

Teaching with CLARIN 2022

Introduction

Francesca Frontini, BoD
Iulianna van der Lek, Training Officer

Programme

Introduction and Overview of Training Activities by Francesca Frontini and Iulianna van der Lek

Processing Europeana Text Collections with Jupyter Notebooks by Twan Goosen and Michal Gawor

Teaching with CLARIN Submissions 2022

- [What's on the Agenda? Topic Modelling Parliamentary Debates before and during the COVID-19 Pandemic](#) by Ajda Pretnar Žagar, Kristina Pahor de Maiti, Darja Fišer, Ljubljana, Slovenia
- [Lithuanian Collocations: Usage, Teaching, Learning, and Translation](#) by [Jurgita Vaičenonienė](#), Vytautas Magnus University, Lithuania
- [Natural Language Processing Methods](#) by [Rachele Sprugnoli](#), Dipartimento di Discipline Umanistiche, Sociali e delle Imprese Culturali – Università degli Studi di Parma, Italy

Discussion & Certificates

Teaching & Training | CLARIN KI



Knowledge Centres



Digital Humanities Course Registry



Tour de CLARIN



Teaching



Annual Conference



Funding



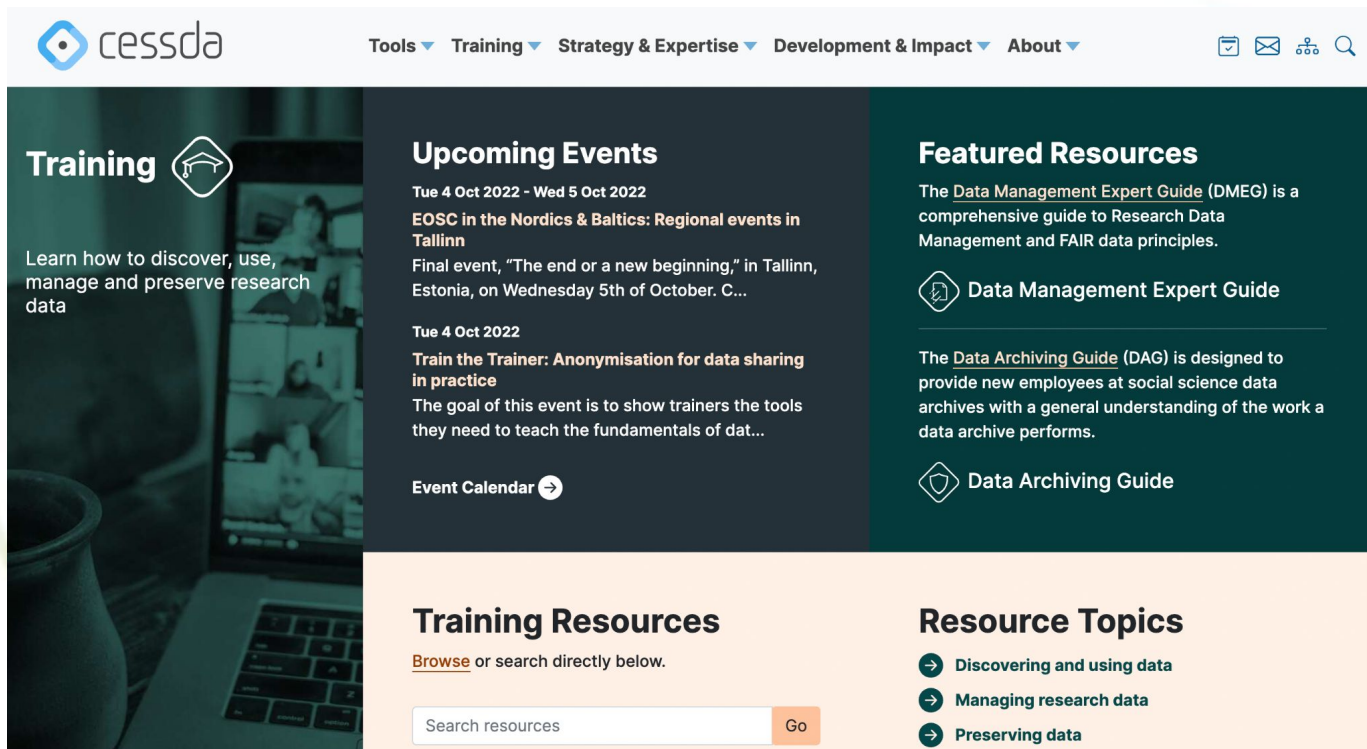
Video Channels




Best-Practice Papers

In alignment with other infrastructures

In alignment with other infrastructures



The screenshot displays the CESSDA website interface. At the top, the CESSDA logo is on the left, and navigation links for Tools, Training, Strategy & Expertise, Development & Impact, and About are in the center. On the right, there are icons for a calendar, email, a group of people, and a search function. The main