CLARIN Training and Education
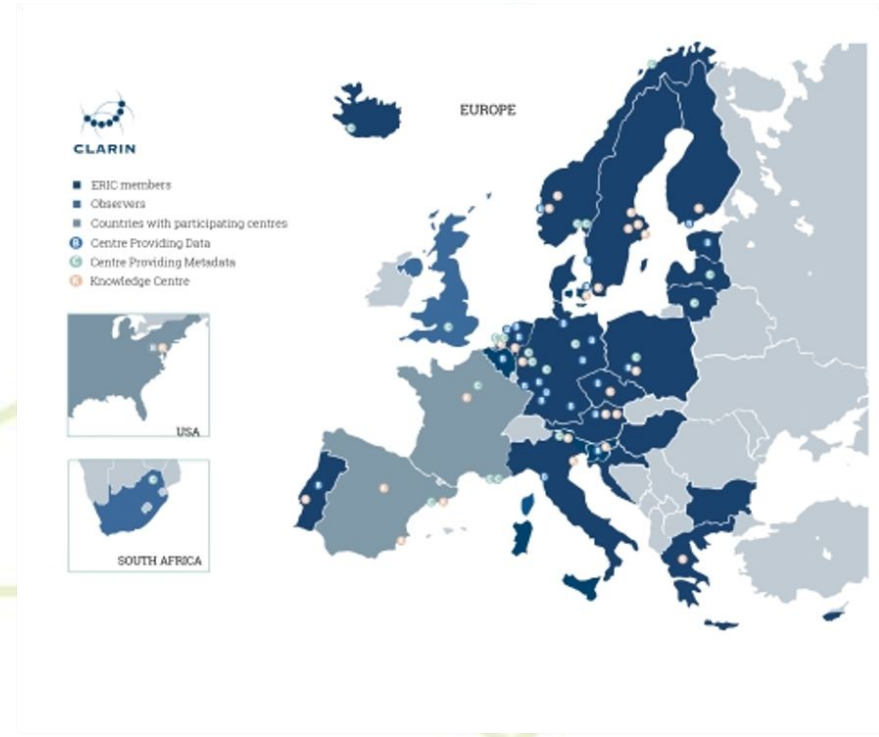
Francesca Frontini, BoD

04 November 2022
Third Hybrid UPSKILLS Multiplier Event:
Guidelines and Best Practices for Research-Based Teaching
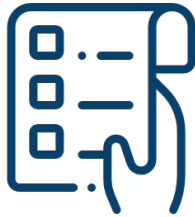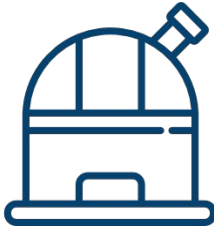
CLARIN Today

- a distributed network of **70 centres**
- **22 members**: AT, BE, BG, CY, CZ, DE, DK, EE, FI, GR, HR, HU,IS, IT, LT, LV, NL, NO, PL, PT, SE, SI
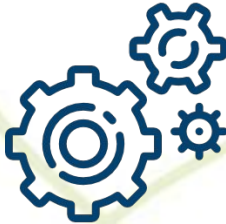- **2 observers**: UK, ZA
- **1 third party**

