

CLARIN-IT

The Italian Common Language Resources and Technology Infrastructure



CLARIN-IT: l'infrastruttura di ricerca per le scienze umane e sociali

Monica Monachini - CLARIN Italian National Coordinator

ILC-CNR
National Representative

1° Ottobre 2015



- L'Italia diventa membro della Infrastruttura CLARIN-ERIC
*Common Language Resource Infrastructure for Social Sciences
and Humanities*



- Una opportunità in più per chi si occupa di Discipline Umanistiche

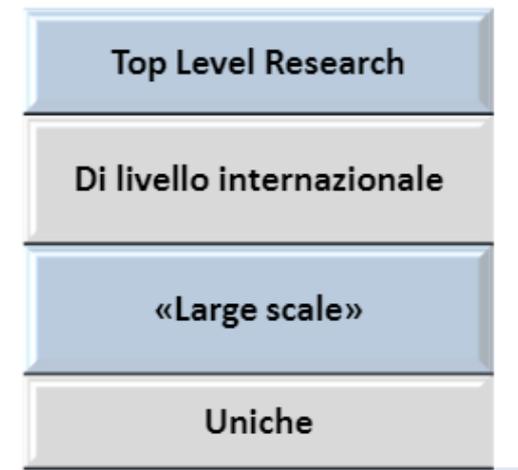
Cosa si intende per IR



- *Insieme di risorse strumentazioni, competenze, dati, servizi, funzioni specializzate e dedicate a realizzare in modo continuativo i flussi necessari alla ricerca scientifica e tecnologica*
- CLARIN è una infrastruttura di tipo ERIC formalmente riconosciuta
 - completato la European Strategy Forum on Research Infrastructures (ESFRI) roadmap
- ha raggiunto lo stato di "land-mark"



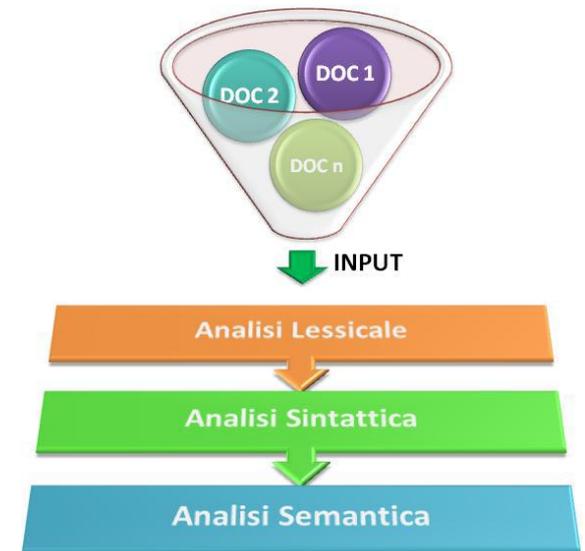
Research Infrastructure (ESFRI)



A cosa servono le Risorse Linguistiche



- A cosa servono:
 - studio della lingua
 - analisi di fonti testuali
 - analisi automatica del linguaggio
 - annotazione automatica dei testi
 - ...



Situazione e bisogni



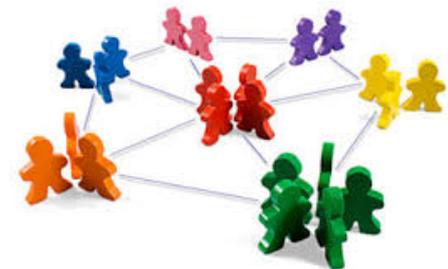
- Le risorse spesso risiedono presso l'istituzione di appartenenza del produttore; spesso relegate agli scopi e ambiti scientifici di un progetto; gravate da vincoli



- Costruire queste risorse richiede tempo e denaro:
 - cruciale garantire un facile accesso alle risorse per i ricercatori di diverse istituzioni e paesi
 - fondamentale garantire che tali risorse rimangano fruibili nel tempo