content area is divided into several sections: a 'Training' section with a laptop background and text about discovering, using, managing, and preserving research data; an 'Upcoming Events' section listing events from October 4th to 5th, 2022, including 'EOSC in the Nordics & Baltics: Regional events in Tallinn' and 'Train the Trainer: Anonymisation for data sharing in practice'; a 'Featured Resources' section highlighting the 'Data Management Expert Guide' and the 'Data Archiving Guide'; a 'Training Resources' section with a search bar and a 'Go' button; and a 'Resource Topics' section with three bullet points: 'Discovering and using data', 'Managing research data', and 'Preserving data'.

Training 

Learn how to discover, use, manage and preserve research data

Upcoming Events

Tue 4 Oct 2022 - Wed 5 Oct 2022

EOSC in the Nordics & Baltics: Regional events in Tallinn

Final event, "The end or a new beginning," in Tallinn, Estonia, on Wednesday 5th of October. C...

Tue 4 Oct 2022


Train the Trainer: Anonymisation for data sharing in practice

The goal of this event is to show trainers the tools they need to teach the fundamentals of dat...


Event Calendar →

Featured Resources

The [Data Management Expert Guide \(DMEG\)](#) is a comprehensive guide to Research Data Management and FAIR data principles.

 **Data Management Expert Guide**

The [Data Archiving Guide \(DAG\)](#) is designed to provide new employees at social science data archives with a general understanding of the work a data archive performs.

