

CLARIN-IT

the Italian Common Language Resources and Technology Infrastructure



CLARIN ERIC and CLARIN-IT

Francesca Frontini - *CLARIN Board of Directors*

Monica Monachini - CLARIN-IT National Coordinator

Istituto di Linguistica Computazionale - ILC CNR

13/06/2022





Materials for this lesson:

<https://docs.google.com/document/d/144SWjq-i-bRF2nbbandg2ZEF5OEW60zN6uDIJaNlp4/>

1. Open Data, SSH, Research Infrastructures
2. The CLARIN Research Infrastructure

Exercise 1 - Finding and processing data

3. Knowledge Sharing

Exercise 2 - The ParlaMint corpora

4. The CLARIN-IT national consortium

Exercise 3 - Explore ILC4CLARIN

5. A Cluster of SSH research Infrastructures

Exercise 4 - The SSHOC Marketplace

Open Data, SSH, Research Infrastructures

A New Paradigm



"Open Science represents a **new approach to the scientific process** based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools ([European Commission, 2016](#))"

Open science encompasses unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives. ([OCSE, 2015](#))



- Humanities Research, more and more data driven
 - *Are there defining characteristics of 19th century European novels, across different literary traditions?*
 - *What are the stylistic differences between Classic French and English Theatre?*
- Quantitative studies, based on corpora

Data and DH: ELTeC





Language	Last update	AUTHORSHIP						LENGTH			TIME SLOT					REPRINT COUNT			E5C
		Texts	Words	Male	Female	1-title	3-title	Short	Medium	Long	1840-59	1860-79	1880-99	1900-20	range	Frequent	Rare		
cze	2021-04-09	100	5621667	88	12	62	6	43	49	8	12	21	39	28	27	1	19	80.00	
deu	2021-04-09	100	12738842	67	33	35	9	20	37	43	25	25	25	25	0	48	46	96.92	
eng	2021-04-09	100	12227703	49	51	70	10	27	27	46	21	22	31	26	10	32	68	100.00	
fra	2021-04-09	100	8712219	66	34	58	10	32	38	30	25	25	25	25	0	44	56	101.54	
hrv	2021-03-22	21	1440018	21	0	4	0	6	12	3	6	12	2	1	11	1	0	23.08	
hun	2021-04-09	100	6948590	79	21	71	9	47	31	22	22	21	27	30	9	32	67	100.00	
ita	2019-11-21	34	3328244	32	2	19	3	13	10	11	5	12	10	7	7	12	0	55.97	
lav	2020-12-20	2	106045	2	0	2	0	0	2	0	0	0	1	1	1	0	1	21.54	
lit	2020-08-20	25	636132	18	7	16	1	19	3	2	5	3	3	14	11	6	18	55.38	
nor	2021-04-09	53	3432676	38	15	18	11	26	18	9	4	2	30	17	28	32	21	67.69	
pol	2021-04-09	100	8500172	58	42	1	33	33	35	32	8	11	35	46	38	39	61	80.00	
por	2021-04-09	100	6799385	83	17	73	9	40	41	19	13	37	19	31	24	26	60	94.62	
rom	2021-04-09	95	5558961	74	16	54	9	46	31	18	4	17	25	49	45	24	71	80.00	
slv	2021-04-09	100	5682120	89	11	26	5	53	39	8	2	13	36	49	47	48	52	78.46	
spa	2021-04-09	84	7147890	67	17	45	5	31	28	25	17	17	25	25	8	42	42	92.31	
srp	2021-04-09	90	3961405	82	8	35	11	56	32	2	2	12	38	38	36	32	58	73.68	
swe	2021-04-09	58	4960085	29	28	18	8	16	24	18	15	3	20	20	17	17	41	76.92	
ukr	2021-04-09	50	1840062	37	13	23	7	34	13	3	5	10	11	24	19	30	20	70.77	

Source <https://www.distant-reading.net/eltec/>

Data and DH: DraCor



 ABOUT ▾ CORPORA ▾ TOOLS ▾ HOW TO ▾ MERCH 			
Fre DraCor	Ger DraCor	Rus DraCor	Ita DraCor
French Drama Corpus	German Drama Corpus	Russian Drama Corpus	Italian Drama Corpus
1,560 Number of plays	554 Number of plays	212 Number of plays	139 Number of plays
15,526 (M: 6860, F: 3720) person + personGrp Number of characters	13,305 (M: 9410, F: 2713) person + personGrp Number of characters	3,707 (M: 2608, F: 871) person + personGrp Number of characters	1,527 (M: 989, F: 413) person + personGrp Number of characters
15,671,574 Text tokens text	9,801,725 Text tokens text	2,316,996 Text tokens text	1,895,476 Text tokens text
445,403 (14,444,024) sp (Tokens)	387,277 (9,350,717) sp (Tokens)	119,332 (2,191,658) sp (Tokens)	66,607 (1,763,669) sp (Tokens)
38,306 (307,262) stage (Tokens)	181,388 (1,146,507) stage (Tokens)	49,440 (215,112) stage (Tokens)	13,522 (62,311) stage (Tokens)
Last update 22/10/2021, 16:31:07	Last update 08/02/2022, 11:06:40	Last update 04/08/2021, 21:18:51	Last update 04/08/2021, 21:10:31

Fischer, Frank, et al. (2019). Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama. In Proceedings of DH2019: "Complexities", Utrecht University, [doi:10.5281/zenodo.4284002](https://doi.org/10.5281/zenodo.4284002).

Not just Data...



Lesson Index

Our lessons are organized by typical phases of the research process, as well as general topics. Use the buttons to filter lessons by category. If you can't find a skill, technology, or tool you're looking for, please [let us know!](#)

ACQUIRE (12)

TRANSFORM (33)

ANALYZE (20)

PRESENT (20)

SUSTAIN (2)

APIS (6)

PYTHON (21)

DATA MANAGEMENT (8)

DATA MANIPULATION (26)

DISTANT READING (12)

SET UP (7)

LINKED OPEN DATA (1)

MAPPING (11)

NETWORK ANALYSIS (5)

WEB SCRAPING (5)

DIGITAL PUBLISHING (11)

RESET TO SEE ALL LESSONS (87)

START SEARCHING

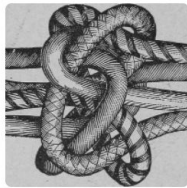
SORT BY PUBLICATION DATE ▲

SORT BY DIFFICULTY ▼

Not just Data...



FILTERING RESULTS: (20) DATE ▼

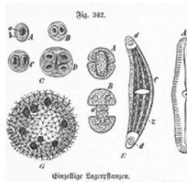


GABI KIRILLOFF

Interactive Fiction in the Humanities Classroom: How to Create Interactive Text Games Using Twine

This lesson provides strategies for incorporating game creation into the classroom. The first half of the lesson discusses the challenges and benefits of teaching game creation while the second half includes a technical tutorial for

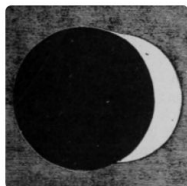
Twine, an open source game creation tool.



THOMAS JURCZYK

Clustering with Scikit-Learn in Python

This tutorial demonstrates how to apply clustering algorithms with Python to a dataset with two concrete use cases. The first example uses clustering to identify meaningful groups of Greco-Roman authors based on their publications and their reception. The second use case applies clustering algorithms to textual data in order to discover thematic groups. After finishing this tutorial, you will be able to use clustering in Python with Scikit-learn applied to your own data, adding an invaluable method to your toolbox for exploratory data analysis.



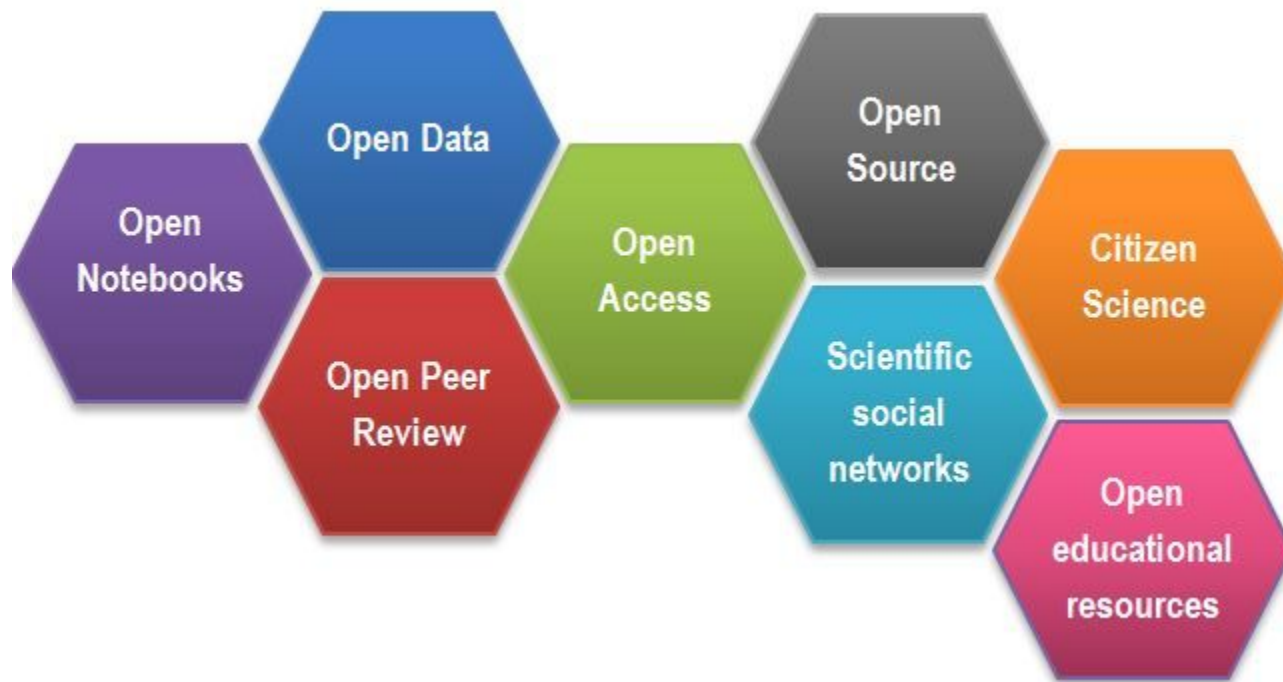
JOHN R. LADD

Understanding and Using Common Similarity Measures for Text Analysis

This lesson introduces three common measures for determining how similar texts are to one another: city block distance, Euclidean distance, and cosine

