

CLARIN-IT

the Italian Common Language Resources and Technology Infrastructure



CLARIN ERIC and CLARIN-IT

Monica Monachini - *National Coordinator CLARIN-IT*

Francesca Frontini - *CLARIN Board of Directors*

Istituto di Linguistica Computazionale - ILC CNR

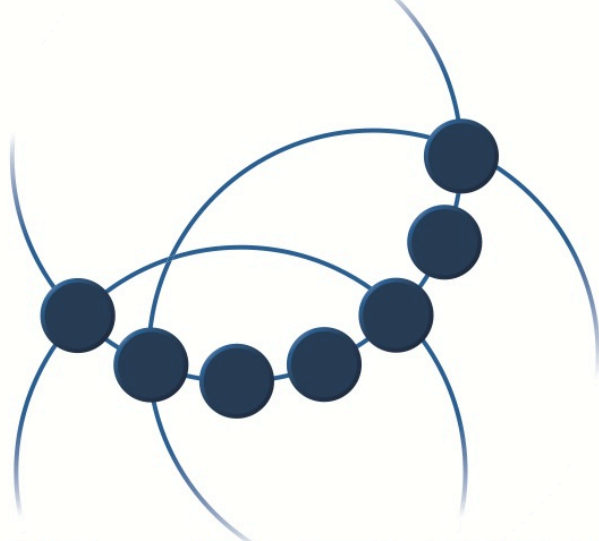
08/06/2021

Scienza aperta e gestione dei dati per le scienze umane e del patrimonio culturale: corso di formazione a distanza



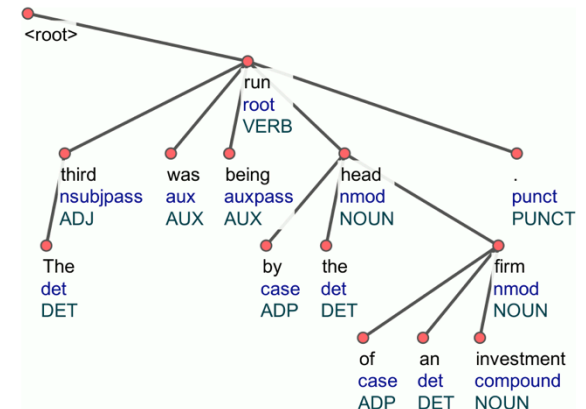
- CLARIN ERIC - di cosa si tratta?
- Cosa offre CLARIN-IT?
- Alcuni esercizi per mostrarvi le varie possibilità e i servizi che offre CLARIN
 - link in chat al documento
 - <https://docs.google.com/document/d/1XZoacpC1wd0YISKIJSLIFKTYNxo6AmYG>

CLARIN ERIC



CLARIN

Common Language Resources and
Technology Infrastructure



CLARIN in due parole



- detiene lo status di **ESFRI ERIC** dal 2012 e **Landmark** dal 2016
- attraverso modalità intuitiva e sostenibile, fornisce agli studiosi delle **scienze umane e scienze sociali** accesso a:
 - dati linguistici digitalizzati (in formato scritto, parlato, video o multimodale)
 - **strumenti avanzati** per scoprire, esplorare, annotare, analizzare o confrontare/combinare tali dati a prescindere di dove si trovino
 - attraverso un sistema di controllo di accesso **single sign-on**
- costituisce un ecosistema ideale per la condivisione di conoscenze
- è parte integrante dell' **European Open Science Cloud**
 - See clarin.eu/eosc

CLARIN: dati e comunità



- 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 oggi



- **21 membri:** (AT, BG, CY, CZ, DE, DK, EE, FI, GR, HR, HU, IS, IT, LT, LV, NL, NO, PL, PT, SE, SI)
- **3 observers:** FR, UK, ZA
- **> 68 centri (25 centri certificati CTS)**



accesso federato



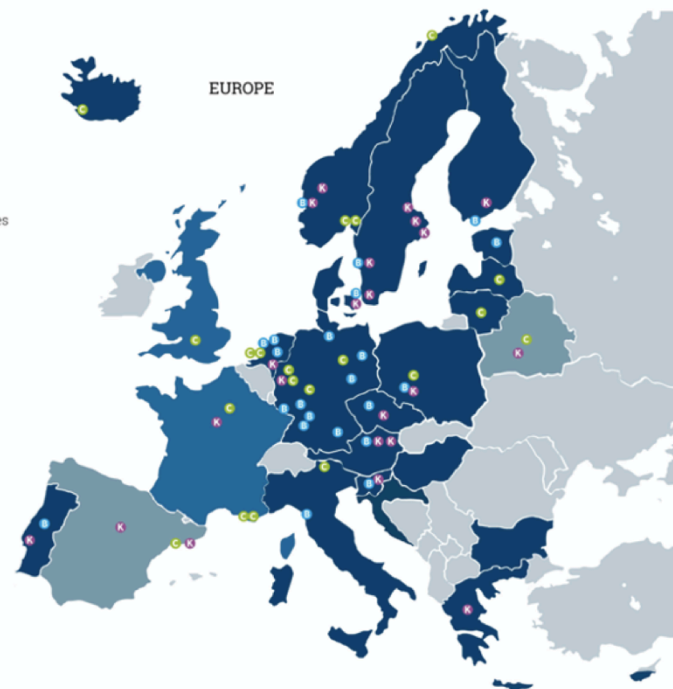
raccolta centralizzata di metadati



servizi concatenati



- ERIC members
- Observers
- Countries with participating centres
- Centre Providing Data
- Centre Providing Metadata
- Knowledge Centre



CLARIN

Common Language Resources and
Technology Infrastructure



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!

