigitization and data modeling);

- at the intermediate level: *Standards for Representing Lexical Data: An Overview* (in order to understand how TEI fits in with the rest of the standardization landscape, and what other options are out there); *Modeling dictionaries in TEI Lex-0* (in order to learn about the best practices of encoding dictionaries); and *LEX3: Mastering ELEXIS Tools for Legacy Dictionaries* (in order to learn how to publish TEI-encoded dictionaries using ELEXIS' Dictionary Viewer); and, finally,
- at the advanced level: *Mastering oXygen XML Editor for Dictionary Nerds* (in order to become efficient encoders who can take advantage of all the advanced feature of the state-of-the art XML editor); *Extracting Lexical Data: XPath for Dictioanry Nerds* (in order to learn how to query TEI-encoded dictionaries); and *Transforming Lexical Data: XSLT for Dictionary Nerds* (in order to learn how to transform TEI-encoded dictionaries into different output formats).

As the courses which are part of the ELEXIS Curriculum get developed over the next two years of the project, members of WP5 should pay special attention to defining different *learning scenarios* in order to enable multiple transversal paths across the curriculum.

Various learning scenarios will be strengthened by:

- setting clear prerequisites at the beginning of each course; and
- linking not only between individual courses but also, when appropriate, referring to specific parts within different courses.

Because not all WP5 partners will work on all the courses, it will be duty of the WP5 Lead (BCDH) to oversee the process of mutual linking between different learning units.

Quality Assurance

ELEXIS training materials will be developed by WP5 partners: Jožef Stefan Institute (JSI), Lexical Computing (LC), Institute of Dutch Language (IVDNT), National University of Ireland at Galway (NUIG), Austrian Academy of Sciences (OeAW), Belgrade Center for Digital Humanities (BCDH), Research Institute for Linguistics Hungarian Academy of Sciences (RILMTA), Faculty of Social Sciences and Humanities at the Nova University of Lisbon (FCSH-UNL), Institute for Computational Linguistics “A. Zampolli” (CNR-ILC), and the University of Trier (UT). Each of these institutions has advanced expertise and a great deal of experience in some of the topics which will be covered by the suite of ELEXIS training materials.

WP5 partners will, nonetheless, try to solicit help from other lexicographic partners, both within and beyond the ELEXIS consortium, in order to test and validate the ELEXIS training materials from the beginning. Specifically, ELEXIS Observes should be granted early access to training materials and asked for feedback.

Finally, whenever possible, some ELEXIS learning units should be tested for their suitability for blended learning, an approach which combines online training materials with traditional face-to-face opportunities. In preparation for future face-to-face workshops, such as those previously organized in cooperation with the DARIAH Lexical Data Masterclass or the Lisbon Summer School in Linguistics, for instance, parts of the ELEXIS Curriculum could be given to workshop participants to study in advance of the actual workshop. During the workshops, students could give feedback and point to particular challenges they encountered while using ELEXIS training materials on their own.

Courses

Introduction to Dictionaries

Level: Introductory

Contributors: BCDH, FCSH-UNL

The goal of this course is to introduce a brief history of dictionaries as a tool for the organization of knowledge about words and their meanings, and to analyze different ways of understanding and classifying the dictionary genre. In order to do so, the course will cover the constituent parts of a dictionary (macrostructure, microstructure and mediostructure) as well as different kinds of dictionary typologies, including those based on source and target languages (monolingual, bilingual, multilingual); types of language(s) and topic(s) covered (general language, encyclopedic, terminological); medium (print and electronic); semantic structure (onomasiological vs. semasiological dictionaries); and target audience (literate adults, language learners, language professionals). At the end of this course, students will have a fundamental understanding of the complexities of the dictionary genre as well as an appreciation of the role played by the medium in which the dictionary is compiled and consumed -- from clay tablets to computer screens.