- Un cambio di paradigma:
 - condivisione, accesso, replicabilità dei risultati

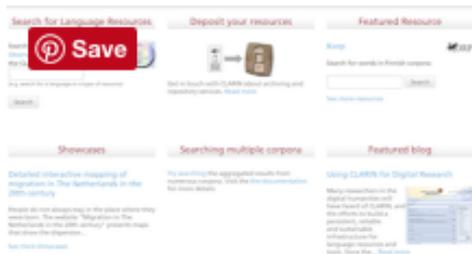


CLARIN-ERIC



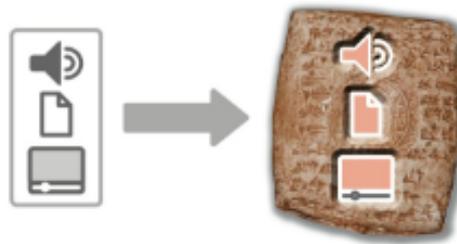
- CLARIN è una infrastruttura digitale, immateriale che mira a fornire:
 - sistema di **descrizione** e **catalogazione** dell'esistente
 - **deposito** per l'archiviazione a lungo termine
 - **accesso** facile e sostenibile da parte degli studiosi dei paesi membri
- CLARIN fornisce inoltre
 - strumenti avanzati per **reperire** le risorse adeguate per la propria ricerca ovunque esse si trovino e **combinarle** per **analizzare**, esplorare, annotare dati digitali

Servizi CLARIN



CLARIN portal

Get an example-based impression of what's currently available



Depositing services

Store language resources in a sustainable repository at a CLARIN centre

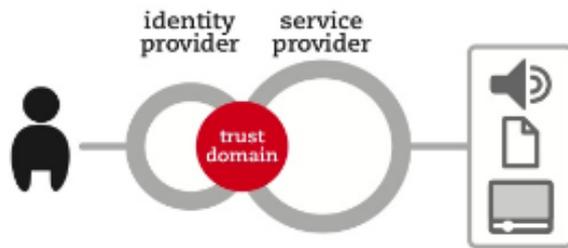


Virtual Language Observatory

Discover language resources using a faceted browser or a map

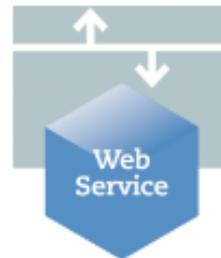
clarin.eu
clarin.eu/content/services

Servizi CLARIN



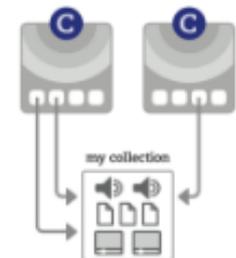
Easy access to protected resources

Get easy access to protected resources, with your institutional username and password.



Web services and applications

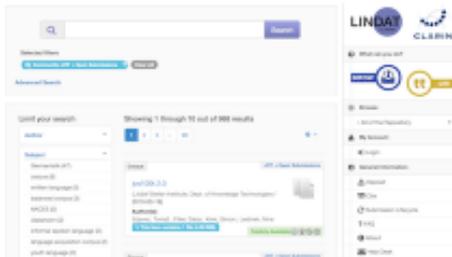
Explore and analyze language data with a wide variety of tools



Virtual Collections

Create your own digital bookmarks, ideal for citing data sets.

Servizi CLARIN



Language Resource Inventory

Submit and access information about language resources relevant to your research.



Content Search (prototype version)

Search different corpora with a single search engine



Consulting Services

Searching for a specific data set or application? Wondering how CLARIN can assist your research? Feel free to contact us!