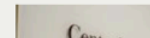
Events

[More](#)



Lancaster
Symposium
on Innovation
in Corpus
Linguistics
2021

Wednesday 23
June, Online



Corpus



Read more about CLARIN-UK

Data & Tools

[More](#)



CorCenCC
 Corpws Cenedlaethol Cymraeg Cyfoes - the National Corpus of Contemporary Welsh


- #LancsBox
- BNC
- CKLD
- CLAWS
- CLiC
- CorCenCC
- CQPweb
- ELAR
- GATE
- GATE Cloud
- Hansard at Huddersfield

Read more about CLARIN-UK


[More](#)

Events

[More](#)



Lancaster Symposium on Innovation in Corpus Linguistics 2021
 Wednesday 23 June, Online



Corpus Linguistics

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

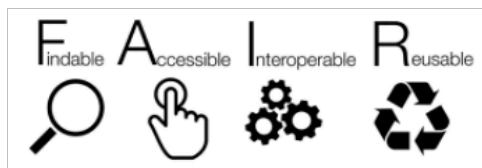
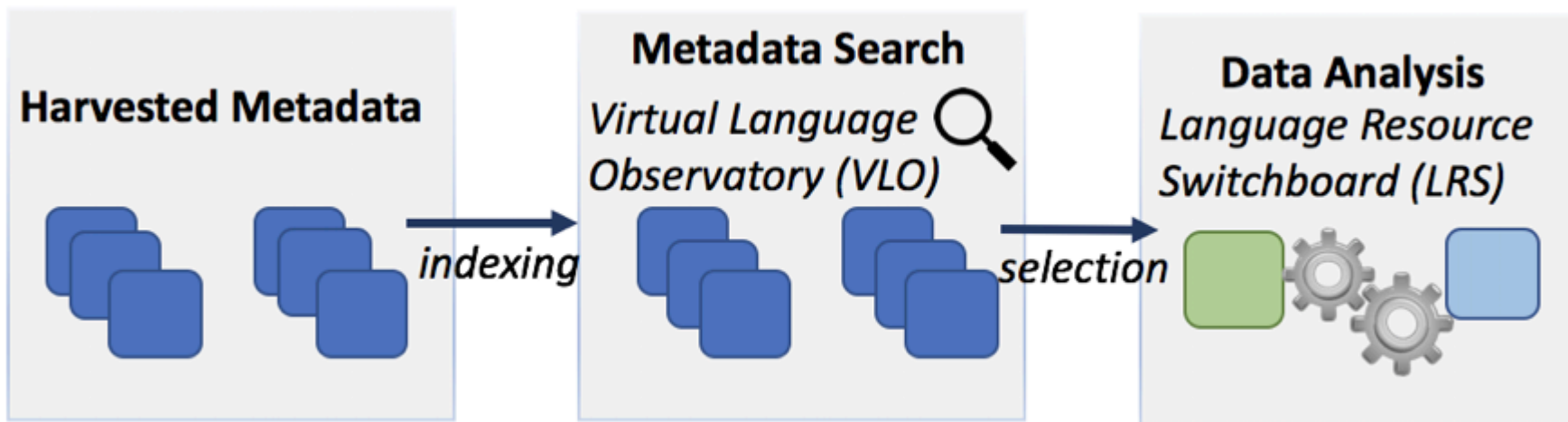
Spoken corpora in the CLARIN infrastructure

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

<https://www.clarin.eu/resource-families>

L'infrastruttura tecnica



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



- IMDI Corpora
- Lund Corpora
 - Eline Visser
 - ESST
 - Eye-Tracked Frog Stories
 - LACOLA
 - LANG-KEY
 - LUNDIC
 - REaChES
 - SpaceH
 - Strömqvist-Richthoff
 - Swedia2000
 - Tactile Reading
 - Test
 - ThaiSweVideo
 - The Barack Obama Corpus**
 - the_barack_obama_corpus_information.txt
 - 2009
 - 2009
 - 2010
 - 2010
 - 2011
 - 2011
 - 2012
 - 2012
 - 2013
 - 2013
 - 2014
 - 2014
 - 2015
 - 2015
 - 2016
 - 2016
 - USE
 - VOKART