Introduction to Dictionary Users

Level: Introductory

Contributors: JSI, OeAW, RILMTA

The goal of this course is to introduce students to the important role played by dictionary usage research when developing and implementing new dictionaries. The course will address the question of how different types of target users (in terms of age, language proficiency and pre-existing skills) or different types of use (encoding, decoding, translation etc.) influence the scope of the dictionary, the lemma selection process or the very structure of a dictionary entry. At the end of this course, students will have a fundamental understanding of the ways in which user research (both commercially and academically) can contribute to the tailoring of lexicographic content. Going beyond the realm of user-centered lexicography, the course will also explore possible user contributions in the creation of content and the increasing importance of crowdsourcing in lexicography.

Introduction to Corpus-Based Lexicographic Practice

Level: Introductory

Contributors: JSI, IVDNT, LC

This course will explore the notion of lexicographic evidence and the limitation of subjective views on language by tracing the changes in lexicographic practice from the extensive use of manually selected citations to the employment of large language corpora. The course will introduce the fundamentals of corpus linguistics and corpus design, as well as the role played by various NLP tools such as taggers and parsers. At the end of this course, the students will have a basic understanding of how corpora can be used in dictionary writing, both for lemma selection, sense disambiguation, composing good definitions, choosing good examples and for automatic term and collocation extraction.

Capturing, Modeling and Transforming Lexical Data: An Introduction

Level: Introductory

Contributors: BCDH

This course will introduce the theories, practices, and methods of digitizing legacy dictionaries for research, preservation and online distribution by focusing on the process of converting paper-based dictionaries to electronic format through image capture, text capture, data modeling and data enrichment. In addition to explaining how various OCR and HCR tools can be used to extract text from images, the course will focus on analyzing, identifying and describing lexicographic data using markup languages such as XML in order to produce semantically structured datasets that can be easily queried, shared and transformed to different outputs. The course will be of interest not only to those who are converting legacy dictionaries, but also those who want to understand the principles and modes of representing structured lexicographic data, which will be a prerequisite for more advanced coursework on TEI Lex-0, OntoLex-Lemon, XPath and XSLT.

LEX2: Mastering ELEXIS Corpus Tools for Lexicographic Purposes

Level: Intermediate

Contributors: LC, IVDNT

The course will introduce corpus tools available in ELEXIS and describe various ways in which they can be exploited in lexicographic research and for compiling dictionaries. At the end of the course, students will be able to build corpora, as well as evaluate their quality and suitability for the respective lexicographic task. Students will also acquire the skills necessary to work with the tools effectively to retrieve valid linguistic information from the corpus. Practical examples will simulate the use of these tools in the lexicographic workflow.

Lexonomy: Mastering the ELEXIS Dictionary Writing System

Level: Intermediate

Contributors: LC, IVDNT, JSI, RILMTA

The course will explore how software tools for dictionary production (so-called dictionary writing systems, or DWS) can be used to streamline and facilitate the structural coherence and quality assurance in a dictionary project by focusing on Lexonomy, a dictionary-writing system developed as part of ELEXIS. At the end of this course the students will know how to use Lexonomy in various stages of the lexicographic workflow, from creating a dictionary, selecting a suitable data model, and setting up different configuration options, to using advanced features such as workflow monitoring, and preparing the dictionary for publication. Selected existing projects will be used as case studies.

Automating the Process of Dictionary Creation

Level: Intermediate

Contributors: LC, IVDNT, JSI

Building upon the material covered in *LEX2: Mastering ELEXIS Corpus Tools for Lexicographic Purposes* and *Lexonomy: Mastering the ELEXIS Dictionary Writing System*, this course will focus specifically on the changes in dictionary production after 2000 and the increasing importance of automation and post-editing in lexicography. The course will focus on the ELEXIS One-Click Dictionary as a corpus-based dictionary-drafting tool and the way Lexonomy DWS can be used to post-edit content which has been automatically pulled from ELEXIS Corpus Tools.