Source: <https://programminghistorian.org/>

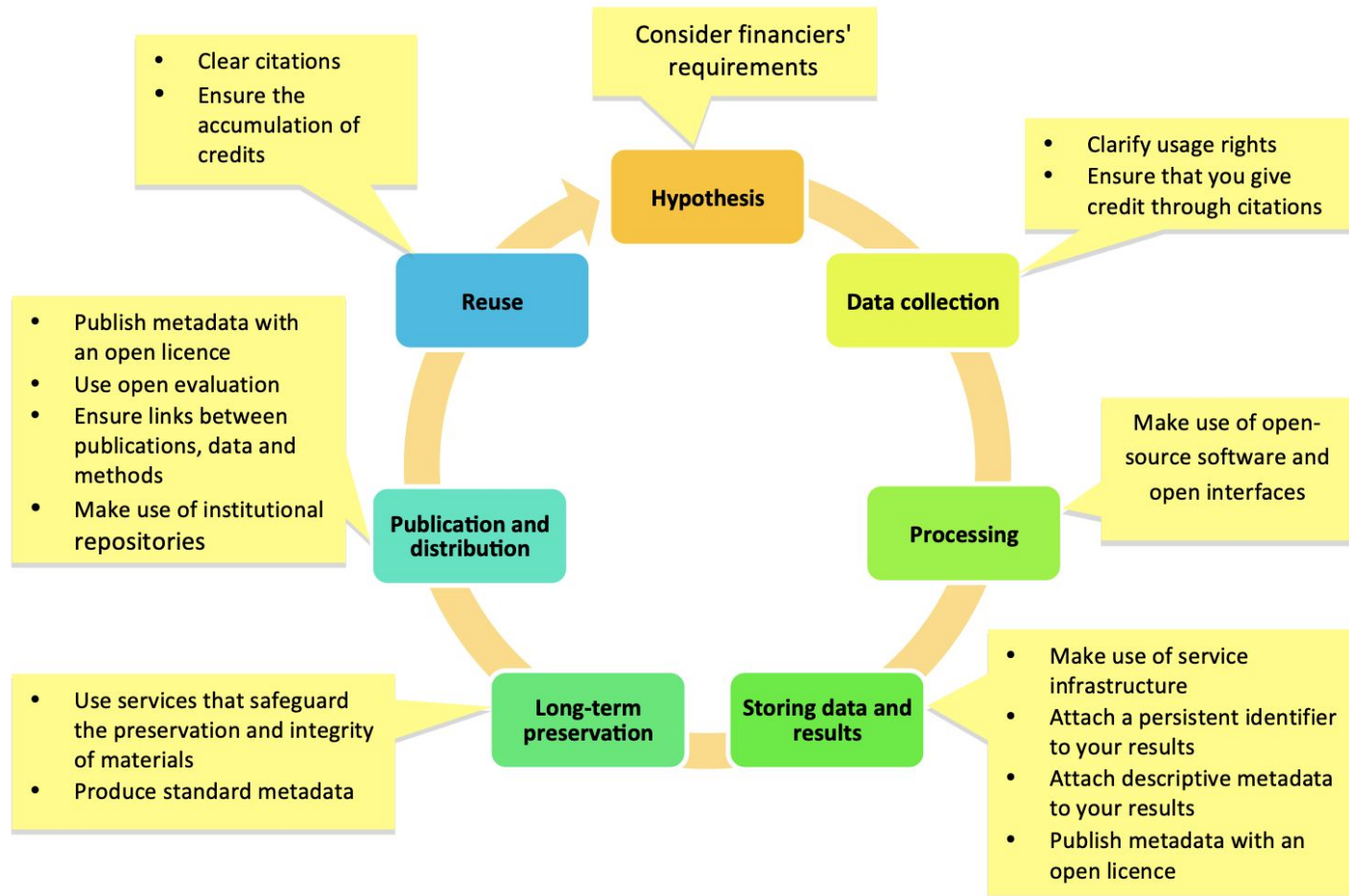
Open Science Practices



Source of the image and further information:

<https://www.fosteropenscience.eu/learning/open-science-at-the-core-of-libraries/#/id/5a01e2d1c2af651d1e3b1b3c>

... in all Areas of Research



Source: The Open Science and Research Handbook v1

https://cdn2.euraxess.org/sites/default/files/open_science_and_research_handbook_v.1.0.pdf



FAIR Principles

Compliance



Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ F1. Resource is uploaded to a public repository.
- ✓ F2. Metadata are assigned a globally unique and persistent identifier.



Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.

- ✓ A1. Resource is accessible for download or manipulation by humans and is ideally also machine readable.
- ✓ A2. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.



Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.

- ✓ I1. Resource is uploaded to a repository that is interoperable with other platforms.
- ✓ I2. Repository meta- data schema maps to or implements the CG Core metadata schema.
- ✓ I3. Metadata use standard vocabularies and/or ontologies.



Reusability

Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

- ✓ R1. Metadata are released with a clear and accessible usage license.
- ✓ R2. Metadata about data and datasets are richly described with a plurality of accurate and relevant attributes.

'FAIR Guiding Principles for scientific data management and stewardship', 2016

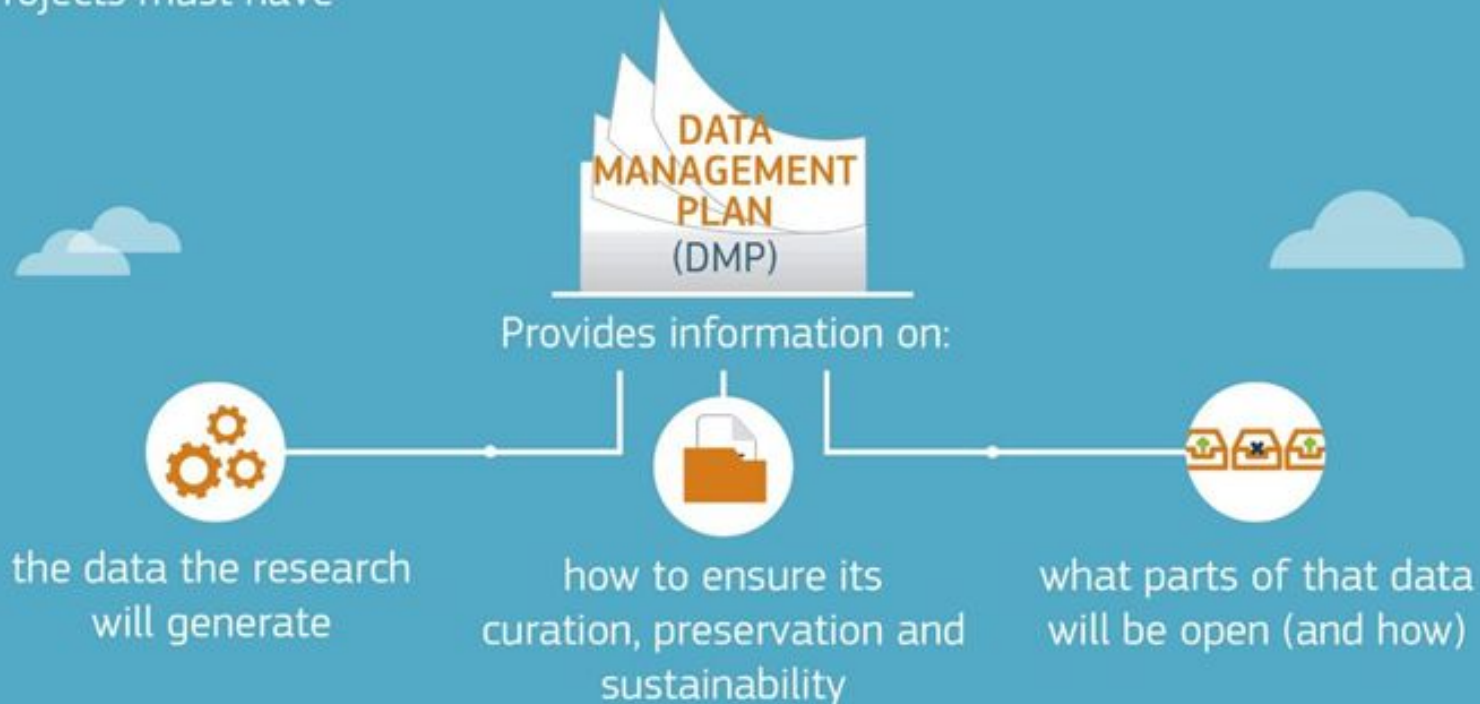
<https://www.go-fair.org/fair-principles/>

A European Strategy



RESEARCH DATA - OPEN BY DEFAULT

Projects must have



How to ensure ...

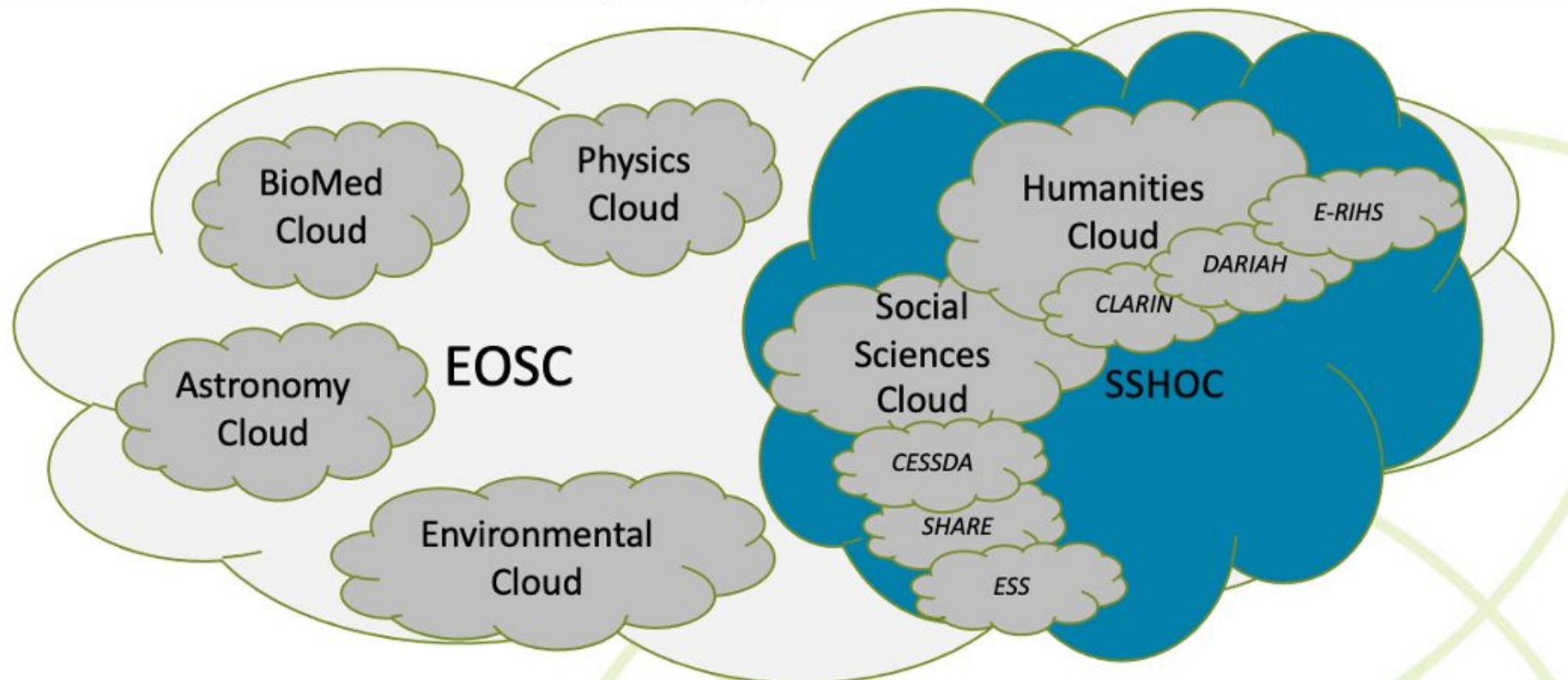


- long term deposit
- metadata quality
- findability
- citability
- clear licenses
- interoperability?