METADATA SEARCH

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>



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

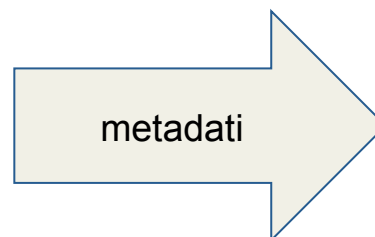
(Part of Lund University Humanities Lab)



 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](http://corpora.humlab.lu.se)



SWE-CLARIN

vlo.clarin.eu

UDPipe

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

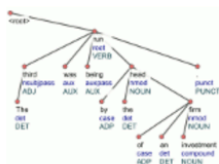
BIBTEX

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 a 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



> UDPipe



> WebLicht Dep Parsing DE



> WebLicht Dep Parsing EN



Esercizio 1

- Andate sul sito clarin.eu
- Cercate il VLO
- Cercate i testi di Robert Louis **Stevenson**
 - Cosa avete trovato?
 - In quale formato è disponibile il corpus?
 - dove sono ospitati/depositati?
- Nella sezione “Links”, cliccate su “(...)” per attivare la funzione Switchboard:
 - Esplorate con **Voyant**
 - Procedete con **UDpipe**

CLARIN Knowledge Infrastructure



Knowledge centres

CLARIN Resource Families

VideoLectures

Digital Humanities Course Registry

Workshops

Trainer Network Programme

Training Suite



<https://www.clarin.eu/content/clarin-for-researchers>
<https://www.clarin.eu/content/knowledge-sharing>

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>

Tour de CLARIN



- Quanto detto non esaurisce la ricchezza dei servizi CLARIN
- Le visite e i blog del Tour de CLARIN permettono di scoprire le risorse dei vari centri

Tour de CLARIN



Tour de CLARIN is a CLARIN *ERIC* initiative that aims to periodically highlight prominent User Involvement (UI) activities of a particular CLARIN national consortium. The highlights include an interview with one or more prominent researchers who are using the work of national consortium's infrastructure and can tell us more about their experience with CLARIN in general; one or more use cases that the consortium is particularly proud of and any relevant user involvement activities carried out. Tour de CLARIN helps to increase the visibility of the national consortia, reveal the richness of the CLARIN landscape, and to display the full range of activities throughout the network. The content is disseminated via the [CLARIN Newsflash](#), [blog posts](#) and [Twitter](#).

<https://www.clarin.eu/Tour-de-CLARIN>

Tour de CLARIN



Overview [Copy Resource Link](#) [Switch to Expert-View](#)

Protokoll der 34. Sitzung des Kongresses von Aachen Digitale Edition

Type(s): [acdh:Resource](#)

PID: <https://hdl.handle.net/21.11115/0000-000C-2018-5>

Creator(s): [Karin Schneider](#)

Contributor(s): [Stephan Kurz](#)

Available Date: 16 Jul 2019

Extent: 75028

Binary Size: 75028

Category: Dataset

Schema: <https://www.tei-c.org/release/xml/tei/schema/relaxng/tei.rng>

License: [Karin Schneider](#)

License: Attribution 4.0 International (CC BY 4.0)

Owner: [Karin Schneider](#)

Access Restriction: public

Identifier(s): <https://id.acdh.oeaw.ac.at/cmdii/619>, <https://arche.acdh.oeaw.ac.at/api/619>, <https://id.acdh.oeaw.ac.at/uuid/5439795c-9758-5a0e-d8b5-61933a463660>, https://id.acdh.oeaw.ac.at/maechtekongresse/editions/Aachen_Prot_34.xml, <https://hdl.handle.net/21.11115/0000-000C-2018-5>

Part of: Editionsdaten

[BibLaTeX service](#) [Switchboard](#) [Custom TEI to HTML transformation](#) [Thumbnail service](#) [RDF access](#)

[RAW CMDI access](#) [Download](#) [CMDI2HTML access](#) [GUI access](#)

Tour de CLARIN: A CLARIN B-centre in Vienna: The Austrian Centre for Digital Humanities and Cultural Heritage with ARCHE

Publication date: 21 April 2021

Austria • Tour de CLARIN • Tour de CLARIN: B-centres



The CLARIN B-centre ARCHE (A Resource Centre for the Humanities) is central to ACDH-CH's mission of fostering the change towards the digital paradigm in the humanities.



Matching Tools

▼ Distant Reading



[Open](#) Voyant Tools

<https://www.clarin.eu/blog/tour-de-clarin-clarin-b-centre-vienna-austrian-centre-digital-humanities-and-cultural-heritage>

CLARIN e la scienza aperta



- Promozione della condivisione e riutilizzo dei data attraverso dei registri sostenibili
- Tutti i dataset sono disponibili in accesso libero per scopi di ricerca
- Aderenza ai principi dei dati FAIR
 - Findable, Accessible, Interoperable, Re-usable (reperibile, accessibile, interoperabile, ri-utilizabile)
 - Interoperabilità attraverso un sistema di metadati comune
- Promozione di una scienza responsabile
- Supporto alla diversità linguistica
 - I dati riguardano più di 1500 lingue
 - Strumenti per diverse lingue
 - Risorse Linguistiche disponibili in varie modalità
- Rafforzamento del supporto per professionisti e ricercatori nelle scienze sociali e materie umanistiche (> 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.