CLARIN Tools and Resources for Lexicographic Work

Level: Intermediate

Contributors: CNR-ILC, OeAW

This course will present an overview of tools and datasets available in CLARIN for supporting lexicographic work. The course will focus on Semantic Web technologies and the linked data paradigm by presenting LexO, a collaborative web editor offering a user-friendly interface to build and manage lexical and terminological resources based on the Lemon model. At the end of the course, students will be able to use the CLARIN infrastructure to find tools for lexicographers and linguistic datasets, depending on their own needs. In particular, they will acquire the ability to model their own linguistic data with the lemon model, without necessarily having fully-developed technical skills in the Semantic Web and Linked Data technologies. They will learn how to create lexical resources ex novo or how to convert the existing ones, even if they are not familiar with the underlying technical details.

Standards for Representing Lexical Data: An Overview

Level: Intermediate

Contributors: FCSH-UNL

This course will impress upon students the importance of shared standards for the productive cooperation among lexicographers in a multicultural and multilingual context by referring to a number of lexicographically relevant standards such as TEI, TEI Lex-0, ISO 1951, ISO 24613 (LMF), OntoLex-Lemon etc. The relevant standards will be classified according to their aims (those dealing with linguistic content; those that are used for annotation purposes; those that handle representation purposes; those addressing interoperability issues etc.) At the end of the course, students will be able to conduct dictionary research and work with dictionaries, taking into account standards in different parts of their lexicographic workflows.

Modeling Dictionaries in TEI Lex-0

Level: Intermediate

Contributors: BCDH, CNR-ILC, FCSH-UNL

The course will focus on modeling dictionaries using TEI Lex-0, a subset of the community standard TEI (Text Encoding Initiative). The course will focus on best-practices and recommendations in view of accuracy, consistency and interoperability of lexicographic data. At the end of this course, the students will become familiar with the underlying principles and the explicit guidelines of TEI Lex-0 by learning how to encode a number of dictionary entries through step-by-step tutorials with the ultimate goal of being able to adopt TEI Lex-0 in their own work.

Modeling Dictionaries in OntoLex-Lemon

NUIG, CNR-ILC

This course describes the OntoLex-Lemon model, a recent standard for the representation of lexical information on the Web as linked data. In addition to providing a basic introduction to linked data and the Resource Description Framework (RDF), the course will cover the core model of OntoLex and how to represent basic lexical information. Additional modules of the OntoLex module for the description of syntax, term decomposition, variation & translation, metadata, lexicography, morphology and corpus information will also be described. At the end of the course, students should be able to express lexicons as linked data using the model.

LEX3: Mastering ELEXIS Tools for Legacy Dictionaries

Level: Intermediate

Contributors: JSI, UT

This course will introduce Elexifier, a cloud-based dictionary service for the conversion of legacy XML and PDF dictionaries into a shared data format based on the ELEXIS Data Model; and the Legacy Dictionary Viewer, a generic, modular dictionary publication tool for retrodigitized dictionaries which offers interfaces for the analysis and profiling of the underlying lexical data. At the end of this course, students will be able to convert XML or PDF dictionaries into TEI-compliant XML files in line with the specification described in the ELEXIS data model; as well as publish user-friendly, online editions of retrodigitized dictionaries using the Legacy Dictionary Viewer without facing extensive technical or financial obstacles.

Lexicography in the Age of Open Data

Level: Intermediate

Contributors: CNR-ILC, NUIG

Technology should not necessarily be seen as the ultimate challenge facing lexicography today: social, cultural and legal obstacles often stand in the way of collaboration and knowledge sharing. This course will explore the principles of open access, open data, FAIR principles and open science as they apply to lexicography including the specific challenges posed by intellectual property rights and copyright issues in the context of lexicographic work.