**EUROPEAN OPEN
SCIENCE CLOUD**

An Ecosystem of Infrastructures



Research Infrastructures



European Commission defines research infrastructures as **facilities** that provide **resources and services for research communities to conduct research and foster innovation**. They can be used beyond research, e.g. for education or public services and they may be single-sited, distributed, or virtual.



In the age of Artificial Intelligence and the so-called “big data society,” large [Research Infrastructures \(RIs\) in the Arts and Humanities](#), such as [CLARIN](#) and [DARIAH](#), provide scholars, researchers, students and educators with the **knowledge, high-quality data, services and technologies** to help them explore the global scientific and societal transformations, and tackle the new challenges.

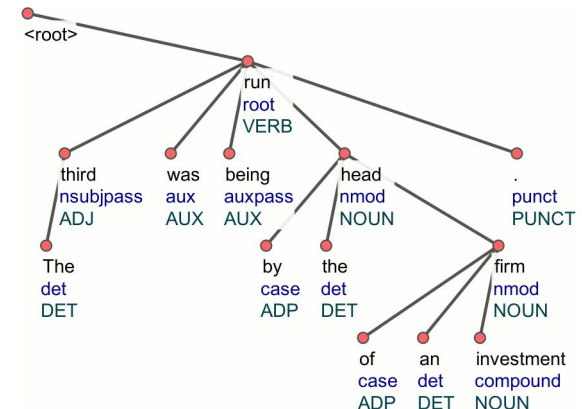
The CLARIN Research Infrastructure for data sharing

CLARIN & Language Resources



CLARIN

Common Language Resources and
Technology Infrastructure



CLARIN in a nutshell



- has the **ESFRI ERIC** status since 2012, **Landmark** since 2016
- provides easy and sustainable access for scholars in the **humanities and social sciences** and beyond
 - to digital language data (in written, spoken, video or multimodal form)
 - and **advanced tools** to discover, explore, exploit, annotate, analyse or combine them, wherever they are located
 - through a **single sign-on** environment
- serves as an ecosystem for **knowledge sharing and training**
- is an integral part of **the European Open Science Cloud**
 - See clarin.eu/eosc

CLARIN data and communities



- Newspaper archives
- Literary texts
- Parliamentary records
- Literary texts
- Historical letters
- Broadcast archives
- Oral History data
- Social Media data
- L-2 Learner Resources
- Survey data
- ...

For the CLARIN Resource Families initiative, see:
<https://www.clarin.eu/resource-families>

- Digital humanities
- Linguistics and Philology
- Translation and Lexicography
- Literary Studies
- History
- Political and Social Sciences
- Media Studies
- Culture, Folklore, Anthropology
- Speech therapy
- General Public
- ...

CLARIN today



- **21 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
- **> 70 centres**



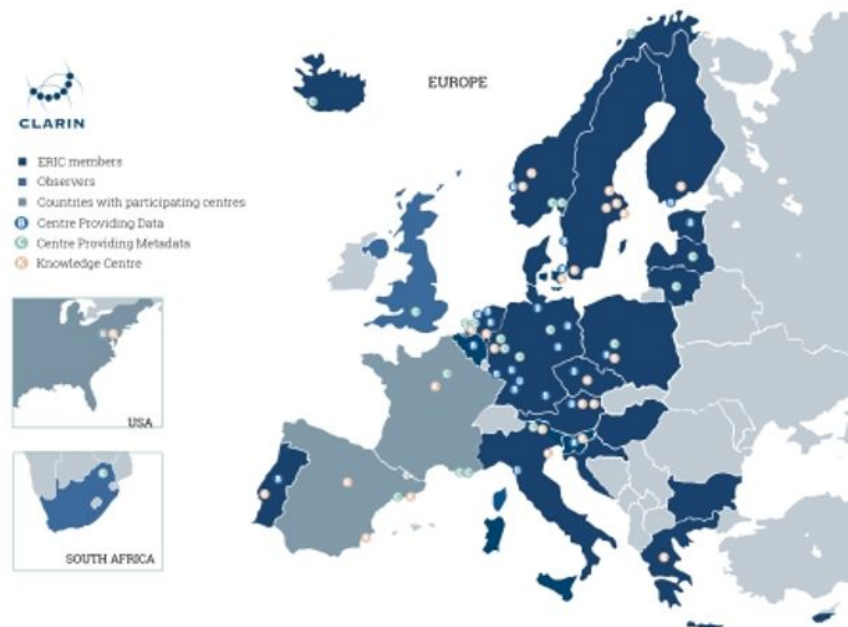
federated login



central harvesting of metadata



chained services



CLARIN types of centres



- **B-Centres**

- These are the distributed data centres that universities or academic institutions can access to search for or deposit language resources, access services and expert knowledge. There are several [certified B-centres](#) that have passed the [CLARIN centre assessment procedure](#). The CLARIN-B centres are required to apply for the [CoreTrustSeal](#) certification.

- **C-Centres**

- These are centres that provide metadata to CLARIN but they do not offer any other services.

- **K-Centres**

- Knowledge centres that share knowledge and expertise on one or more aspects of the domains covered by the CLARIN infrastructure.

....Consortia and Centres offer Data, Tools, Services



Search Catalogue Education Projects Tools Services About ▼

Digital Research Infrastructure for the Language Technologies, Arts and Humanities

Catalogue Corpora Treebanks ČDK Bibliography


|search by type of data



Search

e.g. [corpus](#) or [lexicon](#) or [editor](#)

<https://lindat.cz/>

...hosting repositories

 [Search](#) [Catalogue](#) [Education](#) [Projects](#) [Tools](#) [Services](#) [About ▾](#)



Find

Linguistic Data and NLP Tools

Citation Support (with Persistent IDs)



Search

[Advanced Search](#)

Author	Subject	Language (ISO)
Veselý, Bohumil (787)	People (803)	Nolinguistic content (712)
Hajič, Jan (89)	Galerie osobností (787)	Czech (486)
Aktualita (78)	Places (555)	English (313)
Straka, Milan (67)	machine translation (63)	German (216)
Krátký film (63)	Český zvukový týdení ... (51)	French (113)
... View More	... View More	... View More

<https://lindat.mff.cuni.cz/repository/xmlui/?locale-attribute=en>

.... offering tools for exploration and processing

Tools for Natural Language Processing



[Terms of Use](#)
[Service Status](#)

Enrich your texts
Make them more searchable

Phonetics, Phonology

- [UWebASR](#) (audio transcription, Czech)



Machine Translation

- [LINDAT translation](#)



Morphology and tagging

- [ElixirFM](#) (Arabic)
- [MorphoDiTa](#) (Czech)
- [UDPipe](#)



Natural language processing

- [Treex](#) (Czech, English, Latin)
- [UDPipe](#)



Syntactic parsing

- [Korektor](#) (spellchecker, Czech)
- [Parsito](#)
- [UDPipe](#)



Search

- [CzEngVallex](#) (lexicon, Czech, English)
- [EngVallex](#) (lexicon, English)
- [PDT-Vallex](#) (lexicon, Czech)
- [Dialogy.org](#) (audio-visual corpora search)
- [Internet Language Reference Book](#) (Czech)
- [KonText](#) (concordances, collocations, word frequencies)
- [TEITOK](#) (search, visualize, edit corpora)
- [PML Tree Query search](#) (treebank search)



Discourse, Pragmatics

- [Evald](#) (coherence evaluation, native speakers of Czech)
- [Evald](#) (coherence evaluation, non-native speakers of Czech)
- [KER](#) (keyword extraction, Czech, English)
- [NameTag](#) (named-entity recognition, Czech, English)



<https://lindat.cz/#tools>



CLARIN-UK

Infrastructure for Digital Language Resources and Tools

[Home](#)[Centres](#)[Resources](#)[People](#)[News](#)[Events](#)[Courses](#)[Blog](#)[About](#)

[Read more about CLARIN-UK](#)

Data & Tools

[More](#)

CorCenCC

Corpws
Cenedlaethol
Cymraeg
Cyfoes – the
National Corpus
of
Contemporary
Welsh

Latest News

[More](#)

New
members of
the
CLARIN_UK
consortium

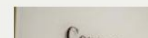
18 November
2020
Welcome
aboard!

Events

[More](#)

Lancaster
Symposium
on Innovation
in Corpus
Linguistics
2021

Wednesday 23
June, Online



Corpus

<https://www.clarin.ac.uk/>



CLARIN-UK

Infrastructure for Digital Language Resources and Tools

[Home](#)[Centres ▾](#)[Resources ▾](#)[People](#)[News](#)[Events](#)[Courses](#)[Blog](#)[About](#)[#LancsBox](#)[BNC](#)[CKLD](#)[CLAWS](#)[CLiC](#)[CorCenCC](#)[CQPweb](#)[ELAR](#)[GATE](#)[GATE Cloud](#)[Hansard at
Huddersfield](#)[Read more about CLARIN-UK](#)

Data & Tools

[More](#)

CorCenCC

Corpws
Cenedlaethol
Cymraeg
Cyfoes – the
National Corpus
of
Contemporary
Welsh

[More](#)

Events

[More](#)

Lancaster
Symposium
on Innovation
in Corpus
Linguistics
2021

Wednesday 23
June, Online



Corpus
Linguistics

<https://www.clarin.ac.uk/>

CLARIN Central Services



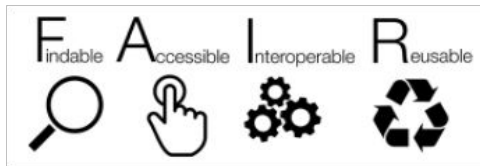
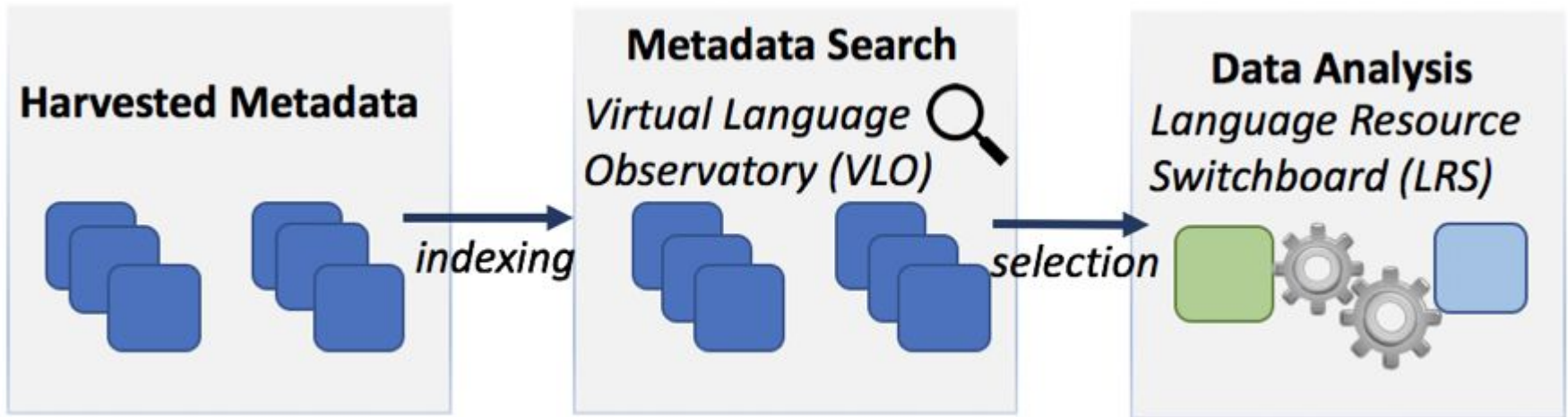
The CLARIN portal:

A single access point to the distributed infrastructure

- Make it easier for researchers, lecturers, students to **find** data, tools, training materials from centres
- Make it easier to **share and publish** data following open science practices
- Allow you to be **cited** for your work and acknowledge that of others
- Promote **knowledge sharing** and networking

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

The technical infrastructure



clarin.eu/fair



vlo.clarin.eu



switchboard.clarin.eu

Barack Obama's identity-building in the health care debate: A corpus-assisted discourse study

AUTHOR

[Katherina Riesner](#)

Summary, in English

In this study, I demonstrate that identity-building is an important discursive strategy for President Barack Obama in the seven-year long debate surrounding the Affordable Care Act (ACA). The data for the study comes from a 6-million word corpus of speeches that were held by Obama between January 2009 and January 2016, all published by the White House. The speeches are classified according to genre, audience, topic and date of delivery. Throughout the paper, I adopt the notion that identity is intentionally constructed by the speaker and strategically exploited for his communicative goals. With the help of two methodological approaches, I investigate what kind of identities Obama builds. The purely qualitative part of the study deals with three central corpus speeches from a discourse-analytic perspective. In the second, more quantitative part, I use a group of seven verbs with epistemic meaning to trace the usage of two predominant discursive identities in the ACA debate. The results suggest that President Obama repeatedly constructs the identities of father and teacher to persuade his audience. I argue that his use of these identities constitutes an attempt to reach the argumentative goals of effectiveness and reasonableness.

Department/s

Master's Programme: Language and Linguistics

Publishing year

2016

Language

English

Full text

[Available as PDF](#) - 2 MB

[Download statistics](#)

Document type

Student publication for Master's degree (two years)

Topic

Languages and Literatures

Q

METADATA SEARCH

Q

CONTENT SEARCH

🔑

MANAGE ACCESS

🔖

BOOKMARK

🌿

REQUEST ACCESS

”

CITATION

Corpus

Name

The Barack Obama Corpus

Title

The Barack Obama Corpus

Description

the_barack_obama_corpus_information.txt

Description

The Barack Obama Corpus (BOC) consists of 6,215,948 words (tokens), which are sourced from nearly 3,500 different texts, dating from January 2009 to January 2016. The texts, all taken from the White House Archives, comprise all speeches held by Barack Obama in his official capacity as 44th President of the United States of America. The earliest speech in the BOC is President Obama's inauguration speech and the last is his final State of the Union speech (January 2016). In total, the corpus includes 34,967 word types, which leads to a type/token-ratio of 0.56.

The files, which display the original titles given to them by the White House, have been tagged for genre, audience type, date and location of delivery, and principal topics. The genres include remarks, addresses, statements, press conferences, debates and question-

Description

How to cite this resource:

Riesner, Katherina (2017). The Barack Obama Corpus [Data set]. <http://hdl.handle.net/10050/00-0000-0000-0003-C53B-4#view>



Riesner, Katherina (2017). The Barack Obama Corpus [Data set].
<http://hdl.handle.net/10050/00-0000-0000-0003-C53B-4@view>

VLO / Faceted search / Search results

obama

Showing 5 results for obama  Results per page: 10 


Use the categories below to limit the search results to those matching the selected value(s).

Language Collection Modality 

The Barack **Obama** Corpus

(Part of Lund University Humanities Lab)



 the_barack_obama_corpus_information.txt; The Barack **Obama** Corpus (BOC) consists of 6,215,948 words (tokens), which are sourced from nearly 3,500 different texts, dating from January 2009 to January 2016. The texts, all taken from the White House Archives, comprise all speeches held by Barack **Obama** in his official capac...



 [Landing page for this record at corpora.humlab.lu.se](https://corpora.humlab.lu.se)

**SWE-CLARIN**

metadata

vlo.clarin.eu

UDPipe

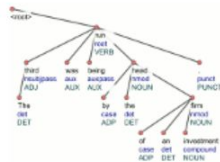
“ Please use the following text to cite this item or export to a predefined format:

[BIBTEX](#)
[CMDI](#)

Straka, Milan and Straková, Jana, 2016, *UDPipe*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-1702>.



Share:  



Authors:

Milan Straka, Jana Straková

Description:

UDPipe is an trainable pipeline for tokenization, tagging, lemmatization and dependency parsing of CoNLL-U files. UDPipe is language-agnostic and can be trained given only annotated data in CoNLL-U format. Trained models are provided for nearly all UD treebanks. UDPipe is available as a binary, as a library for C++, Python, Perl, Java, C#, and as a web service. UDPipe is a free software under [Mozilla Public License 2.0](#) and the linguistic models are free for non-commercial use and distributed under [CC BY-NC-SA license](#), although for some models the original data used to create the model may impose additional licensing conditions.

[Project home](#)
[Run](#)


Tool Inventory

▼ Constituency Parsing



> WebLicht Const Parsing DE 



> WebLicht Const Parsing EN 

▼ Coreference Resolution



> Concraft -> Bartek

▼ Dependency Parsing



> Concraft -> DependencyParser



> MaltParser



> Spacy (hosted by D4Science) - DE



> Spacy (hosted by D4Science) - EN



LINDAT

> UDPipe



> WebLicht Dep Parsing DE 



> WebLicht Dep Parsing EN 



CLARIN Resource families



Corpora

- Computer-mediated communication corpora
- Corpora of academic texts
- Historical corpora
- L2 learner corpora
- Literary corpora
- Manually annotated corpora
- Multimodal corpora
- Newspaper corpora
- Parallel corpora
- Parliamentary corpora
- Reference corpora
- Spoken corpora

Lexical Resources

- Lexica
- Dictionaries
- Conceptual Resources
- Glossaries
- Wordlists

Tools

- Normalization
- Named entity recognition
- Part-of-speech tagging and lemmatization
- Tools for sentiment analysis

Spoken corpora in the CLARIN infrastructure

Corpora with transcriptions and audio recordings

Corpus	Language	Description	Availability
Arabic Speech Corpus Licence: CC BY 4.0	Arabic	The corpus is available for download from a dedicated webpage. For a relevant publication, see Halabi (2016) .	Download
DIALEKT v1: dialectal corpus with multi-tier transcription Size: 100,000 words Annotation: orthographically and phonetically (dialect features) transcribed, MSD-tagged, lemmatised Licence: Academic Licence Agreement for Czech National Corpus Data	Czech	This corpus contains traditional dialectological material, mostly unprepared monologue-type speech. The corpus is available download (upon request) and through the concordancer KonText. For a related publication, see Komrsková et al. (2018) .	Concordancer Download

CLARIN and Open Science



- Promoting the sharing and re-use of data through sustainable data registries
- All integrated datasets available in open access for research purposes
- Adherence to the FAIR data principles
 - Findable, Accessible, Interoperable, Re-usable
 - Interoperability through a common metadata framework
- Promotion of responsible data science
- Support for linguistic diversity
 - Data covering more than 1500 languages
 - Tools for many languages
 - Language resources in all modalities
- Strengthening the support for professional SSH researchers (> 500.000 in Europe)

[CLARIN: Towards FAIR and Responsible Data Science Using Language Resources.](#) In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, May 2018, 3259-3264.

Exercise 1 - Finding and processing data

- Go to clarin.eu
- Find the VLO
- Search for texts by Robert Louis **Stevenson**
 - what can you find?
 - which format is the corpus in?
 - where are they hosted?
- On the Links tab, use the three dots (...) to activate the Switchboard.
 - Explore with **Voyant** or
 - Process with **UDpipe**

CLARIN for knowledge sharing

CLARIN Knowledge Infrastructure



Knowledge Centres

Digital Humanities Course Registry

Tour de CLARIN

Teaching

Annual Conference

Funding

Video Channels

Best-Practice Papers



<https://www.clarin.eu/content/knowledge-infrastructure>

Knowledge Centres



About ▾ Language Resources ▾ Learn & Exchange ▾ Events News Contact

 CLARIN and Ukraine

Home / Learn & Exchange / Knowledge Centres

Knowledge Centres

CLARIN Knowledge Infrastructure

CLARIN Knowledge Centres (K-centres) are a cornerstone of the CLARIN Knowledge Infrastructure (KI), one of the main components ensuring a continuous transfer of knowledge between all players involved in the construction, operation and use of the infrastructure. The mission of the CLARIN KI is to ensure that the available knowledge and expertise does not exist as a fragmented collection of unconnected bits and pieces, but is made accessible in an organised way to both the CLARIN community and the social sciences and humanities research community more widely.

The Role of K-Centres

The focus of CLARIN is on language resources (in all modalities, from all regions and with any topical orientation) and K-centres serve researchers and educators from any discipline where language plays one of its many roles, ranging from object of study, a means of communication or expression, a means to store and extract information, object of learning or teaching activities, to training source for data-driven analytics, and many others. K-centres share their knowledge and expertise on one or more aspects of the domain covered by the CLARIN infrastructure and can be mostly found in CLARIN countries, but also exist elsewhere, and they all have a virtual presence.

Areas of K-Centre Expertise

K-centres all have their own specific areas of expertise, which can belong to many different categories, such as

<https://www.clarin.eu/content/knowledge-centres>

The Knowledge
Centres of CLARIN
can be contacted
through their
Help Desks

TA

Knowledge Centres

List of all 22 CLARIN K-centres with expertise in specific linguistic topics

Click on the full name of the K-centre to go to its landing page, and click on the acronym to see its full organisation details

<u>ACE</u>	<u>CLARIN Knowledge Centre for Atypical Communication Expertise</u>
Areas of competence	Atypical communication encompasses language and speech as encountered during (second) language acquisition and development, and in language disorders, but also more broadly in bilingual language development and in sign language. ACE is specialised in this type of research and concomitant infrastructural issues related to data acquisition, processing and sharing, which is typically highly characterised by sensitivity issues. For data storage and access the centre collaborates with MPI's TLA (The Language Archive) which is a CLARIN B Centre and also based in Nijmegen.
Audiences served	- linguists; - psychologists; - neuroscientists; - computer scientists; - speech and language therapists; - education specialists
Types of services	- how-to documents; - access to document templates; - Access to data; - Depositing; - FAQ; - Helpdesk; - Technical support
Is portal for language(s)	-
Other languages covered	-
Modalities covered	- Audio: speech; - Text; - Video: sign language
Linguistic topics	- Language acquisition (L1 and L2); - language disorders; - Language learning
Language processing	-
Data types	-
Resource families	- Spoken corpora; - Manually annotated corpora; - Multimodal corpora
Generic topics	- Critical Data Management; - Legal and ethical issues
Other keywords	- Language acquisition; - sign language; - language pathologies
Tour de CLARIN	Introduction Interview

<https://www.clarin.eu/content/knowledge-centres>

DIGITAL HUMANITIES COURSE REGISTRY

Committee ISO/TC 37/SC 04/WG 04 Lexical...
<https://livelink.din.de/livelink/livelink?func=ll&o...>