Progetti infrastrutturali



PARTHENOS

Pooling Activities, Resources and Tools
for Heritage E-research Networking,
Optimization and Synergies



Domande?

Esercizi 1 e 2

Pausa

CLARIN-IT - Consorzio

CLARIN-IT (dal 2015)



- 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](#)



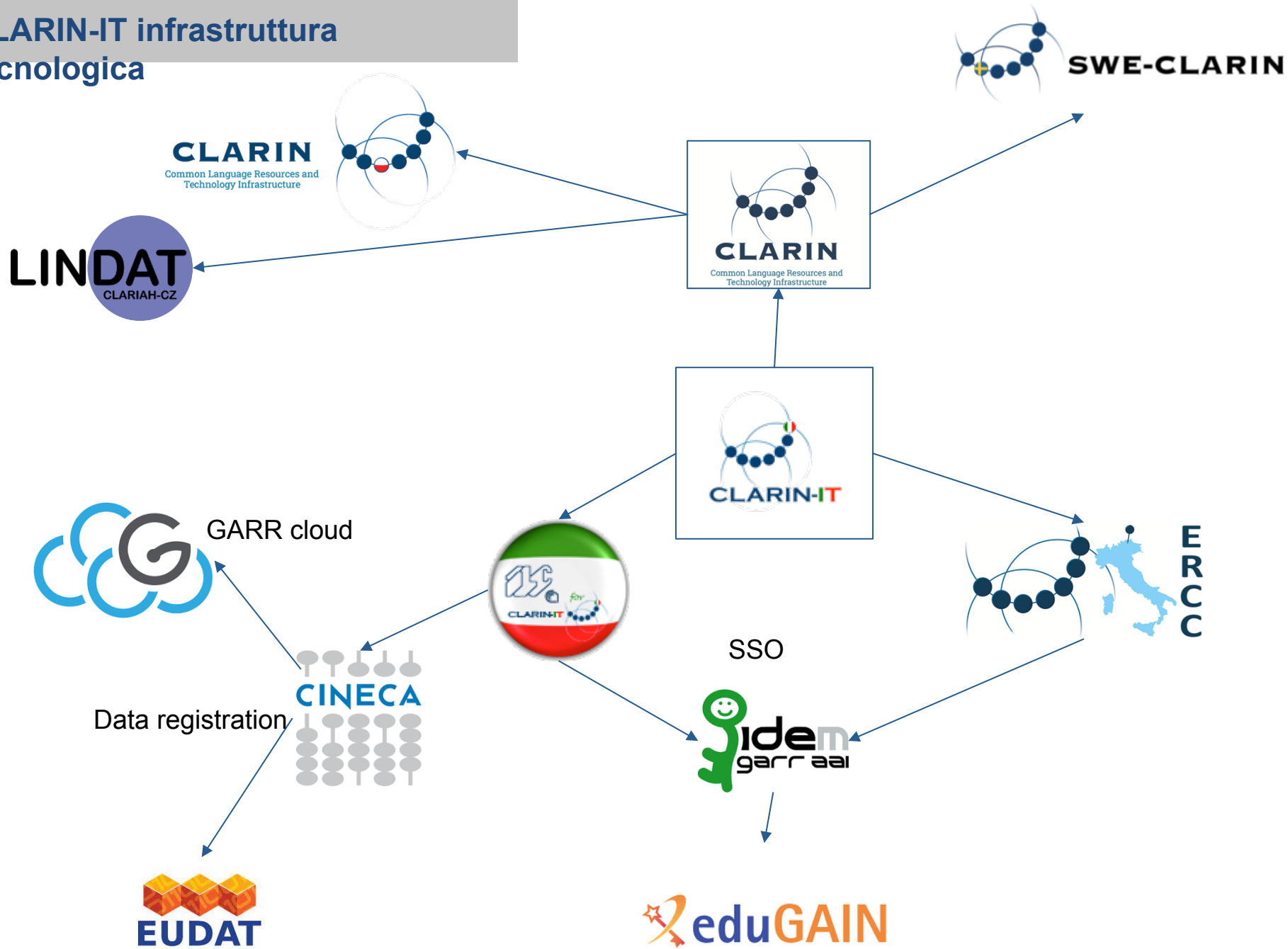
07/12/2020 - [Registration](#) to follow the Webinar via Zoom

CLARIN-IT Consorzio



- Il Dipartimento di scienze della formazione, scienze umane e della comunicazione interculturale presso l'Università degli Studi di Siena
- Il Dipartimento di Filologia e Critica Letteraria presso l'Università degli Studi di Siena
- L'associazione Eurac Research (Bolzano)
- La fondazione Bruno Kessler (Trento)
- La Soprintendenza Archivistica e Bibliografica della Toscana (Firenze)
- Il Dipartimento di Ingegneria Elettrica e delle tecnologie dell'informazione e Il centro interdipartimentale di ricerca "URBAN/ECO" de l'Università di Napoli Federico II
- L'Università Cattolica del Sacro Cuore (Milano) ha avviato la procedura per diventare membro.