Mastering LEX1: The Dictionary Matrix

Level: Advanced

Contributors: JSI, NUIG, FCSH-UNL

This course will introduce the concept and the ELEXIS implementation of the dictionary matrix, a universal repository of linked senses, and other types of lexical information found in existing lexicographic resources. Students who complete the course will become a) familiar with the techniques used to compile the dictionary matrix; b) capable of searching the dictionary matrix using the GUI (with input fields, dropdowns etc.) and/or using a SPARQL endpoint; and c) aware of the steps needed to contribute to the matrix by linking monolingual dictionaries at a sense level using the tools provided by LEX1; or extending this to multilingual linking using BabelNet.

Mastering oXygen XML Editor for Dictionary Nerds

Level: Advanced

Contributor: BCDH

This course will help users learn how to use oXygen XML, a versatile, professional-grade XML editor to edit, validate, query and transform lexicographic data. The course will focus on a range of practical and time-saving functionalities and features, such as the schema-based tool tips, document and code templates, customizable shortcuts as well as different editing modes (text, grid and CSS-styled Author) which can be used to adapt the display of dictionaries to one's particular editing needs. At the end of this course, the students will have become accustomed to the oXygen editing environment and ready to work productively and efficiently in it.

Extracting Lexical Data: XPath for Dictionary Nerds

Level: Advanced

Contributor: BCDH

This course will cover the fundamentals of XPath (XML Path Language), a standard query language for selecting nodes from XML documents. After explaining the basic syntax of XPath (axes, nodes and predicates), the course will guide the students through a number of real-life dictionary-specific scenarios (for instance: how do you select only those entries whose translation equivalents are missing the xml:lang attribute? how do you select only those entries whose etymological information contains etymons from Latin?) in order to help them hone their skills. At the end of the course, students will be able to write their own XPath expressions in order to navigate around XML-encoded dictionaries and select only those bits of data that they are interested in.

Transforming Lexical Data: XSLT for Dictionary Nerds

Level: Advanced

Contributor: BCDH

The course builds upon *Extracting Lexical Data: XPath for Dictionary Nerds* and introduce the basics of XSL Transformations (XSLT), a standard language for transforming XML documents. After explaining the basic syntax and processing model of XSLT (stylesheet declarations, templates, pattern matching etc.), the course will guide students through a number of real-life dictionary-specific scenarios (renaming, adding or removing elements and attributes, rearranging and sorting elements, performing tests, hiding and showing portions of the dictionary content etc.) in order to help them improve their skills. At the end of this course, students will be able to write their own XSLT stylesheets to transform lexicographic data.

Trends in Computational Lexicography: A Pathfinder for Developers

Level: Advanced

Contributors: IVDNT

This learning resource will provide an overview of the trending topics in computational lexicography ranging from automatic extraction of good examples and word-sense disambiguation to machine learning. The course will be aimed at developers who are new to lexicography and who quickly want to familiarize themselves with the most important literature and/or available learning resources which can bring them up to speed in the fast-changing field.



Hosting platform: DARIAH-Campus

Rationale

As we already saw in ELEXIS D5.1, most informants imagined ELEXIS as a central hub or reference point for collecting and developing openly accessible training materials through a “specific website” or a “platform”. In addition, in the DoW, ELEXIS committed to embed and sustain the learning resources it develops in cooperation with the Digital Research Infrastructure for Arts and Humanities (DARIAH) by “leverage[ing] and build[ing] upon the significant experience of DARIAH VCC2 Training and Education”.

In November 2019, DARIAH launched DARIAH-Campus, a discovery framework and hosting platform for training and education. The goal of DARIAH-Campus (D-C) is to widen access to open, inclusive, high-quality learning resources that aim to enhance creativity, skills, technology and knowledge in the digital arts and humanities. While not being specific to lexicography, D-C recognizes the diversity of scholarly communities using computational methods in the humanities, and welcomes contributions from different fields and disciplines.