Come funziona CLARIN-ERIC



- CLARIN è organizzato sotto forma di consorzio di nazioni (e organismi sovranazionali)
- Ciascuna nazione:
 - paga una tassa annua di iscrizione
 - offre servizi in-kind per il mantenimento del consorzio
 - mette a disposizione le proprie risorse e servizi tramite uno o più **CENTRI CLARIN**

Come funziona un CLARIN nazionale



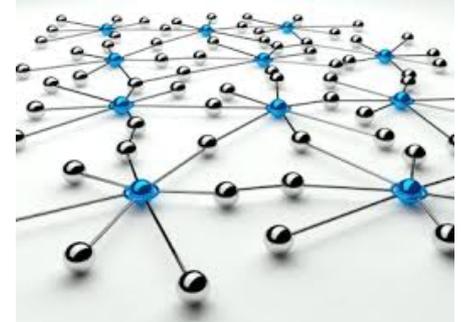
- CLARIN raggruppa 18 paesi (più alcuni membri osservatori)
- I CLARIN nazionali sono anche essi organizzati in consorzi:
- raggruppano varie istituzioni (centri di ricerca, università, biblioteche, istituzioni di supporto alla ricerca...)
- sotto forma di centri, forniscono dati, strumenti e servizi
- supportati finanziariamente dalla propria nazione grazie a progetti strategici nazionali

Members	National Consortium (NC)	Leading NC partner	Number NC partners	National coordinator
Austria	CLARIN Austria	CLARIN Centre Vienna	4	Karlheinz Mörth
Bulgaria	CLARIN Bulgaria	Bulgarian Academy of Sciences	10	Kiril Simov
Czech Republic	LINDAT Clarin	Charles University Prague	4	Eva Hajičová
Denmark	CLARIN Denmark	University of Copenhagen	4	Bente Maegaard
Dutch Language Union	Dutch Language Union	Dutch Language Union	4	Jan Theo Bakker
Estonia	CLARIN Estonia	Center of Estonian Language Resources	3	Kadri Vider
Finland	FIN-CLARIN			Krister Lindén
Germany	CLARIN-D	University of Tuebingen	12	Erhard Hinrichs
Greece	CLARIN Greece			Stelios Piperidis
Italy	CLARIN Italy			Monica Monachini
Lithuania	CLARIN-Lithuania	Vytautas Magnus University	3	Jurgita Vaičėnionienė
The Netherlands	CLARIN-NL	Utrecht University	24	Jan Odijk
Norway	CLARINO	University of Bergen	8	Koenraad De Smedt
Poland	CLARIN Poland	Wroclaw University of Technology	6	Maciej Piasecki
Portugal	CLARIN-Portugal	University of Lisbon	22	António Branco
Slovenia	CLARIN.SI	Jožef Stefan Institute	12	Tomaž Erjavec
Sweden	SWE-CLARIN	Språkbanken	9	Lars Borin
Observer	National Consortium (NC)	Leading partner NC	Number partners NC	National coordinator
United Kingdom	CLARIN-UK	Oxford University	12	Martin Wynne

CLARIN-ERIC: i pilastri



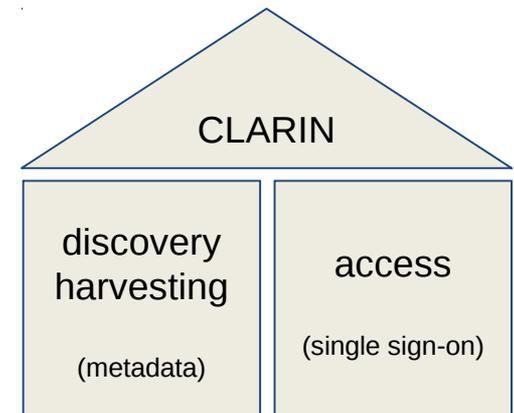
- CLARIN si configura quindi come un network di centri di vario tipo che sono i pilastri dell'infrastruttura



- I pilastri fondamentali a cui ogni centro deve attenersi:

- **documentazione**: le risorse devono essere documentate con un formato standard di **metadati** (CMDI),
- **esposizione** di metadati per **harvesting in un meta-catalogo centralizzato** (VLO)

- **accesso effettivo alle risorse**:
 - **completamente libero**
 - **sistema unificato di autenticazione single sign-on** (questo è particolarmente importante per tutte quelle risorse digitali che sono disponibili soltanto per scopi accademici)



CLARIN-ERIC: governance



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

General Assembly: responsabili designati dal ministero di ogni nazione. Massimo organismo decisionale.