CLARIN-IT infrastruttura tecnologica



CLARIN-IT Tematiche di ricerca



- **Risorse e Strumenti per la Lingua Italiana**
 - Creazione di nuove risorse attraverso l'arricchimento dei corpora esistente
 - Creazione di lessici da Linked Open Data
 - Corpora specializzati per la comunicazione computer-mediated.
 - strumenti di elaborazione e analisi del linguaggio naturale, offerti come [servizi web](#) and integrati in [Weblicht](#).
- **Risorse per Italiano regionali e corpora multilingue**
 - Corpora di apprendenti (learner corpus) per il tedesco, l'italiano e il ceco.
- **Archivi Orali**
 - Grafo, Archivio Caterina Bueno
- **Classici digitalizzati**
 - Risorse per [greco antico](#) e [latino](#) (versione LOD del dizionario TEI-dict Perseus Liddell-Scott Jones)
 - Latinità durante il medioevo italiano
 - Edizioni digital di antichi frammenti

ILC4CLARIN



CLARIN
B CENTRE



CLARIN-IT CLARIN

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



ILC4CLARIN

REPOSITORY

Easy to be found | Easy to be cited

ILC4CLARIN: risorse



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 Comput
<http://hdl.>

ALIM

Share

Home

Authors

Item iden

Project U

Demo UR

Date issu

Type

Size

Language

Descript

Biblio

▶ Entra n

▶ Elenco

▶ Elenco

▶ Ricerca

▶ Ricerca

▶ Docum

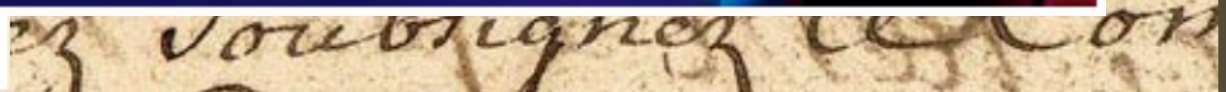
▶ Segnalazioni


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



ITA
EN

Q



ILC4CLARIN: deposito



CLARIN
B CENTRE



CLARIN-IT CLARIN

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



[OPEN](#)

[Data, Tools and Services from other Italian institutions](#)

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 Weblight)

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](#)

Servizi lessicali



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 \[6\]](#))

Servizi TAL



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

i [Come usare il servizio](#)

Prova questo servizio attraverso il form sottostante.



Run service

Inputs

input

<input type="text"/>	<input type="radio"/> <i>as URL</i>
<input type="text"/>	<input checked="" type="radio"/> <i>direct data or local file</i>
<input type="button" value="Choose file"/> No file chosen	