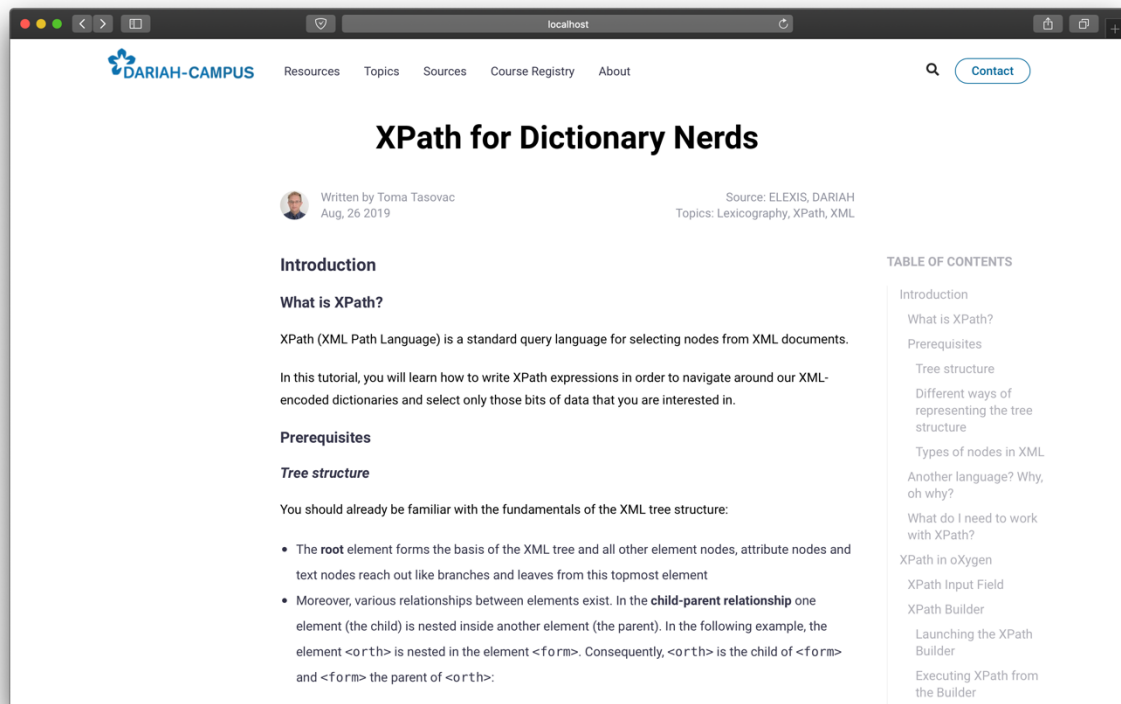
DARIAH-Campus is the platform of choice for ELEXIS learning resources for a number of reasons:

- D-C is a data-centric overlay site built with modern web technologies (specifically Gatsby and React.js) on top of the content which is stored and versioned in a GitHub repository;
- the separation of data and presentation layers on D-C allows content creators to focus on content creation, while delegating the uniform and user-friendly styling of the output to the automatic build process;
- the versioning of the learning resources on D-C provides a significant advantage for ELEXIS training materials in terms of updatability and citeability;
- D-C content is written in Markdown, an open text format which is very easy to learn, providing intuitive workflows and fast turnaround; while multimedia and interactive content can be added using easy-to-understand MDX (JSX in Markdown) components;
- because of its plaintext syntax, Markdown content, hosted and versioned in a digital repository, should be easier to reuse and repurpose than only the fully generated website in HTML;
- D-C comes with humanist-friendly documentation;
- D-C learning resources contain structured metadata written in YAML, a human-readable data serialization format;
- D-C metadata requirements will be aligned with major European infrastructural initiatives such as the European Open Science Cloud (EOSC) and Social Sciences and Humanities (SSHOC) Marketplace, guaranteeing future discoverability and interoperability of ELEXIS learning resources;
- by outsourcing hosting of the ELEXIS training materials to D-C, members of ELEXIS WP5 can fully focus on curriculum development and content creation;
- D-C learning resources are by default licensed under a Creative Commons CC BY license. This ensures that they are available for free; can be re-mixed and repurposed; and are easy to combine with other materials and courses, while requiring that proper credit be given to the original authors;

- D-C is maintained by DARIAH, which is a European Research Infrastructure Consortium (ERIC) and an ESFRI Landmark. The legal status of DARIAH offers a stable framework for sustaining the ELEXIS learning resources beyond the end of ELEXIS as a funded H2020 project.