CLARIN Technical Infrastructure

FAIR sharing of Language Resources

Meta catalogue for finding LRs wherever they are

Interoperability framework to match LRs and services

CLARIN Knowledge Infrastructure

Knowledge Centres

Digital Humanities Course Registry

Tour de CLARIN

Teaching

Annual Conference

Funding

Video Channels

Best-Practice Papers

CLARIN Knowledge Infrastructure

Knowledge Centres

Digital Humanities Course Registry

Tour de CLARIN

Teaching

Annual Conference

Funding

Video Channels

Best-Practice Papers

DIGITAL HUMANITIES COURSE REGISTRY
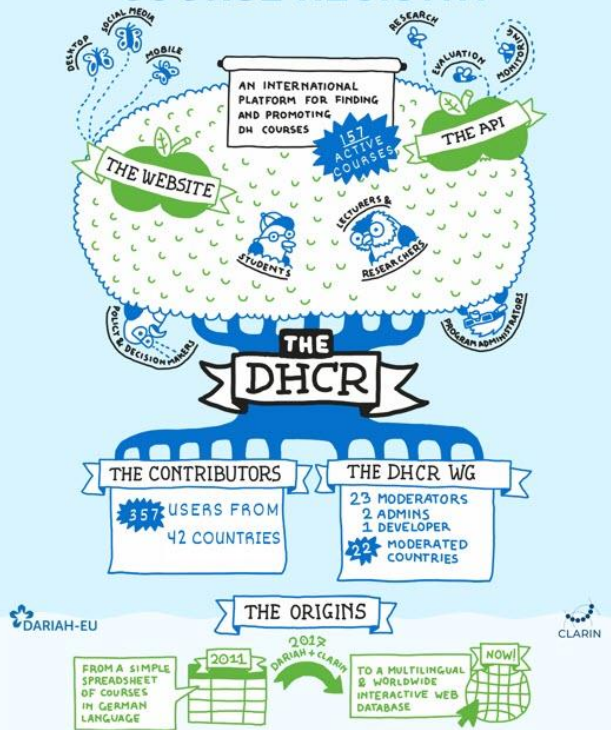
NAME	UNIVERSITY	PLACE	DATE	TYPE
Web development - Javascript	Université de Lausanne	Lausanne, Switzerland	20 Sept ⟳ 1 semester	Bachelor Programme
Methodenkurs - Digitale Methoden und bewährte Werkzeuge für ...	Universität Wien	Vienna, Austria	1 Mar ⟳ 1 semester	Course
Methodenkurs - Digitale Annotation im geschichtswissenschaf ...	Universität Wien	Vienna, Austria	1 Oct ⟳ 1 semester	Course
Masterstudium "Digitale Geisteswissenschaften" (Master Prog ...	Karl-Franzens-Universität Graz	Graz, Austria	2 Oct ⟳ 2 years	Master Programme

https://dhcr.clarin-dariah.eu/

Teaching with CLARIN initiative

Training Materials

Applied Language Technology
Author: Tuomo Hiippala
Faculty of Arts, University of Helsinki, Finland
Keywords: *language technology, digital humanities, tutorial, beginner, spaCy, Stanza, Universal Dependencies, introduction*

Archilochus of Paros: Elegiac Fragments - XML Archive
Author: Anika Nicolosi and Beatrice Nava
University of Parma, Italy
Keywords: *Ancient Greek, Fragmentary poetry, Textual criticism, Text annotation, Data science*

Computational Morphology with HFST
Author: Erik Axelson
Faculty of Arts, University of Helsinki, Finland
Keywords: *morphology, weighted finite-state networks, two-level rules, xfst, lexc, twolc*

GATE, an Open-Source Toolkit for Natural Language Processing
Author: Diana Maynard
Faculty of Engineering, University of Sheffield
Keywords: *Natural Language Processing; Machine Learning; GATE; social media analysis; disinformation; online abuse detection; Python; Deep Learning; information extraction; digital humanities; corpus linguistics; annotation*

CLARIN 2021

Teaching with CLARIN Award

Sharing teaching materials as digital objects

Description of the Training Materials

(Sub)discipline, topic, language(s)	**Topics**: Language technology, digital humanities **Language**: English
Keywords	language technology, digital humanities, tutorial, beginner, spaCy, Stanza, Universal Dependencies, introduction
Project URL	Applied Language Technology (mooc.fi) YouTube channel
CLARIN resources	The course materials build on various resources distributed through CLARIN, such as Universal Dependencies corpora. The materials refer to the CLARIN website for further study, highlighting the digital humanities course registry.
Structure and duration	The learning materials constitute a 10 ECTS module, consisting of two 5 credit courses. The materials are divided into two parts, in which each section corresponds to one week of studying.

- quality metadata
- visibility
- citability
- persistent identifiers
- versioning
- licensing

CLARIN and FAIR Training Materials

CoP Task Force - OpenAire

Joint publication to be submitted in

Scientific Data

How to make your training materials FAIR

1

Visibility on SSH Open Marketplace

Social Sciences & Humanities Open Marketplace

Discover new and contextualised resources for your research in Social Sciences and Humanities: tools, services, training materials, workflows and datasets. Read more…