language

it.cfg ▾

multiword

yes no

ner

--Use Default-- ▾

output_format

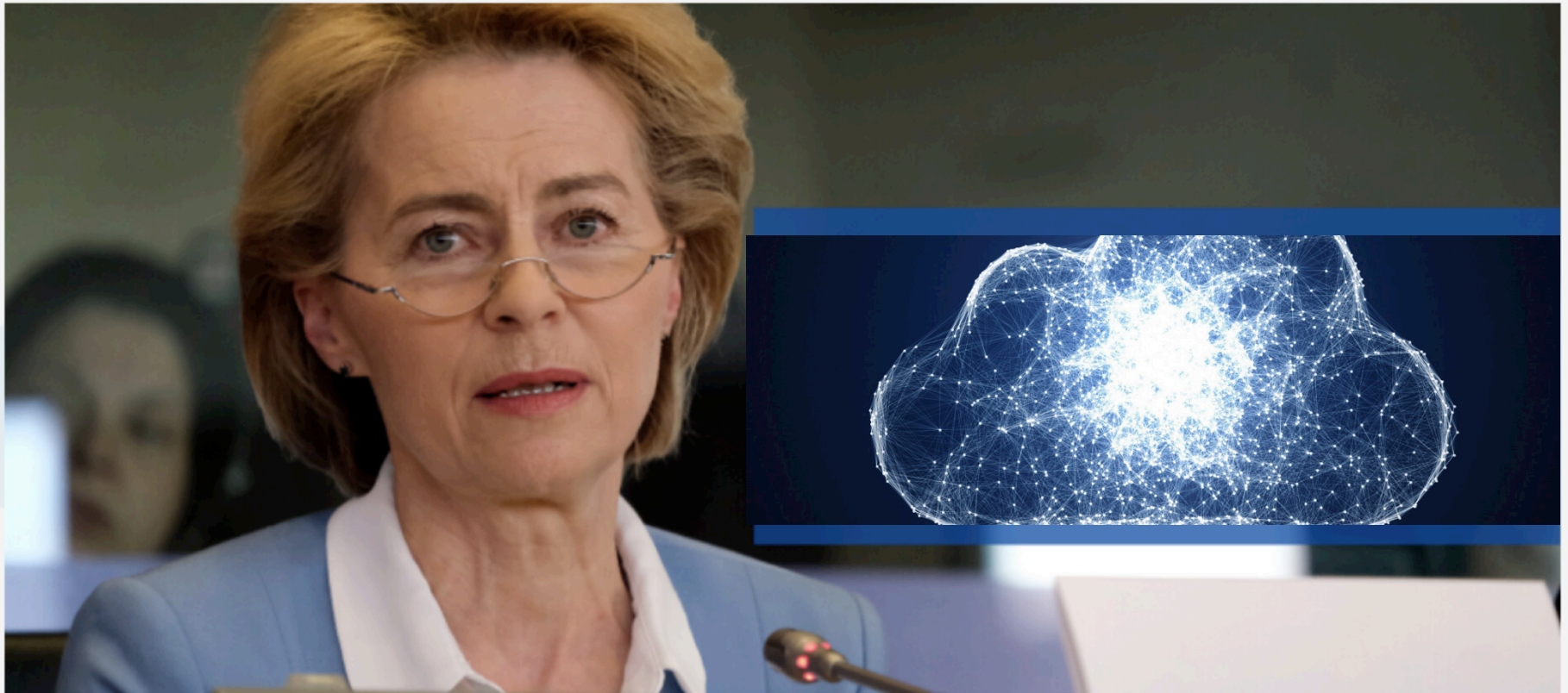
--Use Default-- ▾

Report

mandatory
optional

Esercizio 2

- Cercate il deposito/repository ILC4CLARIN
 - Quale è la lingua più rappresentata in termini numero di voci?
 - Che tipo di dati sono disponibili in Arabo?
 - Come si cita CophiWordNet?
- Provate ad accedere con il vostro profilo istituzionale attraverso la funzione Login in alto a destra



22 Jan 2020

EC President Ursula von der Leyen talks EOSSC in Davos

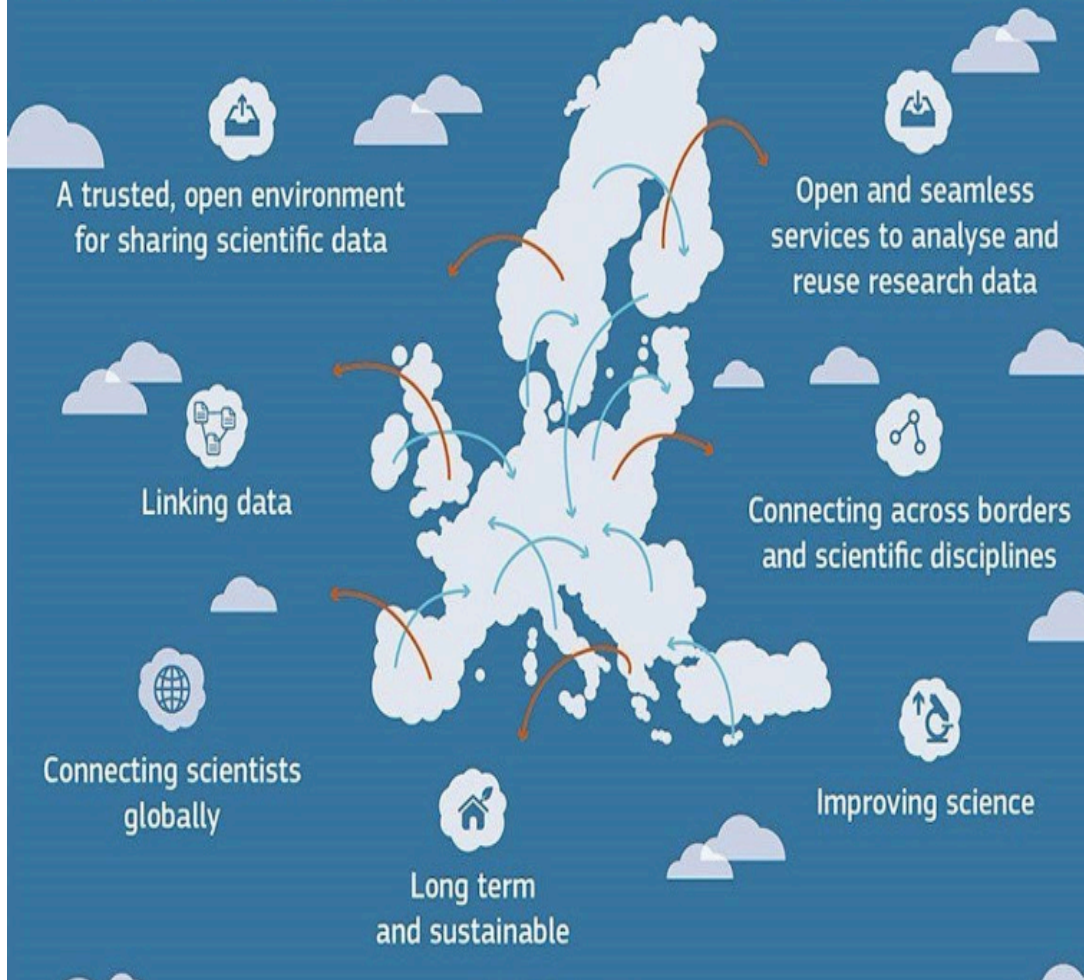
**the open and trusted environment
for accessing and managing research data and related services**

La sfida dell' EOSC



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.