Clarin office + Board of directors: gestione esecutiva



National coordinators' forum: capi dei consorzi nazionali + **Standing committee for technical centers:** integrazione e interoperabilità a livello trans-nazionale



CLARIN-nazionali



CLARIN ERIC non impone un modello di governance ai vari consorzi nazionali.

Richiede di:

- Creare una comunità, coinvolgendo gli attori principali
- Impiantare un centro (o più), per mettere a disposizione varie tipologie di dati, strumenti e servizi
- Partecipare alle attività centralizzate, comitati, etc



[Calls ▾](#)

[Media ▾](#)

[Events](#)

[Showcases](#)

[Projects ▾](#)

[Blog](#)

[About ▾](#)

CLARIN Conference 2016: Call for Papers

Call for papers **** CLARIN 2016 ****

CLARIN is happy to announce the 5th CLARIN Annual Conference and calls for the submission of papers. The series of CLARIN Annual

[Read more >](#)

 [@Twitter](#)

[UWuttke](#)

“On my way to Vienna for workshop with [@CLARINERIC](#) at [@oeaw](#)! Maybe also some sightseeing;-)”

[4 days 3 hours ago.](#)



CLARIN-NL in a nutshell



The CLARIN-NL project is a large national project in the Netherlands (2009-2015) which aims to make a significant contribution to the European CLARIN infrastructure.

The CLARIN infrastructure is a **research infrastructure** intended for humanities researchers that work with language data and tools. The researcher will be able to find data and tools relevant for his/her research via the CLARIN infrastructure and get access to them. We also aim to make it possible to apply tools to data in such a way that no technical background is needed or ad-hoc adaptations to the tools or the data are necessary.

The researcher's research project may result in new **data** and **tools**, which he/she can make available to other researchers via the CLARIN infrastructure. The CLARIN research infrastructure will thus make humanities research related to language easier, faster, better, and in some cases even possible for the first time.

CENTRE	FOCUS ON RESOURCES	RESOURCE EXAMPLES
MEERTENS INSTITUTE	relevant for the study of function, meaning and coherence of cultural expressions and resources relevant for the <u>structural, dialectological and sociolinguistic study</u> of language variation within the Dutch language.	typological databases for Dutch dialects, descriptions of Dutch songs, databases of Dutch names, etc. For more examples, see here
MPI	related to the study of psychological, social and biological foundations of language.	documentation of endangered languages, resources for sign languages, phonetic resources for the study of phone perception, speech error databases, tools for creating and annotating resources, etc. For more examples, see here
HUYGENS ING	related to the study of history and literature of the Netherlands.	historical and literary manuscripts and their annotations, tools to annotate such manuscripts and create scholarly editions, etc. For more examples see here
INL	that are relevant to the lexicological study of the Dutch language and on resources relevant for research in and development of language and speech technology.	lexicons, lexical databases, text corpora, speech corpora, language and speech technology tools, etc. For more examples, see here
DANS	sustained access to digital research data. The DANS data archive contains thousands of datasets in the fields of humanities including oral history, archaeology, geospatial sciences and behavioural and social sciences. The data archive has acquired the <u>Data Seal of Approval</u> . Regarding the conditions under which access to the data is granted, the DANS motto is "Open Access when possible, Restricted Access when necessary".	text resources, databases, spreadsheets, audio and audio-visual resources. For more examples see here

