

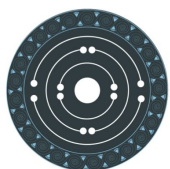
# CLARIN e la scienza aperta



- 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



**EOSC-hub**



**ERIC  
FORUM**



**SSHOC**  
social sciences & humanities open cloud



**Triple**



**ENRIITC**



**EOSC FUTURE**

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](#)



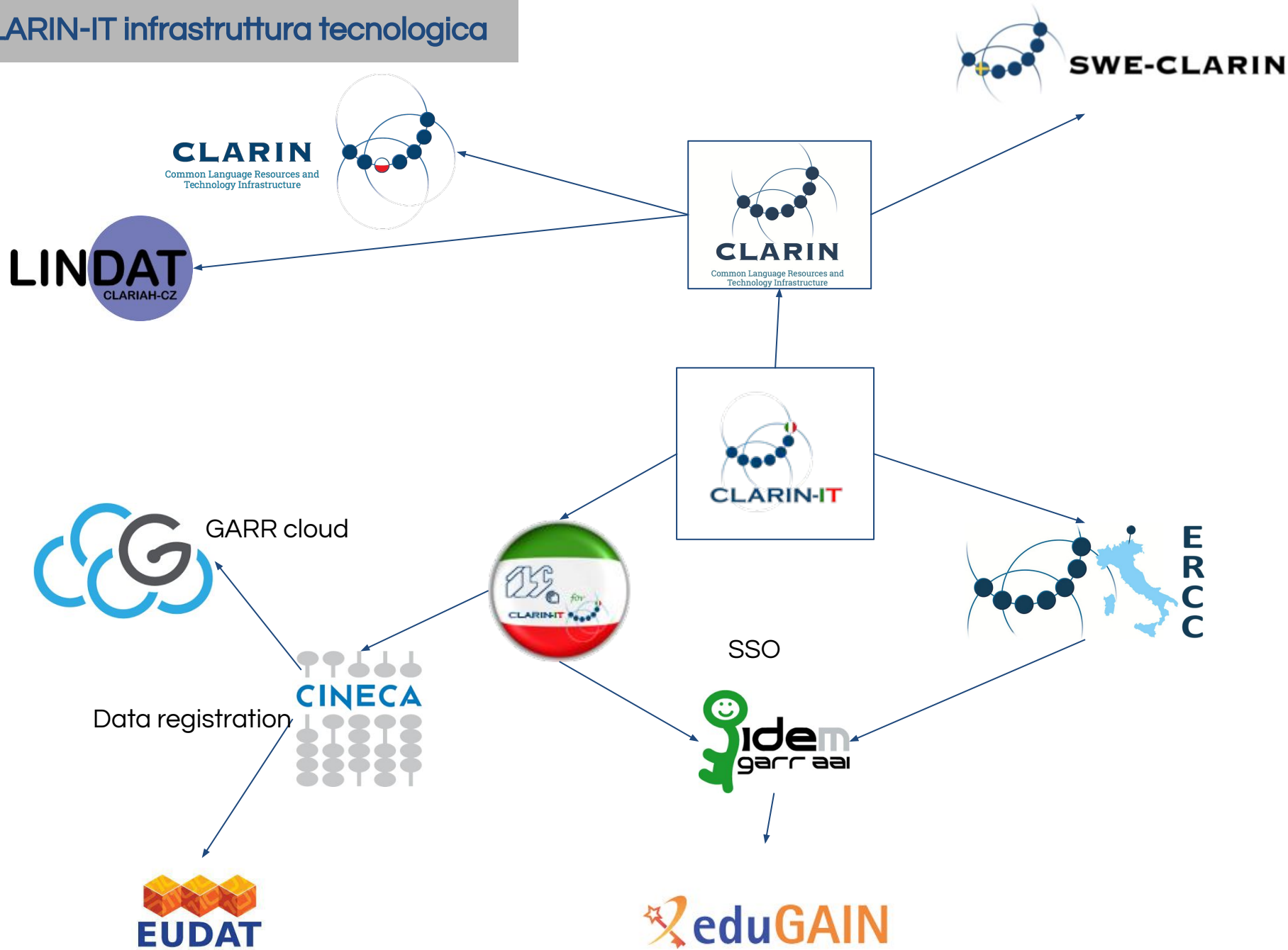
07/12/2020 - [Registration](#) to follow the Webinar via Zoom

# 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" dell'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**



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



# ILC4CLARIN

## REPOSITORY

Easy to be found | Easy to be cited

# ILC4CLARIN: risorse



Repository

About



## 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

Home

Authors

Item identifier

Project URL

Demo URL

Date issued

Type

Size

Language

Description

Bibliography

Entrata

Elenco

Elenco

Ricerca

Ricerca

Documenti

Segnalazioni

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



ITA  
EN

# ILC4CLARIN: deposito



**CLARIN**  
**B CENTRE**



[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 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](#)

# 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])



URL del servizio: [freeling\\_it](#) (WSDL)

## **i** Come usare il servizio

Prova questo servizio attraverso il form sottostante.



Run service

### Inputs

**input**

☐ as URL

☒ direct data or local file

Choose file No file chosen

**language**

it.cfg ▾

**multiword**

yes ☐ no ☒

**ner**

--Use Default-- ▾

**output\_format**

--Use Default-- ▾

Reset fields

### 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