Figures



The screenshot shows a web browser window displaying a page from DARIAH-CAMPUS. The page title is "XPath for Dictionary Nerds". The author is Toma Tasovac, dated August 26, 2019. The source is ELEXIS, DARIAH, and the topics are Lexicography, XPath, and XML. The page content includes an introduction, a section on "What is XPath?", and a "Prerequisites" section with a "Tree structure" subsection. A table of contents is located on the right side of the page.

Introduction

What is XPath?

XPath (XML Path Language) is a standard query language for selecting nodes from XML documents.

In this tutorial, you will learn how to write XPath expressions in order to navigate around our XML-encoded dictionaries and select only those bits of data that you are interested in.

Prerequisites

Tree structure

You should already be familiar with the fundamentals of the XML tree structure:

- The **root** element forms the basis of the XML tree and all other element nodes, attribute nodes and text nodes reach out like branches and leaves from this topmost element
- Moreover, various relationships between elements exist. In the **child-parent relationship** one element (the child) is nested inside another element (the parent). In the following example, the element `<orth>` is nested in the element `<form>`. Consequently, `<orth>` is the child of `<form>` and `<form>` the parent of `<orth>`:

TABLE OF CONTENTS

- Introduction
- What is XPath?
- Prerequisites
 - Tree structure
 - Different ways of representing the tree structure
 - Types of nodes in XML
- Another language? Why, oh why?
- What do I need to work with XPath?
- XPath in oXygen
- XPath Input Field
- XPath Builder
 - Launching the XPath Builder
 - Executing XPath from the Builder

Figure 1: D-C learning resource as it appears to the user. The table of contents is generated automatically from Markdown headings.