EOSC: i progetti preparatori



The creation of EOSC is aimed at removing technical, policy and human barriers, leading to knowledge creation and economic prosperity in Europe.

The European Open Science Cloud is intended to federate existing scientific data infrastructures that are now spread across disciplines and EU member states. will make access to scientific data easier and more efficient This



The [EOSCpilot project](#) is supporting the first phase in the development of the EOSC. The project brings together stakeholders from research infrastructures and e-Infrastructure providers and will engage with funders and policy makers to propose and trial EOSC's governance framework.



The [EOSC-hub project](#) Jan 2018 aims to create the integration and service management structure of the European Open Science Cloud enabling an open access to research resources from a myriad of scientific disciplines via a digital hub. The hub will act as an entry point for researchers and innovators to discover, access, and use a variety of advanced data-driven resources.

EOSC: i progetti cluster



ENVIRONMENT



PHOTON and NEUTRON



ASTRONOMY and
ASTROPARTICLES



SOCIAL SCIENCES
and HUMANITIES



BIOLOGY and
MEDICINE

Le Scienze Umane & Sociali nell'EOSC



SSHOC



INFRAEOSC-04-2018 – EU Cluster projects – Connecting ESFRI infrastructures through Clusters



19 beneficiaries, 42 organizations, 14.5ME, 1759 PMs, Duration 40 months started 1-1-2019



Main consortium partners: **SSH ERICS**: CESSDA (coordinator), **CLARIN**, **DARIAH**, **SHARE**, **ESS**, the **E-RIHS** network and their partner organizations

Impatti di SSHOC



The Social Sciences and Humanities are seamlessly integrated in the European Open Science Cloud



Availability of an EU wide, easy to use SSH Open Marketplace, where tools and data are openly available



EU wide availability of high quality "cloud ready" SSH tools and high quality SSH data



EU wide availability of trusted and secure access mechanisms for SSH data, conforming to EU legal requirements



State of the art advanced through dedicated SSH data pilots cluster projects



Data sharing is accepted practice (the "new normal") among the different SSH communities

TRIPLE: in sintesi



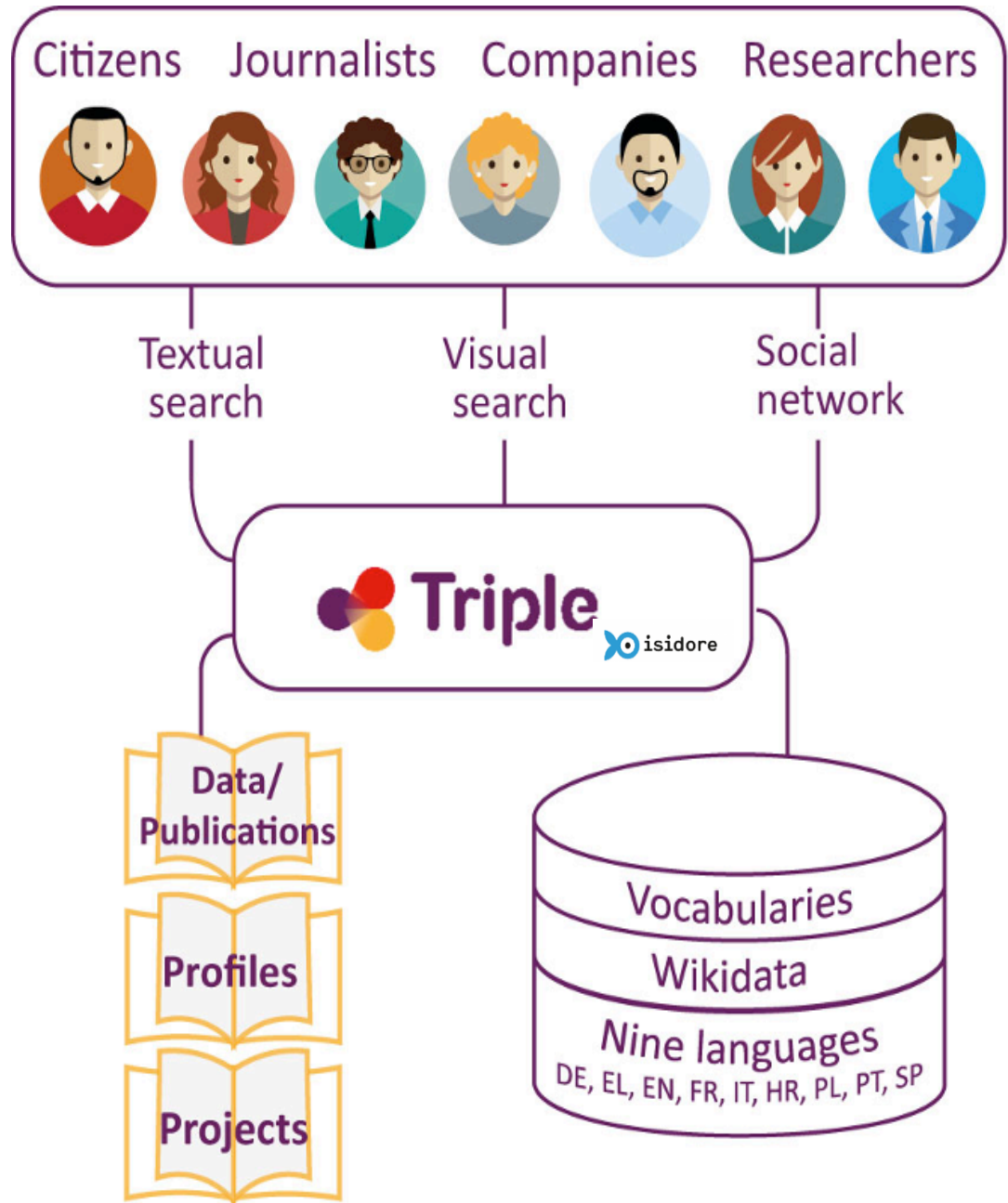
- launched on 1 October 2019
- Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration
- duration of 42 months (2019–2023)
- Horizon 2020 framework with approx. 5.6 million Euros
- consortium of 19 partners from 13 countries
- Coordinated by France by Huma-Num, (CNRS)
- currently 87 staff members