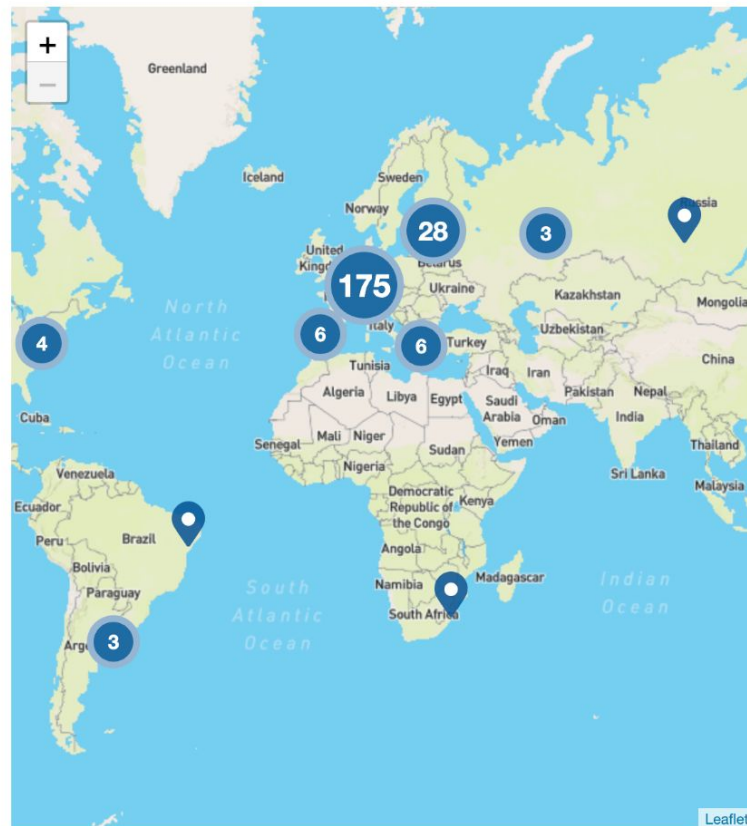
INFO



FILTER

SORTING

NAME	UNIVERSITY	PLACE	DATE	TYPE
MA in Digital Humanities	Lancaster University	Lancaster, United Kingdom	4 Oct ↻ 12 months	Master Programme
MSc Digital Heritage	University of York	York, United Kingdom	1 Oct ↻ 1 years	Master Programme
MSc Digital Archaeology	University of York	York, United Kingdom	1 Oct ↻ 1 years	Master Programme
MA in Digital Humanities	King's College London	London, United Kingdom	24 Sept ↻ 1 years	Master Programme
Introduction to Computational Linguistics (DHS)	University of Latvia	Riga, Latvia	7 Feb 2023 1 semester	Course
Introduction to Computational Linguistics	University of Latvia	Riga, Latvia	9 Feb ↻ 4 months	Master Programme
Media and Information	Universiteit van Amsterdam	Amsterdam, Netherlands	1 Sept ↻ 3 years	Bachelor Programme
Laboratorio di Informatica per Umanisti (Computing for the ...)	Unviersità degli Studi di Udine	Udine, Italy	11 Feb ↻ 10 weeks	Bachelor Programme
Digital Humanities Laboratory - Laboratorio di Informatica ...	University of Palermo	Palermo, Italy	2 months	Course
PhD in Humanities and Technologies	University of Macerata	Macerata, Italy	3 years	PhD Programme
Digital and Public Humanities	Università Ca' Foscari	Venice, Italy	1 Sept ↻ 2 years	Master Programme
First Level Master in Digital		Venice		



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

Impact Stories:

Showcases high-quality and innovative research that uses CLARIN tools and resources.

IceTaboo: Offensive Word Database with Commercial Application

The IceTaboo database can be used to flag contextually inappropriate words in texts, and is already being used as part of an automatic proofreading software by an Icelandic online news website.

Read more



<https://www.clarin.eu/content/clarin-impact-stories>

Video Lectures, training materials

Training Materials

Applied Language Technology

Author: Tuomo Hiippala

Faculty of Arts, University of Helsinki, Finland

Archilochus of Paros: Elegiac Fragments - XML Archive

Author: Anika Nicolosi and Beatrice Nava

University of Parma, Italy

Computational Morphology with HFST

Author: Erik Axelsson

Faculty of Arts, University of Helsinki, Finland

GATE, an Open-Source Toolkit for Natural Language Processing

Author: Diana Maynard

Faculty of Engineering, University of Sheffield

Introduction to Digital Humanities

Author: Zuzana Neverilova

Faculty of Arts, Masaryk University, Czech Republic

Introduction to Speech Analysis

Author: Mietta Lennes

Faculty of Humanities, University of Helsinki, Finland



Past CLARIN Cafés

TOPIC

CLARIN Café on Text+ : A New Research Data Initiative in Germany

Organised by Erhard Hinrichs (University of Tübingen), Thorsten Trippel (IDS Mannheim), and Andreas Witt (IDS Mannheim)

CLARIN Café: Towards Guidelines for Integrating CLARIN into Teaching

Organised by **Darja Fišer** (University of Ljubljana), Iulianna van der Lek
CLARIN *ERIC*

CLARIN Café on Text and Data Mining Exceptions in the Directive on Copyright in the Digital Single Market

CLARIN Café: Interactive Q&A Session for newcomers in CLARIN from the SSH domain

<https://www.clarin.eu/content/teaching-clarin>

<https://www.clarin.eu/content/clarin-cafe>

CLARIN Funding Hub

On this page you can find all CLARIN funding opportunities, as well as information about additional support that CLARIN can offer researchers preparing project proposals for EU-funded projects or organising virtual events. Below, you can see some recently approved, funded projects, along with a comprehensive list of all current and past projects funded by CLARIN.



CLARIN Calls

CLARIN offers funding to address strategic priorities that require international collaboration, exchange of expertise, training or mobility. If you would like to suggest a new topic for which funding may be beneficial, please get in touch.



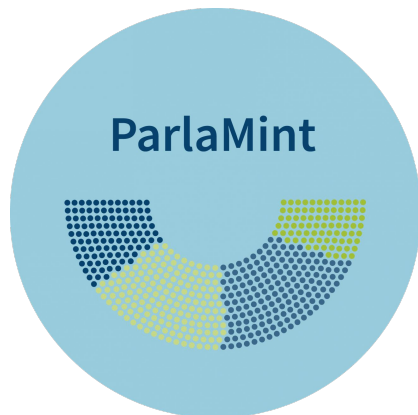
FAIR by design



Thanks to its network of **K-centres** and its **support instruments**, CLARIN not only allows researchers to share and find LRs, but also to create LRs that are

- FAIR by design
- interoperable
- comparable

CLARIN ParlaMint Project



Towards Comparable Parliamentary Corpora

17 european parliaments
including Senato della Repubblica italiana

Word list
Corpus: ParlaMint-GB 2.0 (British parliament)
Subcorpus: COVID_GB

Reference corpus: ParlaMint-GB 2.0 (British parliament)
Reference subcorpus: MPRegReference
Switch focus and reference subcorpus

Page 1

		ParlaMint-GB 2.0 (British parliament) : COVID_GB		ParlaMint-GB 2.0 (British parliament) : MPRegReference		
		frequency	frequency/mill	frequency	frequency/mill	Score
covid		9,667	412.9	0	0.0	413.9
pandemic		10,341	441.7	1	0.1	402.4
coronavirus		5,698	243.4	0	0.0	244.4
Covid		5,626	242.4	0	0.0	243.4
lockdown		5,571	237.9	2	0.2	199.1
furlough		2,129	90.9	0	0.0	91.9
PPE		1,564	66.8	0	0.0	67.8
virus		6,009	256.7	40	4.0	51.5
distancing		1,853	79.1	6	0.6	50.1
CHIS		860	36.7	0	0.0	37.7
lockdowns		666	28.4	0	0.0	29.4
Coronavirus		652	27.8	0	0.0	28.8
SAGE		572	24.4	0	0.0	25.4
tracing		1,048	44.8	8	0.8	25.4
isolate		1,579	67.4	20	2.0	22.8
quarantine		834	35.6	7	0.7	21.5
masks		908	38.8	9	0.9	20.9
vaccine		4,856	207.4	91	9.1	20.6
trace		1,748	74.7	27	2.7	20.5
shielding		492	21.0	1	0.1	20.0
Virtual		297	17.0	0	0.0	18.0
furloughed		296	16.9	0	0.0	17.9
Trace		389	16.6	0	0.0	17.6
Inaudible		320	15.8	0	0.0	16.8
COVID		363	15.5	0	0.0	16.5
Infection		1,697	72.5	36	3.6	16.0
Azzolina	107	31.0	1	0.0	30.9	--
virus	910	263.9	208	7.6	30.9	--
sierologici	100	29.0	1	0.0	28.9	--

<https://www.clarin.eu/content/parlamint-towards-comparable-parliamentary-corpora>

Helsinki Digital Humanities Hackathon 2021: 'Parliamentary Debates in COVID Times'

Contributors: Contributors: Isabella Calabretta, Courtney Dalton, Richard Griscom, Marta Kołczyńska, Matej Klemen, Kristina Pahor de Maiti, Ajda Pretnar Žagar, Ruben Ros

ZBIRKA
PARLAMENTARIA

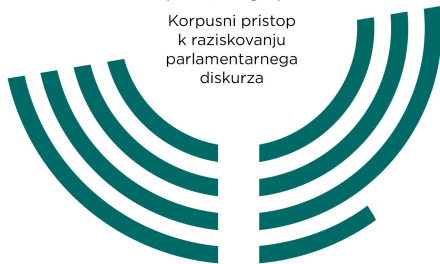
2

Darja Fišer
Kristina Pahor de Maiti

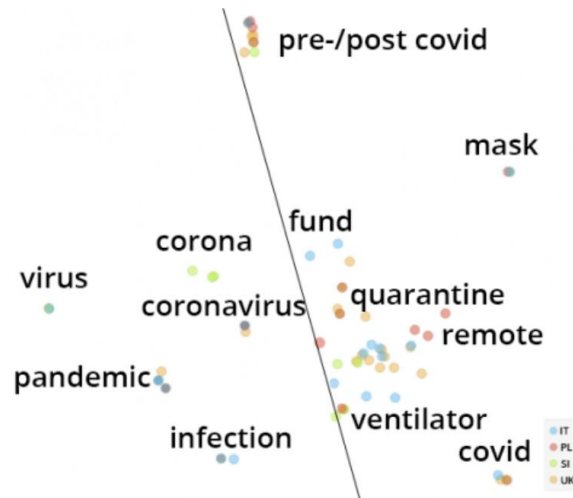
**Voices of the Parliament
A Corpus Approach to
Parliamentary Discourse Research**

»Prvič, sem političarka in
ne politik, drugič pa ...«

Korpusni pristop
k raziskovanju
parlamentarnega
diskurza



INSTITUT INSTITUTE
ZA NOVEJŠO ZDODOVINO OF CONTEMPORARY HISTORY



'Compiling a corpus is already a big project, so being able to skip this step was a huge privilege. Also, knowing that the corpus was granted permission to be included in the CLARIN repository already gives you some idea of its quality.'