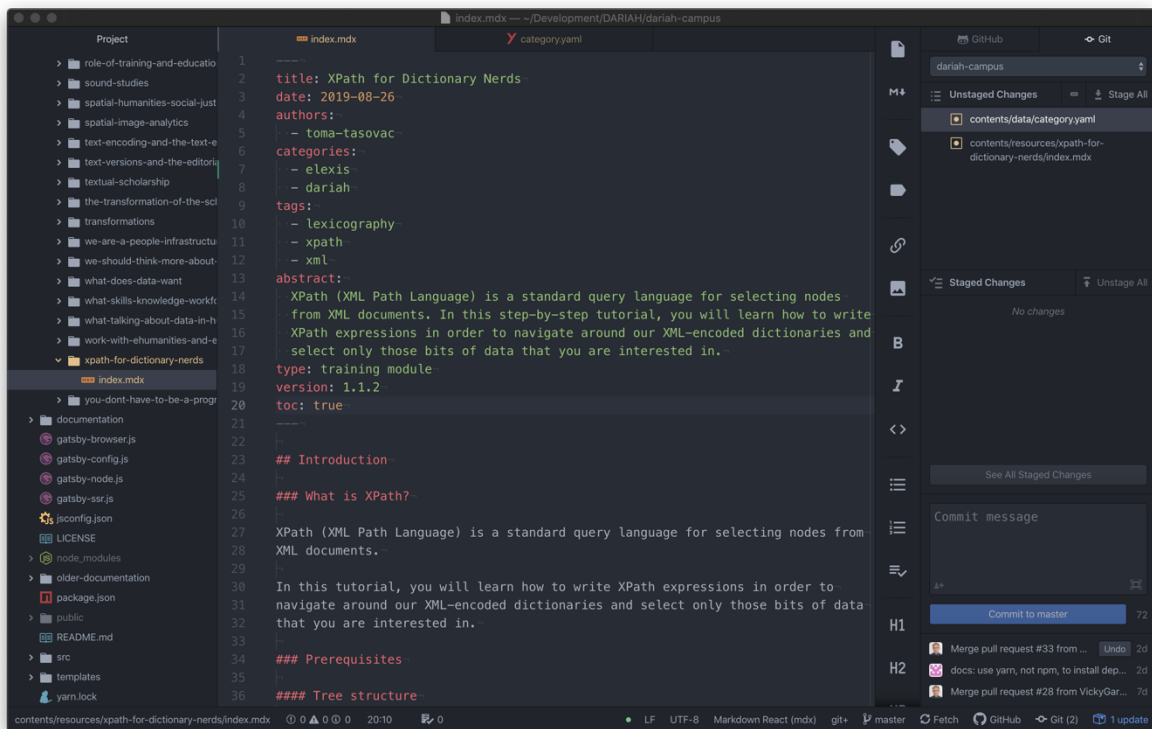


Figure 2: D-C resource as it appears to the content creator in a text editor starting with YAML metadata fields, and followed by Markdown content.

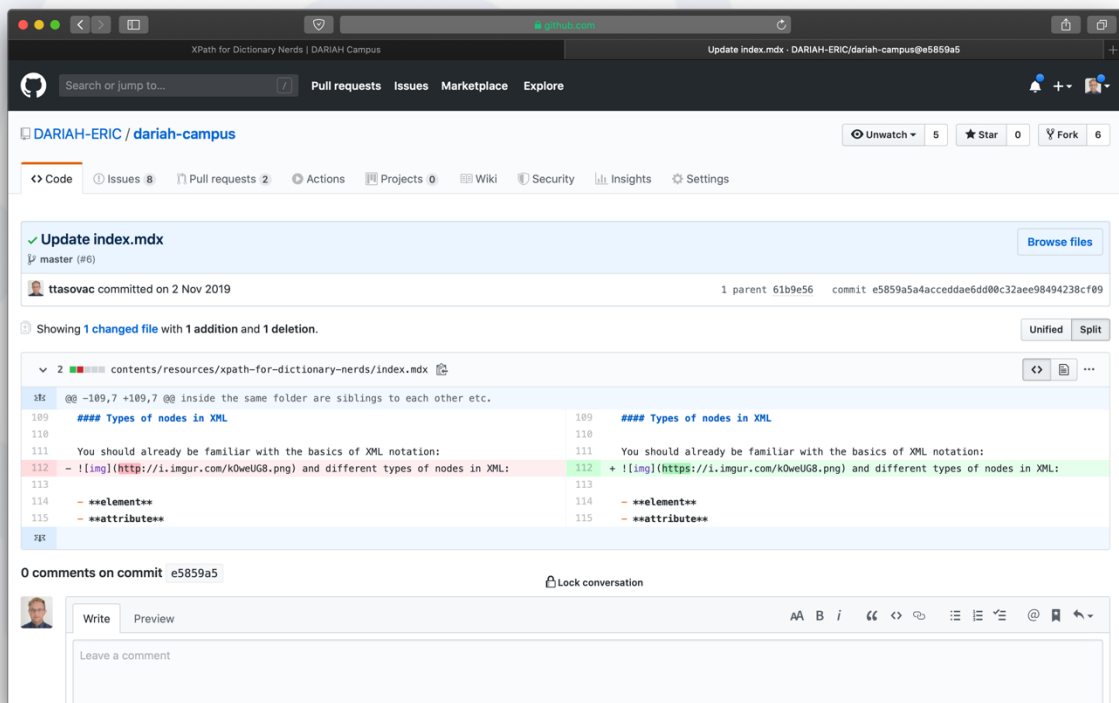


Figure 3: History of changes of a D-C resources is available on GitHub