CENTRE	FOCUS ON RESOURCES	RESOURCE EXAMPLES
MEERTENS INSTITUTE	relevant for the study of function, meaning and coherence of cultural expressions and resources relevant for the <u>structural, dialectological and sociolinguistic study</u> of language variation within the Dutch language.	typological databases for Dutch dialects, descriptions of Dutch songs, databases of Dutch names, etc. For more examples, see here
MPI	related to the study of psychological, social and biological foundations of language.	documentation of endangered languages, resources for sign languages, phonetic resources for the study of phone perception, speech error databases, tools for creating and annotating resources, etc. For more examples, see here
HUYGENS ING	related to the study of history and literature of the Netherlands.	historical and literary manuscripts and their annotations, tools to annotate such manuscripts and create scholarly editions, etc. For more examples see here
INL	that are relevant to the lexicological study of the Dutch language and on resources relevant for research in and development of language and speech technology.	lexicons, lexical databases, text corpora, speech corpora, language and speech technology tools, etc. For more examples, see here
DANS	sustained access to digital research data. The DANS data archive contains thousands of datasets in the fields of humanities including oral history, archaeology, geospatial sciences and behavioural and social sciences. The data archive has acquired the <u>Data Seal of Approval</u> . Regarding the conditions under which access to the data is granted, the DANS motto is "Open Access when possible, Restricted Access when necessary".	text resources, databases, spreadsheets, audio and audio-visual resources. For more examples see here

CENTRE	FOCUS ON RESOURCES	RESOURCE EXAMPLES
MEERTENS INSTITUTE	relevant for the study of function, meaning and coherence of cultural expressions and resources relevant for the <u>structural, dialectological and sociolinguistic study of language variation within the Dutch language.</u>	typological databases for Dutch dialects, descriptions of Dutch songs, databases of Dutch names, etc. For more examples, see here
MPI	related to the study of psychological, social and biological foundations of language.	documentation of endangered languages, resources for sign languages, phonetic resources for the study of phone perception, speech error databases, tools for creating and annotating resources, etc. For more examples, see here
HUYGENS ING	related to the study of history and literature of the Netherlands.	historical and literary manuscripts and their annotations, tools to annotate such manuscripts and create scholarly editions, etc. For more examples see here
INL	that are relevant to the lexicological study of the Dutch language and on resources relevant for research in and development of language and speech technology.	lexicons, lexical databases, text corpora, speech corpora, language and speech technology tools, etc. For more examples, see here
DANS	sustained access to digital research data. The DANS data archive contains thousands of datasets in the fields of humanities including oral history, archaeology, geospatial sciences and behavioural and social sciences. The data archive has acquired the <u>Data Seal of Approval</u> . Regarding the conditions under which access to the data is granted, the DANS motto is "Open Access when possible, Restricted Access when necessary".	text resources, databases, spreadsheets, audio and audio-visual resources. For more examples see here

CENTRE	FOCUS ON RESOURCES	RESOURCE EXAMPLES
MEERTENS INSTITUTE	relevant for the study of function, meaning and coherence of cultural expressions and resources relevant for the <u>structural, dialectological and sociolinguistic study</u> of language variation within the Dutch language.	typological databases for Dutch dialects, descriptions of Dutch songs, databases of Dutch names, etc. For more examples, see here
MPI	related to the study of psychological, social and biological foundations of language.	documentation of endangered languages, resources for sign languages, phonetic resources for the study of phone perception, speech error databases, tools for creating and annotating resources, etc. For more examples, see here
HUYGENS ING	related to the study of history and literature of the Netherlands.	historical and literary manuscripts and their annotations, tools to annotate such manuscripts and create scholarly editions, etc. For more examples see here
INL	that are relevant to the lexicological study of the Dutch language and on resources relevant for research in and development of language and speech technology.	lexicons, lexical databases, text corpora, speech corpora, language and speech technology tools, etc. For more examples, see here
DANS	sustained access to digital research data. The DANS data archive contains thousands of datasets in the fields of humanities including oral history, archaeology, geospatial sciences and behavioural and social sciences. The data archive has acquired the <u>Data Seal of Approval</u> . Regarding the conditions under which access to the data is granted, the DANS motto is "Open Access when possible, Restricted Access when necessary".	text resources, databases, spreadsheets, audio and audio-visual resources. For more examples see here

Concretamente....(1)



- Un ricercatore vuole fare una ricerca sulla sintassi delle varie lingue
- In particolare è interessato a studiare il giapponese....
- Cerca uno strumento per la ricerca di corpora annotati sintatticamente (tree-bank)

Virtual Language Observatory



Explore the world of language resources and technology from different perspectives



[VLO](#) > [Faceted search](#) > [Search: "querying annotated treebanks"](#) ✕

SEARCH



SEARCH RESULTS

10 results

Showing 1 to 10

[PML Tree Query](#)

[Expand](#)

System for **querying annotated treebanks** in PML format. The **querying** uses its own query language with graphical representation. It has two different implementations (SQL and Perl) and several clients (TrEd, browser-based, command line interface).