GOTRIPL: la piattaforma



- is the heart of the project and will be released as a prototype in autumn 2021;
- is an innovative **multilingual and multicultural discovery solution** for the social sciences and humanities (SSH);
- will provide a **single access point** that allows you to explore, find, access and reuse materials such as literature, data, projects and researcher profiles at European scale;
- is based on the **Isidore search engine** developed by Huma-Num (unit of CNRS)
- will be one of the dedicated **services of the EOSC Market Place** for the social sciences and humanities in the European Research Area.

GOTRIPLE: Platform (contents, processes, services)



CLARIN-ERIC Opportunità



<https://www.clarin.eu/content/funding-opportunities>

CLARIN ParlaMint



Towards Comparable Parliamentary Corpora

17 european parliaments including Senato della Repubblica italiana

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

Word list
 Corpus: ParlaMint-IT 2.0 (Italian parl)
 Subcorpus: COVID

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

Page 1

word	ParlaMint-GB 2.0 (British parliament) - COVID_GB		ParlaMint-GB 2.0 (British parliament) - MPregReference		Score
	frequency	frequency/mill	frequency	frequency/mill	
covid	2,667	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	2,329	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
Covid-19	860	36.7	0	0.0	37.7
Covid	666	28.4	0	0.0	29.4
coronavirus	652	27.8	0	0.0	28.8
lockdown	572	24.4	0	0.0	25.4
mascherine	1,048	44.8	8	0.8	25.4
recovery	1,579	67.4	20	2.0	22.8
WES	834	35.6	7	0.7	21.5
distanzamento	908	38.8	9	0.9	20.9
tamponi	4,856	207.4	31	3.1	20.6
fund	1,748	74.7	27	2.7	20.5
contagi	492	21.0	1	0.1	20.0
generation	327	17.0	0	0.0	18.0
mascherina	396	16.9	0	0.0	17.9
DPCM	389	16.6	0	0.0	17.6
FIBP	320	15.8	0	0.0	16.8
Azzolina	363	15.5	0	0.0	16.5
virus	1,697	72.5	36	3.6	16.0
sierologici	31.0	1	0.0	0.0	30.9
	263.9	208	7.6	30.9	
	29.0	1	0.0	28.9	

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

Esercizio 3

- Esplorate i corpora ParlaMint utilizzando la piattaforma NoSketch Engine (public): clarin.si/noske/index-en.html.
- Selezionate il corpus italiano
 - Guardate le info del **Corpus** - quanti subcorpora sono disponibili?
 - Create una lista di parole/wordlist per il subcorpus - COVID
- Estrazione di parole chiavi
 - Estraiete le parole chiave per il subcorpus “COVID” utilizzando il pulsante “REFERENCE” as **Reference (sub)corpus**

CLARIN per i ricercatori



- Contattate CLARIN essere assistenza e aiuto per la vostra ricerca
- Visitate la pagina
 - <https://www.clarin.eu/content/clarin-researchers>
- Prendere parte agli eventi CLARIN (anche virtuali)
 - <https://www.clarin.eu/events>
- Tour de CLARIN
 - <https://www.clarin.eu/Tour-de-CLARIN>
- Utilizzate il CLARIN Training and le registrazioni video.

Contacts



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

- CLARIN IT
 - www.clarin-it.it
 - @CLARIN_IT
 - coordination@clarin-it.it