 **Data Archiving Guide**

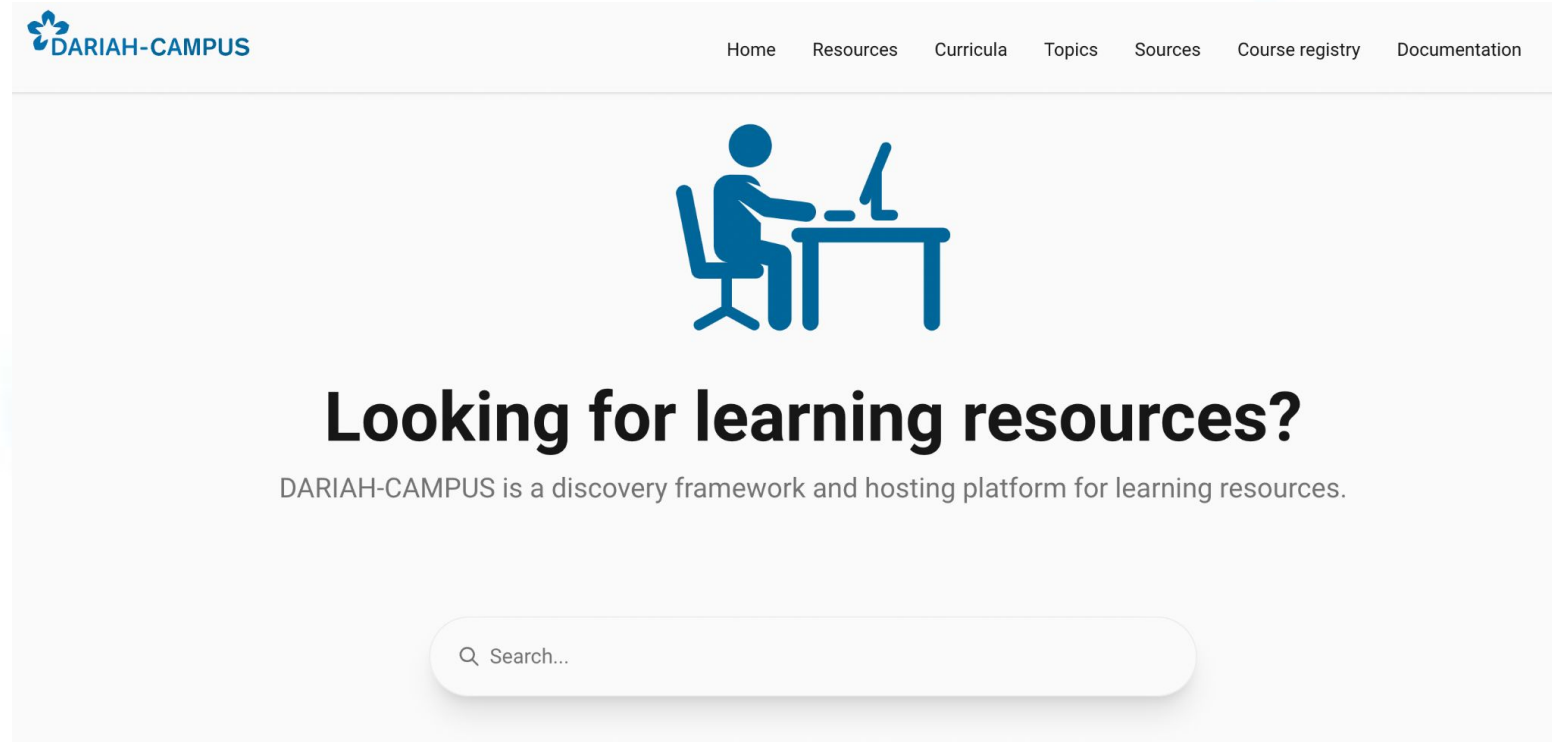
Training Resources

[Browse](#) or search directly below.

Resource Topics

- [Discovering and using data](#)
- [Managing research data](#)
- [Preserving data](#)

In alignment with other infrastructures



and with the SSHOC vision



[Report an issue](#)

[Tools & services](#)

[Training materials](#)

[Publications](#)

[Datasets](#)

[Workflows](#)

[Browse](#)

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

2021: Teaching with CLARIN initiatives

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 Axelsson

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



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.

Open issues:

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

CLARIN and FAIR Training Materials

CoP Task Force - OpenAire

Joint publication to be submitted in

[Scientific Data](#)



How to make your training materials
FAIR

Abstract	2
Introduction	4
The evolving FAIR landscape	4
Research communities and FAIR	4
FAIR training materials	5
Research communities and FAIR training materials	7
Methodology	8
Selection of case studies	8
Structure of case studies and analysis	9
Implementing FAIR standards for training resources: lessons learned from case studies	10
Analysis and consolidation	10
Discussion	11
Further reading and appendices	19
CLARIN	19
Context, scope and objectives	19
Implementation	19
Challenges and Lessons Learnt	20
ELIXIR (work in progress! Just bullets for now)	23
ELIXIR Training	23
ELIXIR and FAIR training	23
Implementation	24
Challenges and Lessons learnt	25
EMBL Bio-IT	26
Bio-IT: introduction	26
Scope and objectives	26
Implementation	26
Lessons learnt	27
OpenAIRE	28
SSHOC - Training Discovery Toolkit	30
CESSDA Training resources catalogue- work in progress	33
FORRT (work in progress)	35
Context, scope and objectives	35
Implementation	36
Challenges and Lessons Learnt	38
EOSC Synergy (in progress)	39
Background	39

2022: Teaching CLARIN

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

1) Focus on core CLARIN services:

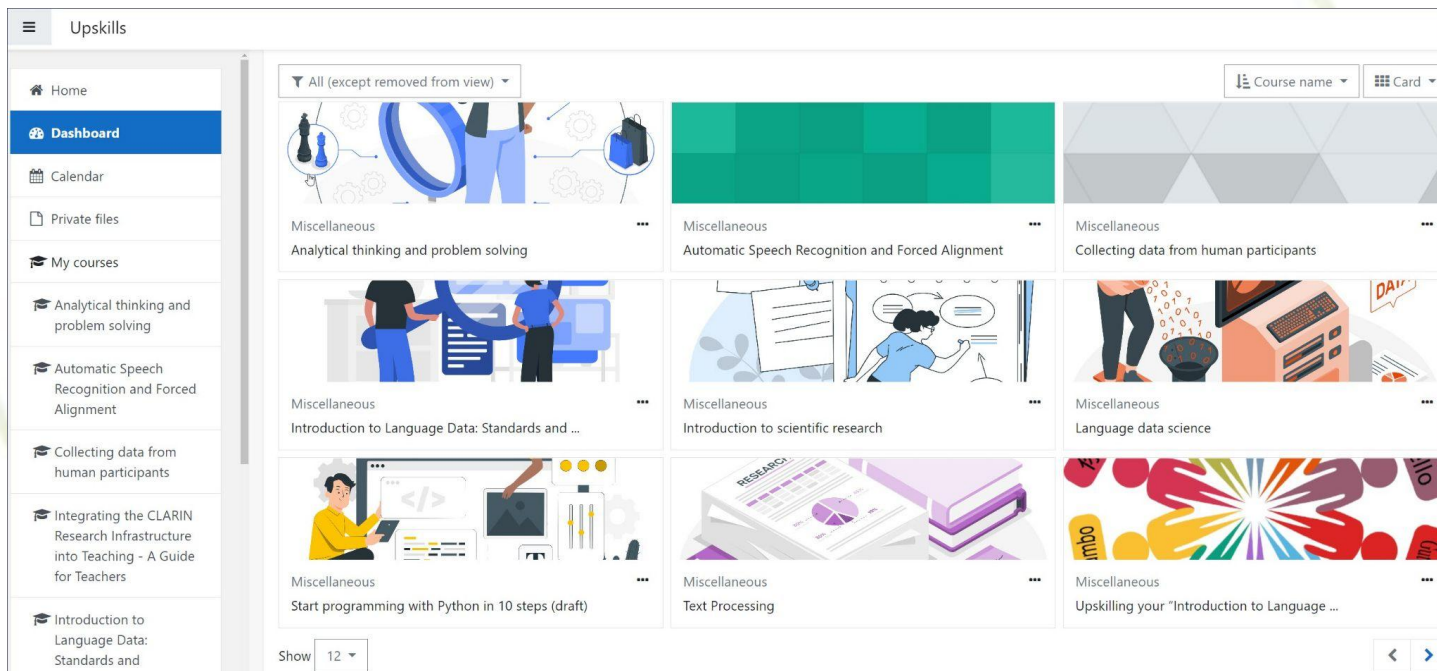
- VLO
- Switchboard
- Federated content search
- Deposit and Citation

2) Automatic Speech Recognition & Force Alignment



What's next

Finalise the UPSKILLS learning content by February 2023 ->
Testing, translation & piloting



What's next

Student internships with FORLI from the University of Bologna



The screenshot displays the official website of the University of Bologna. At the top, the university's name is written in Latin and Italian, accompanied by its seal and a logo for 'for Zaky' featuring a red ribbon and glasses. A navigation bar in red contains links to various university sections. Below this, a large banner features a stylized map of Italy with the 'FORLÌ CAMPUS' highlighted. The main content area shows the breadcrumb trail: Home > Forlì Campus > Research at Forlì > Department of Interpretation and Translation- DIT. A sidebar on the left lists campus-related services, with 'Research at Forlì' highlighted in red. The main text describes the Department of Interpretation and Translation (DIT), emphasizing its role in language and culture studies and its commitment to quality in translation.

Information for ▼

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA *for Zaky*

IT EN

HOME UNIVERSITY ▼ TEACHING ▼ RESEARCH ▼ UNIVERSITY AND SOCIETY ▼ INTERNATIONAL ▼ SERVICES AND OPPORTUNITIES ▼

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI FORLÌ

FORLÌ CAMPUS

Home > Forlì Campus > Research at Forlì > Department of Interpretation and Translation- DIT

Forlì Campus

Presentation and organisation

Studying at Forlì

Campus Services

"R. Ruffilli" Central Library

Research at Forlì

Department of Interpretation and Translation- DIT

Communications between different languages and cultures occupy a place in the Department of Interpretation and Translation's scientific and teaching programmes.

The whole translation process calls for in-depth study and specific criteria of quality. Translating appropriately and consciously is a holistic skill embracing many scientific, productive, cultural and artistic factors

Department of Interpretation and Translation- DIT

What's next

More text processing tutorials with Jupyter Notebooks

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).


Next presentation

[Processing Europeana Text](#)


[Collections with Jupyter](#)

[Notebooks](#) by Twan Goosen
and Michal Gawor

<https://www.clarin.eu/notebooks>



Upgrading the Skills of Linguistics and Language Students



Why

Growing gap between the knowledge and skills acquired at universities and those required by the labour market, e.g.

- Analysis and handling of large data collections
- Knowledge of programming languages
- Design, evaluate and adjust AI systems to the needs of new registers, styles and languages
- Project management

How

Needs analysis

- A survey of curriculum
- A literature review (institutional, academic, professional)
- A corpus-based analysis of job advertisements
- A questionnaire for the language industry
- Interviews with industry representatives