Kristina Pahor de Maiti

<https://www.clarin.eu/impact-stories/helsinki-digital-humanities-hackathon-2021-parliamentary-debates-covid-times>

Exercise 2 - ParlaMint

- Explore the ParlaMint corpora using NoSketch Engine (public):
clarin.si/noske/index-en.html.
- Find the Italian corpus
 - Explore the **Corpus info** - how many subcorpora are there?
 - Create a wordlist for the COVID subcorpus
- Keyword extraction
 - extract **keywords** for the “COVID” subcorpus using the “REFERENCE” as **Reference (sub)corpus**

The CLARIN-IT national consortium

CLARIN-IT (2015-onwards)



About

Governance

Consortium

Centres

Join

Access

Events

Initiatives

News

Home

the Italian Common Language Resources and Technology Infrastructure



English



ONLINE EVENTS

[EUPORIA 2021 Webinar - Encoding a Critical Apparatus](#)



07/12/2020 - [Registration](#) to follow the Webinar via Zoom

CLARIN-IT Consortium



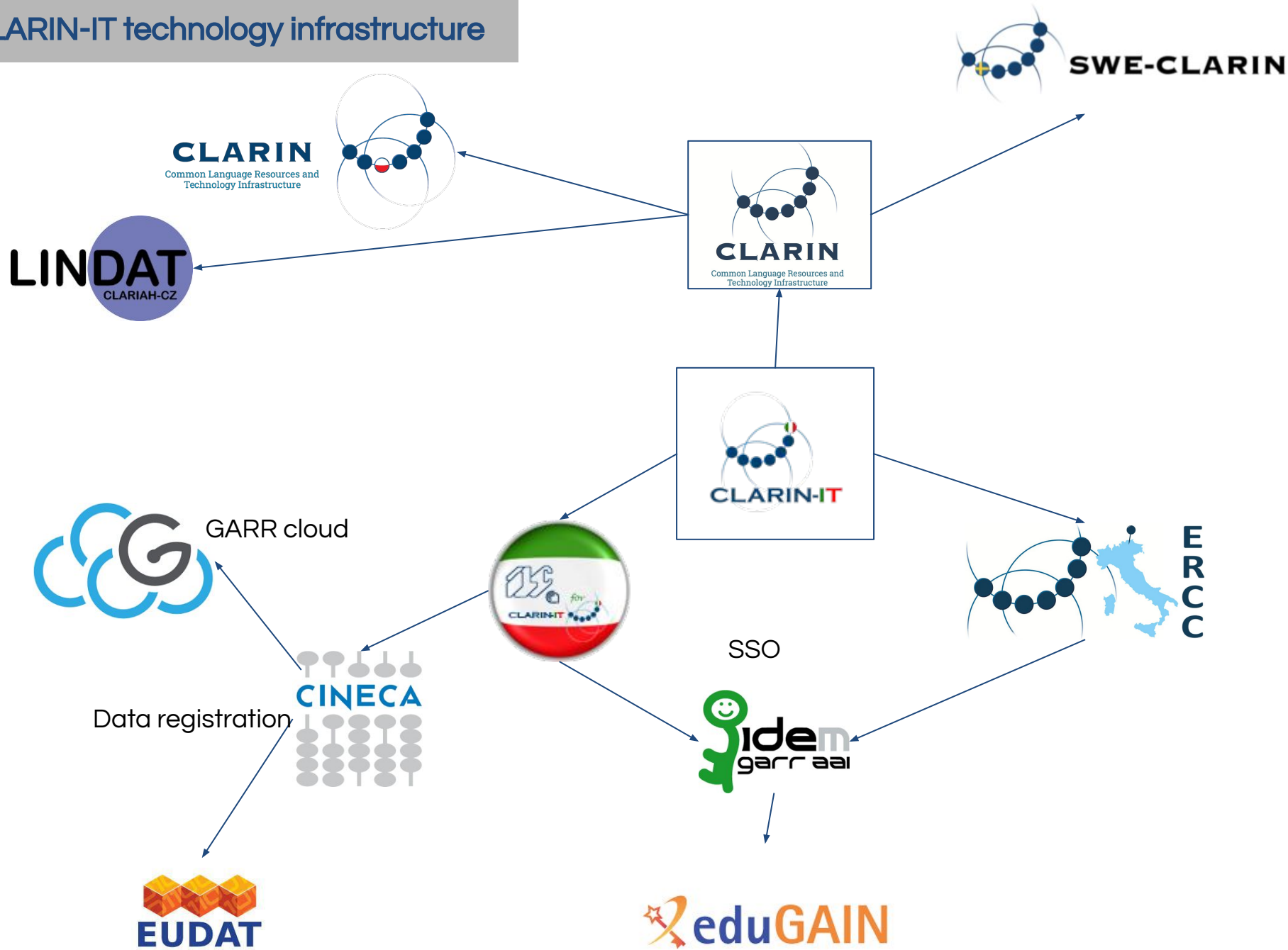
- The Department of Education, Human Sciences and Intercultural Communication of the University of Siena
- The Department of Philology and Literary Criticism of the University of Siena
- The Eurac Research Association (Bolzano)
- The Bruno Kessler Foundation (Trento)
- The Archival and Bibliographical Superintendence of Tuscany (Firenze)
- The Department of Electrical Engineering and Information Technology and the Interdepartmental Research Center "URBAN/ECO" of the University of Naples Federico II
- The Catholic University of the Sacred Heart (Milano)
-
- The University of Parma (Parma) has started the membership procedure.

CLARIN-IT Research Topics



- **Resources and Tools for the Italian Language**
 - create new resources by enriching existing corpora
 - lexical datasets with Linked Open Data
 - specialized corpora for computer-mediated communication.
 - natural language processing and analysis tools, offered as [web services](#) and integrated into [Weblicht](#).
- **Resources for Regional Languages and Multilingual corpora**
 - learner corpus for German, Italian and Czech,
- **Speech Archives**
 - Grafo, Caterina Bueno Archive
- **Digital Classics**
 - resources for Ancient Greek and Latin (LOD version of the TEI-dict Perseus Liddell-Scott Jones dictionary)
 - Italian Latinity of the Middle Ages
 - digital editions of ancient fragmentary texts

CLARIN-IT technology infrastructure



ILC4CLARIN



CLARIN
B CENTRE



[Chi siamo](#) [Organizzazione](#) [Repository](#) [Servizi](#) [Eventi CLARIN](#)



ILC4CLARIN

REPOSITORY

Easy to be found | Easy to be cited

ILC4CLARIN: data types



CLARIN-IT Repository About CLARIN

ItalWordNet v.2

Please use the following text to cite this item or export to a predefined format: BIBTEX CMDI

Roventini, Adriana; Marinelli, hosted at Institute for Computational Linguistics
<http://hdl.handle.net/11362/44444>

Share

ALIM

Home

Authors

Item identifier

Project URL

Demo URL

Date issued

Type

Size

Language

Description

Bibliography

- ▶ Entrate
- ▶ Elenco
- ▶ Elenco
- ▶ Ricerca
- ▶ Ricerca
- ▶ Documenti
- ▶ Segnalazioni

ARCHIVIOVi.Vo.
Conservazione e diffusione degli archivi orali e audiovisivi

REGIONE TOSCANA
Superintendenza Archivistica e Filologica della Toscana
UNIVERSITÀ DI SIENA
CASENTINO

ITA
EN

CLARIN-IT: i servizi



Editing lessicale



LexO

Editor Web collaborativo utilizzato per la creazione e la gestione di risorse lessicali e terminologiche (multilingue) come risorse di dati collegati.

[Approfondisci](#)

Annotazione/analisi del testo



Freeling IT

Eseguibile di freeling solo per la lingua italiana, up to tagged.

[Prova questo strumento](#)



DeSR

Parser shift/reduce di dipendenze per la lingua italiana.

[Prova questo strumento](#)



LTFW (Linguistic Tools For Weblicht)

Porting Java del tokenizzatore basato su perl sviluppato all'interno del progetto OpeNER.

[Prova questo strumento](#)

Merging



LMF Merger

Merger LMF di due lessici italiani.

[Approfondisci](#)



LMF ML Merger

Merger LMF multilivello.

[Approfondisci](#)



Multiword Extractor: Average Frequency Filter

Estrattore di parole multiple che applica un filtro di frequenza media su MWE.

[Prova questo strumento](#)



Multiword Extractor: Merge Overlapping Strings

Estrattore di parole multiple che applica un filtro che unisce stringhe sovrapposte.

[Prova questo strumento](#)



Multiword Extractor: Remove Substrings

Estrattore di parole multiple che applica un filtro che rimuove le sottostringhe.

[Prova questo strumento](#)



Multiword Extractor: Reorder

Estrattore di parole multiple che applica un filtro che riordina i dati ExtractorMW per coppia_frequenza.

[Prova questo strumento](#)

Lexical services



ItalWordNet

Parola: Mostra tutte le relazioni ☐

casa, Nome

- [1] - edificio o parte di esso in cui si abita;
([abitazione](#) [1], [casa](#) [1], [dimora](#) [2], [magione](#) [1], [ostello](#) [2], [tetto](#) [4])
- [2] - edificio con particolari funzioni.
([casa](#) [2])
- [3] - dinastia, casa regnante; l'insieme dei sovrani, appartenenti a una stessa famiglia, che si succedono sul trono
([casa](#) [3], [casa regnante](#) [1], [dinastia](#) [1])
"la casa di Savoia non ha regnato a lungo sull'Italia"
- [4] - la famiglia alla quale si appartiene
([casa](#) [4])
"sta pensando di mettere su casa"
- [5] - ditta, compagnia, in particolare nel settore dell'editoria e della moda
([casa](#) [5])
"casa editrice"
- [6] - casella nel gioco degli scacchi
([casa](#) [6])

NLP services



URL del servizio: [freeling_it](#) (WSDL)

i [Come usare il servizio](#)

Prova questo servizio attraverso il form sottostante.



Run service

Inputs

Report

input

☐ *as URL*
☒ *direct data or local file*
 No file chosen

language

it.cfg ▼

multiword

yes ☐ no ☒

ner

--Use Default-- ▼

output_format

--Use Default-- ▼

mandatory
optional

Easy to Cite



Author

- Anonymus (79)
- Nahli, Ouafae (31)
- Khalfi, Mustapha (30)
- Zarghili, Aرسالane (30)
- Multiple Authors (11)
- ... View More

[Advanced Search](#)

Subject

- Latin (413)
- Middle Ages (405)
- Prosa (318)
- Fonti Letterarie (310)
- Storiografia (82)
- ... View More

Language (1)

- Latin (413)
- English (1)
- Arabic (3)
- Italian (24)
- Ancient Greek (to 1453) (10)
- ... View More

Search

Sign in to ILC for CLARIN-IT Repository
 Login via Your home institution (e.g. university)

ILC4CLARIN: findability, access



Virtual Language Observatory Search Contributors Help

VLO / Faceted search / Search results

ALIM

Showing 1 to 10 of 355 results within selection for **ALIM** **Latin** Results per page: 10

Use the categories below to limit the search results to those matching the selected value(s).