| Training materials ⌄ | 🔍 | Search |
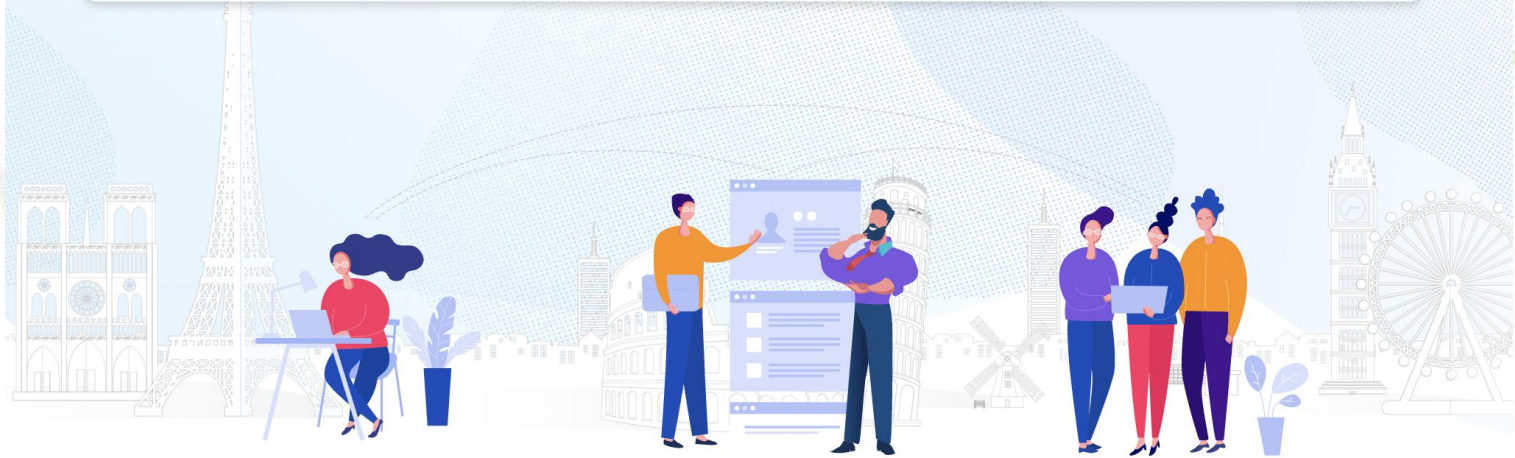
2022: Teaching CLARIN

Developing new learning content to help teachers and lecturers to integrate CLARIN into their courses.

Focus on core CLARIN services:
- VLO
- Switchboard
- Federated content search
- Deposit and Citation

Innovative types of training materials: notebooks

Computational notebooks provide a popular environment for literate programming, especially in education contexts.

Notebooks

Computational notebooks (in short notebooks) provide a popular environment for literate programming, especially in education contexts. This also applies to Natural Language Processing. On this page a first overview is provided of the use of Notebooks in connection with the CLARIN infrastructure.