Reports available at <https://upskillsproject.eu/>


Language data and project specialist — knowledge, skills and competences —

DESCRIPTION	RESEARCH	DATA ANALYSIS
DESCRIPTION OF RESEARCH DESIGN Description of research design, including: - Research objectives - Research questions - Research methodology - Research instruments - Research procedures	RESEARCH DESIGN Description of research design, including: - Research objectives - Research questions - Research methodology - Research instruments - Research procedures	DATA ANALYSIS Description of data analysis, including: - Data collection - Data cleaning - Data management - Data analysis - Data interpretation
INFORMATION	RESEARCH	DATA ANALYSIS
INFORMATION Description of information, including: - Information sources - Information types - Information formats - Information quality	RESEARCH Description of research, including: - Research objectives - Research questions - Research methodology - Research instruments - Research procedures	DATA ANALYSIS Description of data analysis, including: - Data collection - Data cleaning - Data management - Data analysis - Data interpretation
TECHNICAL	RESEARCH	DATA ANALYSIS
TECHNICAL Description of technical, including: - Technical specifications - Technical requirements - Technical constraints - Technical risks	RESEARCH Description of research, including: - Research objectives - Research questions - Research methodology - Research instruments - Research procedures	DATA ANALYSIS Description of data analysis, including: - Data collection - Data cleaning - Data management - Data analysis - Data interpretation

Reaching classrooms

REACHING CLASSROOMS
Introduction to specific research
methodologies and practices
involving linguistic data analysis
Project management

REACHING CLASSROOMS
Introduction to specific research
methodologies and practices
involving linguistic data analysis
Project management



REACHING CLASSROOMS
Introduction to specific research
methodologies and practices
involving linguistic data analysis
Project management

UPSKILLS PROJECTS
Introduction to specific research
methodologies and practices
involving linguistic data analysis
Project management

UPSKILLS PROJECTS
Introduction to specific research
methodologies and practices
involving linguistic data analysis
Project management

Key features of the Research/Teaching materials

Format:

- Modular Moodle unit blocks intended for use by lecturers
- Incorporation of contents into educational games
- Access: Creative Commons License (CC BY)

Dissertation:

- [Digital Humanities Course Registry](#)
- [SH Open Marketplace](#)

Integrating Research Infrastructure into Teaching

Introduction to Language Data: Standards and Repositories (4 ECTS)

Designers: Iuliana van der Lek & David Faria

Contributors by: Francesca Frontini, Alexander König, Willem Elbers

CLARIN in the Classroom

A Guide for Teachers and Students









	Unit 1: Introduction to Research Infrastructures of Language Resources and Technology	
	Unit 2: Finding, Accessing and Using Language Resources Repositories, Catalogues, Resources, Content Search	
	Unit 3: Tools for Linguistic Processing, Annotation and Analysis Switchboard, Weblight, Concordancers	
	Unit 4: Archiving and Sharing Language Resources Subject project in collaboration with University of Bologna	

Illustration sources:
Illustrations by Storvick: <https://istockphoto.com>


CLARIN is a trademark of the European Union


Follow us


Contact
E-mail: iuliana@clarin.eu


Funding
grant agreement no. 2020-1-MT01-KA202-074246

Partners







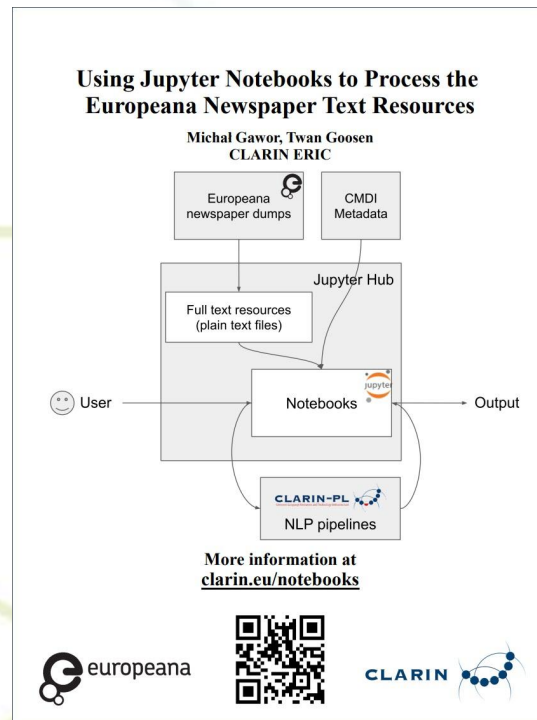


Download poster

<https://upskillsproject.eu/>

<https://twitter.com/upskillsproject>

UPSKILLS PROJECTS
Introduction to specific research
methodologies and practices
involving linguistic data analysis
Project management



Get involved

Training and Education Task Force

training@clarin.eu

Subscribe to our trainers' network mailing list:

training@lists.clarin.eu