Language
Latin

Collection

Resource type

Format

<< < 1 2 3 4 5 6 7 8 9 10 > >>

Chronicon Vulturense, III
(Part of **ALIM** Literary Sources)
compilato da **ALIM** HomePage del progetto: <http://it.alim.unisi.it/il-progetto/> Documentazione: <http://alim.unisi.it/documentazione>
Latin
Landing page for this record

Chronicon Vulturense, I
(Part of **ALIM** Literary Sources)
compilato da **ALIM** Edizione in formato TEI XML, livello ALIM2_1, a cura di Jan Ctibor HomePage del progetto: <http://it.alim.unisi.it/il-progetto/> Documentazione: <http://alim.unisi.it/documentazione>
Latin

Chronicon Vulturense, III

Please use the following text to cite this item or export to a predefined format:

BIBTEX CMDI

Iohannes Monachus, 2006, *Chronicon Vulturense, III*, ILC-CNR for CLARIN-IT repository hosted at Institute for Computational Linguistics "A. Zampolli", National Research Council, in Pisa, <http://hdl.handle.net/20.500.11752/OPEN-432>.

Share: [f](#) [t](#) [g+](#)

OPEN

Authors	Iohannes Monachus
Item identifier	http://hdl.handle.net/20.500.11752/OPEN-432
Project URL	http://it.alim.unisi.it
Demo URL	http://it.alim.unisi.it/di/resource/594
Date issued	2006-11-29
Type	corpus
Language(s)	Latin
Description	compilato da ALIM HomePage del progetto: http://it.alim.unisi.it/il-progetto/

ILC4CLARIN: interoperability, reuse



Resources

nn_nrhz120_1848.TEI-P5.xml 106.14 KiB



Mediatype

text/plain

Language

German

Matching Tools

Group by task ☒

Search for tool

▼ Constituency Parsing



Open

WebLicht Const Parsing DE

▼ Dependency Parsing



Open

Spacy (hosted by D4Science) - DE



Open

UDPipe



Open

WebLicht Dep Parsing DE

▼ Distant Reading



Open

Voyant Tools

K-Centres & DH: IMPACT



ABOUT ▼ TOOLS & RESOURCES ▼ KNOWLEDGE ▼ COMMUNITY ▼ CKC HELPDESK ▼

Tools & Resources

SEARCH OVER

0

TOOLS FOR TEXT DIGITISATION



CitAttest

Attesting Word Forms in Dictionary Citations. With this **tool**, occurrences of a headword of a historical dictionary are automatically marked in the quotations belonging to that headword in the dictionary.

MORE TOOLS ➔

LEXICA FOR

0

DIFFERENT LANGUAGES



Diasearch – Diachronic corpus search service

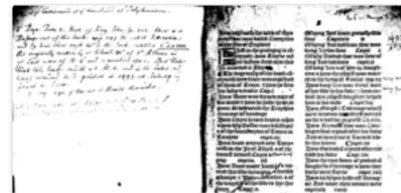
Diasearch is an online service which enables users to perform linguistically enriched queries on a collection of historical texts.

MORE LEXICA ➔

ACCESS

0

IMAGE AND GROUND TRUTH



VIDA DE LAZARILLO DE TORMES

High quality images with **ground truth** associated from the collection of Biblioteca Nacional de España.

MORE RESOURCES ➔

IMPACT is
focused on
digitization

<https://www.digitisation.eu>

K-Centres & DH: DiPText-KC

[ABOUT](#)[PARTNERS](#)[PEOPLE](#)[KNOWLEDGE](#)[HELPDESK](#)[EVENTS](#)[NEWS](#)[CONTACT](#)

DiPText-KC

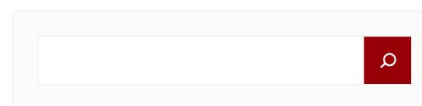
CLARIN Knowledge Centre for Digital and Public Textual Scholarship

DiPText-KC offers expertise on methods, data, instruments and technologies relevant in the field of Philological and Literary Studies, History, Art History and Cultural Heritage.

Its actions aim at:

- sharing information with scholars and students about the state of the art in digital scholarly editing and text annotation through domain-specific languages;
- supporting scholars and students in the creation and publication of digital scholarly editions and resources;
- organizing training activities (for instance webinars, workshops and summer schools).

DiPText-KC is one of the Centres of [CLARIN-IT](#), the Italian node of [CLARIN](#) (Common Language Resources and Technology Infrastructure), a digital infrastructure of pan-European interest identified by [ESFRI](#) (European Strategy Forum on Research Infrastructures) and classified as a Landmark Research Infrastructure for the Social Sciences and Humanities (ESFRI Landmarks SSH RI).



HIGHLIGHTS



The Digital and Public Textual Scholarship Knowledge Centre is focused on **digital philology**

<https://diptext-kc.clarin-it.it>

Activities of the DiPText-KC

[ABOUT](#)[PARTNERS](#)[PEOPLE](#)[KNOWLEDGE](#)[HELPDESK](#)[EVENTS](#)[NEWS](#)[CONTACT](#)

Consortia, Associations, Centers

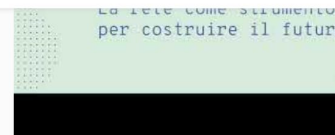
- Consortia
 - [Unicode Consortium](#)
 - [TEI Consortium](#)
- National and International Associations
 - [ADHO](#)
 - [EADH](#)
 - [AIUCD](#)
- DH Centres and Labs
 - Italy
 - [CIRCSE](#)
 - [LabCD](#)

Training

- Summer Schools
 - Venice Centre of Digital and Public Humanities, Department of Humanities, Ca' Foscari University of Venice
 - [Summer Camp 2020](#)
 - University of Pisa
 - [Digital Tools for Humanists 2022](#)
(past editions: [2021](#) | [2020](#) | [2019](#) | [2018](#) | [2017](#))
 - [Digitising, Cataloguing, Searching and Sharing the Medieval and Early-Modern Image](#)
- Seminars / Webinars
 - [VeDPH Seminars in Digital and Public Humanities – January-May 2022](#)
(past editions: [October 2019 – May 2020](#) | [September 2020 – December 2021](#))
 - [Humanities Horizons – History, Hacktivism and Genetic Criticism, Solstice Seminar in DPH](#)
 - [The Public Staging of Gender in Shakespearean Theatre Discussion with Pamela Allen Brown](#)

Digital Libraries

- Zotero Collections
 - [DiPText-KC Library](#)
 - [CLARIN Library](#)



CNR-ILC CoPhiLab @ GARR 2022

The Collaborative and Cooperative Philology Lab (CoPhiLab, CNR-ILC): data, applications, services and infrastructures (5:50:11-5:59:00)

BREAKING NEWS

› Fourth Appointment of the Workshop Cycle
"Digital Philology meets Computational Linguistics"

› Third Appointment of the Workshop Cycle
"Digital Philology meets Computational Linguistics"

› Concluded the First Cycle of the Permanent Seminar Series "A bridge between two worlds"

› CNR-ILC CoPhiLab @ GARR 2022

› Second Appointment of the Workshop Cycle
"Digital Philology meets Computational Linguistics"

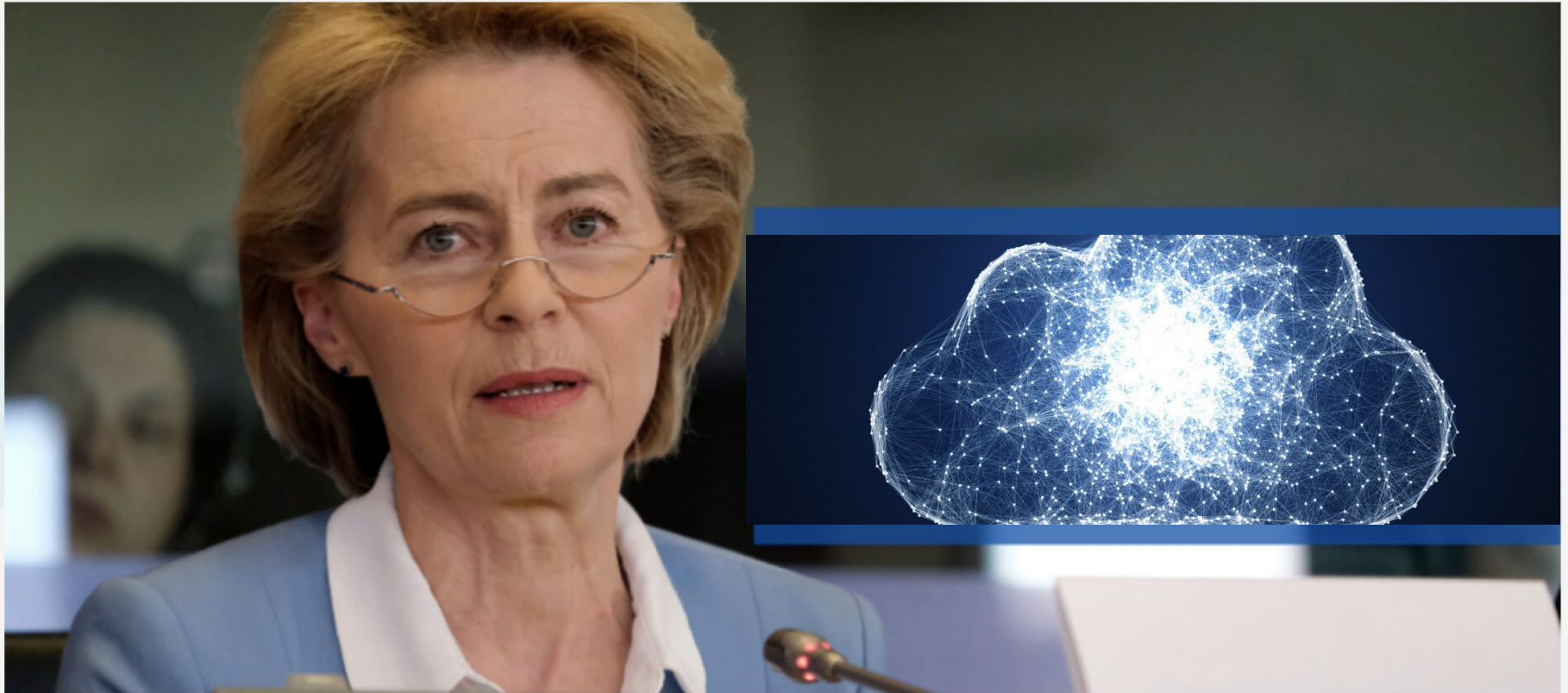
The Digital and Public Textual Scholarship Knowledge Centre keeps you informed on Consortia, Associations, Centres, Training Schools, and Digital Libraries relevant for digital philologists

<https://diptext-kc.clarin-it.it>

Exercise 3 - Explore ILC4CLARIN

- Find the ILC4CLARIN repository
 - Which is the most represented language in terms of records?
 - What kind of data can you find in Arabic?
 - How do you cite CophiWordNet?
- Try to log in with your institutional identifier, using the Login function (top right)