Notebooks provided by CLARIN centres

The following notebooks are provided by CLARIN centres:
- LINDAT/CLARIAH-CZ:
  - POS Tagging and Lemmatization (Czech model)
  - Sentiment Analysis (Czech Model)
- SAFMORIL:
  - Online tutorial for Computational Morphology with HFST
  - Source Code repository on github
- ILC4CLARIN:
  - SSHOC Data Stewardship terminology and Metadata SKOSifying mapping
- PORTULAN CLARIN:
  - Tokenization: Segmentation of texts into lexical tokens.
  - Syllabification: Syllabification of expressions.
  - Sentence splitting: Segmentation of texts into sentences and paragraphs.
  - UPOS tagging: Tokenization and morphosyntactic tagging of expressions in texts with Universal Dependencies POS tagset.
  - LXPOS tagging: Tokenization and morphosyntactic tagging of expressions in texts with LX POS tagset.
  - UÉvoraPOS tagging: Tokenization and morphosyntactic tagging of expressions in texts with UÉvora POS tagset.
  - Universal Sub-syntactic analysis: Tokenization, lemmatization, inflection analysis and morphosyntactic tagging of expressions in texts within the Universal Dependencies framework.
  - LX Sub-syntactic analysis: Tokenization, lemmatization, inflection analysis and morphosyntactic tagging of expressions in texts within the LX framework.
  - Named entity recognition: Detection and semantic classification of names in texts.
  - Universal Dependency parsing: Analysis of grammatical functions in sentences within Universal Dependencies framework.
  - LX Dependency parsing: Analysis of grammatical functions in sentences within LX framework.
  - Constituency parsing: Analysis of syntactic constituents in sentences.
  - Grammatical quantitative analysis: Occurrence counting of grammatical elements in texts.
  - Machine Translation: Translation of a sentence from a source language to a target language (Portuguese-Chinese).
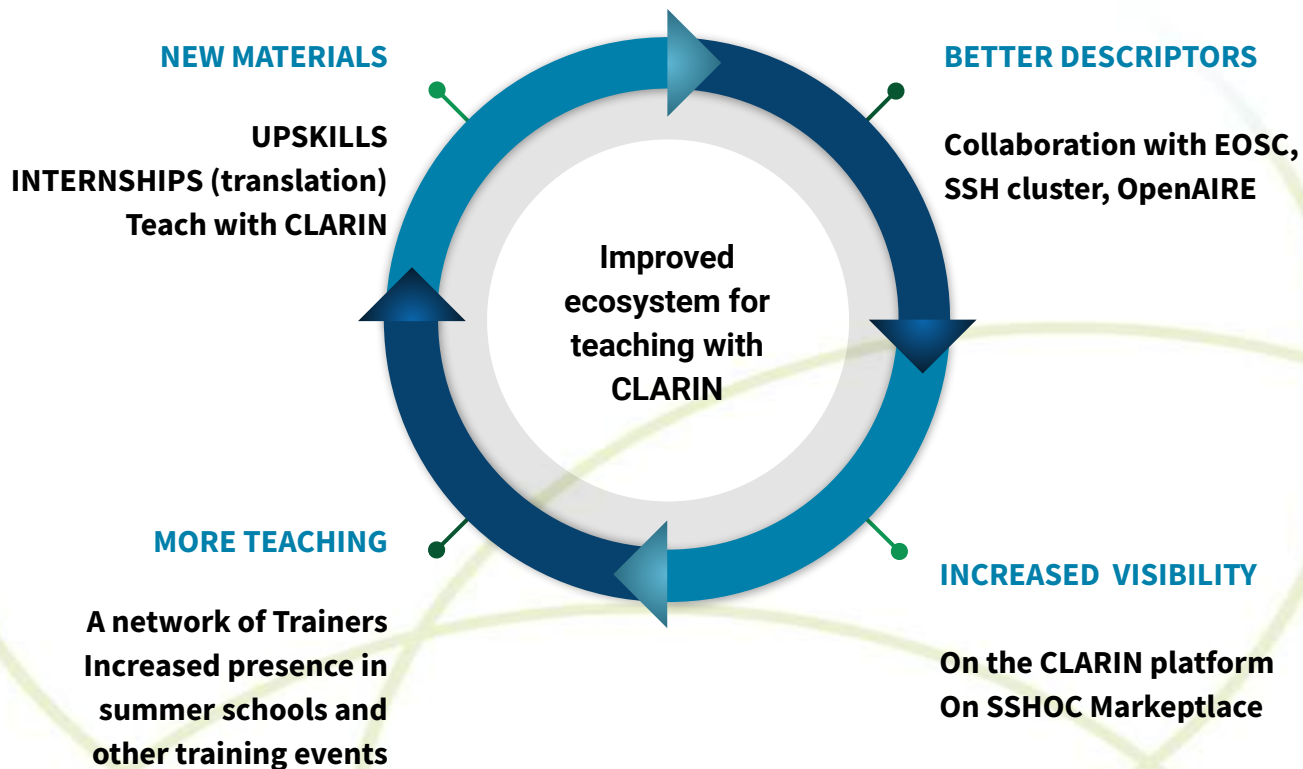
Processing Europeana Text Collections with Jupyter Notebooks by Twan Goosen and Michal Gawor

https://www.clarin.eu/notebooks

Future goals

NEW MATERIALS

**UPSKILLS
INTERNSHIPS (translation)
Teach with CLARIN**

BETTER DESCRIPTORS

**Collaboration with EOSC,
SSH cluster, OpenAIRE**

**Improved
ecosystem for
teaching with
CLARIN**

MORE TEACHING

**A network of Trainers
Increased presence in
summer schools and
other training events**

INCREASED VISIBILITY

**On the CLARIN platform
On SSHOC Markeptlace**

Get involved with CLARIN

Get in touch with your national consortium

https://www.clarin.eu/content/participating-consortia

Events

https://www.clarin.eu/events

Support and funding

https://www.clarin.eu/funding

Get involved CLARIN for training

Subscribe to our **trainers' network** mailing list:

training@lists.clarin.eu

Submit materials to the **Teaching with CLARIN call**

https://www.clarin.eu/content/call-submissions-teaching-clarin-call

Contact the **Training and Education Task Force**

training@clarin.eu