A Cluster of SSH Research Infrastructures



22 Jan 2020

EC President Ursula von der Leyen talks EOSEC in Davos

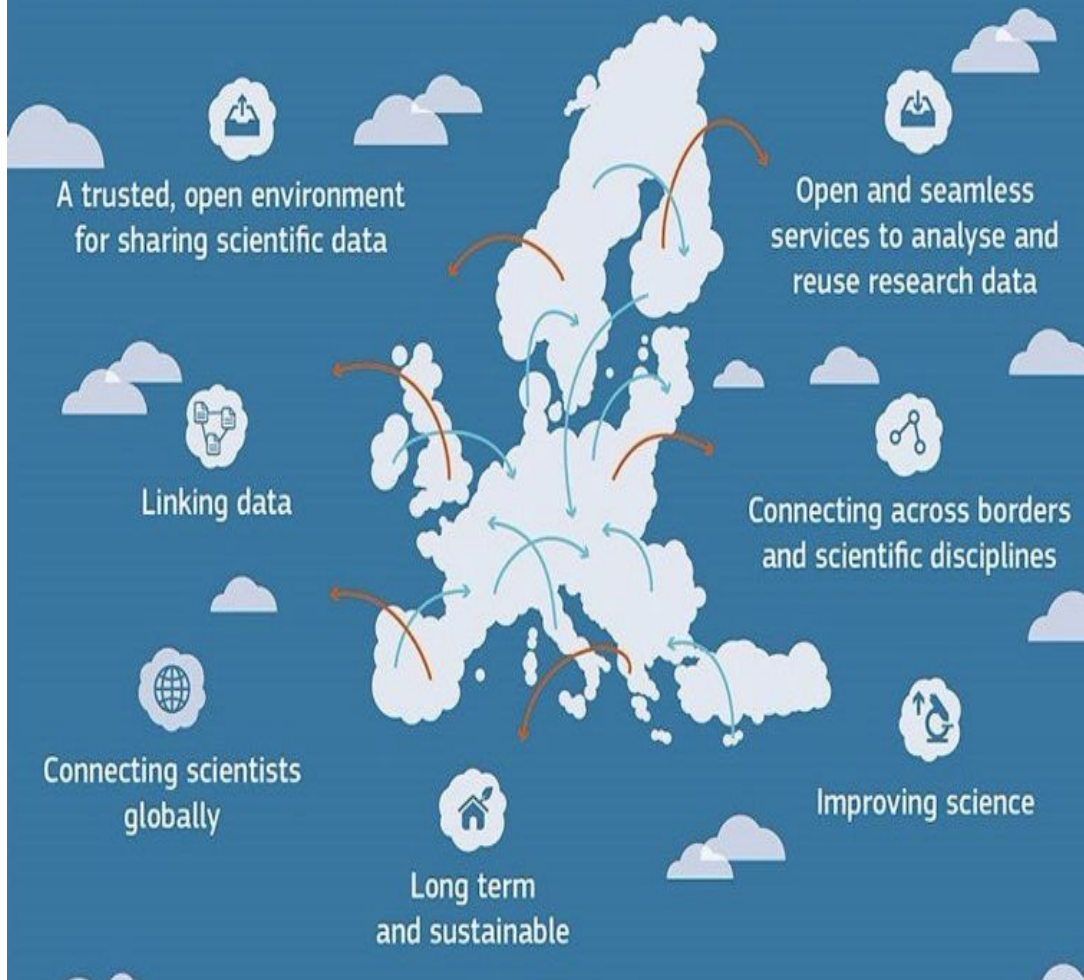
the open and trusted environment
for accessing and managing research data and related services

EOSC: a challenge



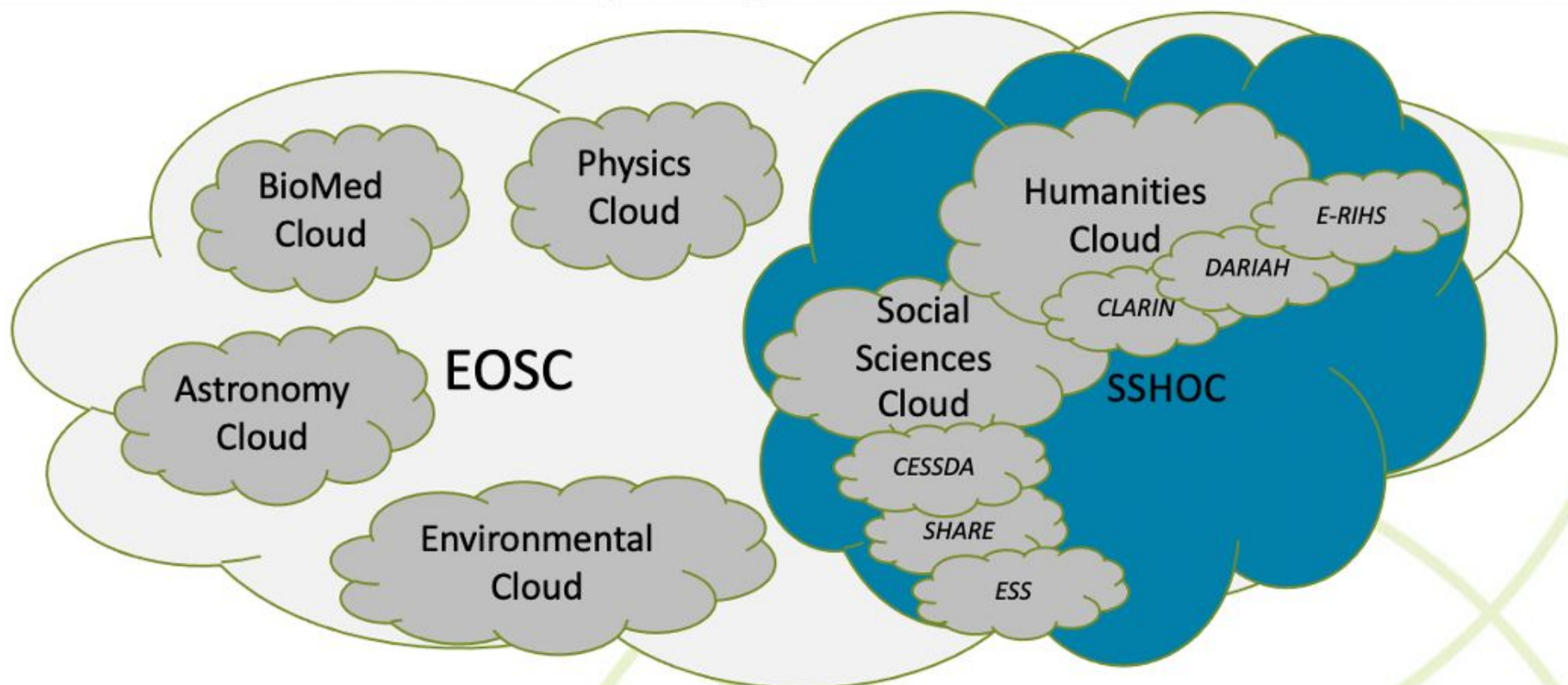
EUROPEAN OPEN SCIENCE CLOUD

BRINGING TOGETHER CURRENT AND FUTURE DATA INFRASTRUCTURES



- The European Union has launched the European Open Science Cloud as a supporting landscape to foster open science and open innovation
- The EOSC is aimed at removing technical, policy and human barriers, leading to knowledge creation and economic prosperity in Europe.

The SSH Open Cluster





Realising the **Social Sciences and Humanities** for the European Open Science Cloud

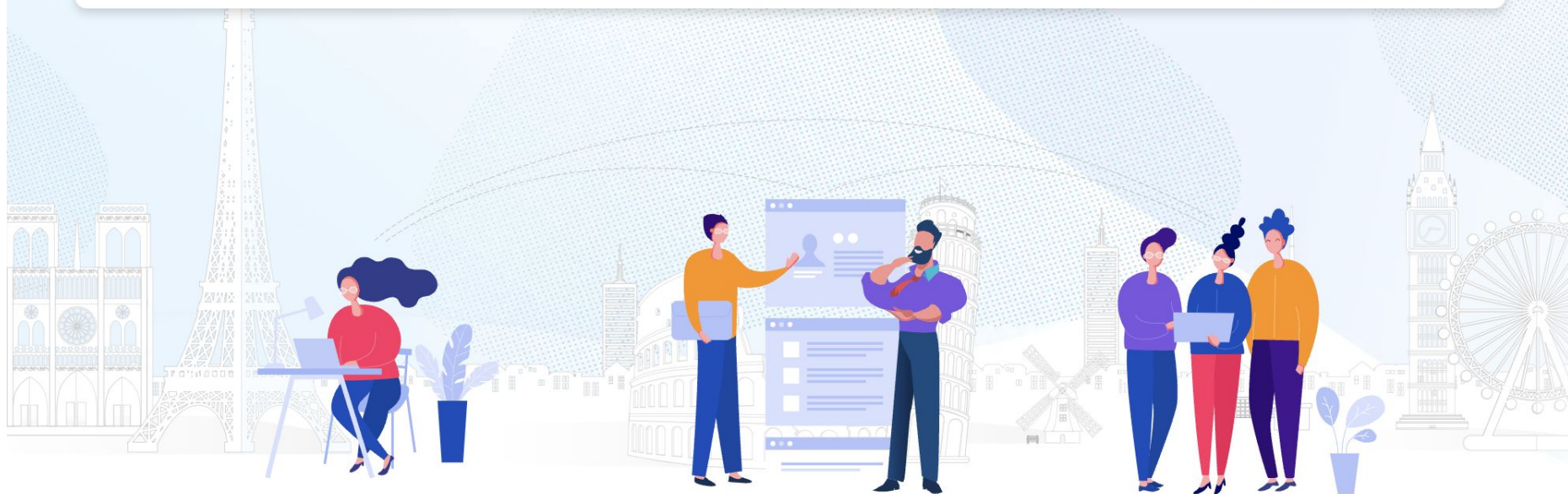
Social Sciences & Humanities Open Marketplace

Discover new and contextualised resources for your research in Social Sciences and Humanities: tools, services, training materials, workflows and datasets. [Read more...](#)

All categories



Search



Search results (5361)

Refine your search [Clear filters](#)

Sort by Relevance ▼

CATEGORIES ▲

☐ Tools & services

1731

☐ Training materials

340

☐ Publications

2953

☐ Datasets

310

☐ Workflows

27

ACTIVITIES ▲

☐ Analyzing


641

☐ Visual Analysis

314

☐ Context Analysis

250




COLLADA Schema Version 1.5.0

[Link icon](#)

No description provided.

[Read more](#)



COLLADA -- Digital Asset Schema Release 1.5.0 - Specification

[Link icon](#)

This document describes the COLLADA schema. COLLADA is a COLLABorative Design Activity that defines an XML-based schema to enable 3D authoring applications to freely exchange digital assets without loss of information, enabling multiple software packages to be combined into extremely...

[Read more](#)

Exercise 4 - The SSHOC Marketplace

- Explore the training materials on the SSHOC Marketplace
- Can you find anything that can be useful for your research?
- If so, can you tell us why?

CLARIN for researchers



- Contact CLARIN for help with your research
- Visit this page
 - <https://www.clarin.eu/content/clarin-researchers>
- Participate in CLARIN (virtual) events
 - <https://www.clarin.eu/events>
- Tour de CLARIN
 - <https://www.clarin.eu/Tour-de-CLARIN>
- Use CLARIN Training and videolectures

Contacts



- CLARIN ERIC
 - Newsletter <https://www.clarin.eu/news>
 - @CLARINERIC
- CLARIN IT
 - www.clarin-it.it
 - @CLARIN_IT
 - coordination@clarin-it.it