Resources: | 1 text document |



- Fa una ricerca nel catalogo centrale di CLARIN, il VLO
- E scopre uno strumento, il Tree Query

PML Tree Query

Tool for searching and browsing treebanks online

 [Browse Treebanks](#)

 [Login](#)

- viene ridiretto al centro Lindat CLARIN (Repubblica Ceca)

Filter treebanks ...



Show only publicly accessible treebanks

LANGUAGES:

Ancient Greek 3	Arabic 2	Basque 1	Bengali 0	Bulgarian 2	Catalan 1	Chinese 1	
Croatian 1	Czech 9	Danish 1	Dutch 3	English 7	Estonian 2	Finnish 2	French 1
Galician 1	German 1	Gothic 1	Greek 1	Hebrew 1	Hindi 1	Hungarian 1	Indonesian 1
Irish 1	Italian 1	Japanese 0	Kazakh 1	Latin 7	Latvian 1	Norwegian 1	Old Church Slavonic 1
Persian 2	Polish 2	Portuguese 3	Romanian 2	Russian 2	Slovak 0	Slovenian 4	Spanish 2
Swedish 3	Tamil 2	Telugu 0	Turkish 1				

TAGS:

CoNLL 1	HamleDT 13	PDT 2	Penn Treebank 4	Treex 1	Universal Dependencies 52
---------	------------	-------	-----------------	---------	---------------------------

Reset filter

- Lo strumento è liberamente accessibile, ...

Filter treebanks ...



Show only publicly accessible treebanks

LANGUAGES:

Ancient Greek 3	Arabic 2	Basque 1	Bengali 0	Bulgarian 2	Catalan 1	Chinese 1	
Croatian 1	Czech 9	Danish 1	Dutch 3	English 7	Estonian 2	Finnish 2	French 1
Galician 1	German 1	Gothic 1	Greek 1	Hebrew 1	Hindi 1	Hungarian 1	Indonesian 1
Irish 1	Italian 1	Japanese 0	Kazakh 1	Latin 7	Latvian 1	Norwegian 1	Old Church Slavonic 1
Persian 2	Polish 2	Portuguese 3	Romanian 2	Russian 2	Slovak 0	Slovenian 4	Spanish 2
Swedish 3	Tamil 2	Telugu 0	Turkish 1				

TAGS:

CoNLL 1	HamleDT 13	PDT 2	Penn Treebank 4	Treex 1	Universal Dependencies 52
---------	------------	-------	-----------------	---------	---------------------------

Reset filter

- ma alcuni dei corpora non sono disponibili liberamente

Login

To sign in, you can use any account with an **Identity Provider** that is a member of **EduGAIN** federation. If you don't have an academic account that works with us, let us know at lindat-help@ufal.mff.cuni.cz. We will make you a local account.

Academic login

OR

Local account

Username:

Password:

Remember me

Sign in with local account

- Al momento del login, si trova davanti ad una richiesta di autenticarsi tramite la federazione CLARIN

Sign in to **PML Tree Query**
Login via Your home institution (e.g. university)

Univerzita Karlova v Praze
Czech Republic

Bocconi University
Italy

University of Genoa
Italy

CSP - innovation in ICT
Italy

Università di Ferrara
Italy

UniServity: CAVLE

IMT Lucca - Identity Provider
Italy

Università di Napoli Parthenope
Italy

Urbino University
Italy

IZS Sicilia

or search for a provider, such as Example University

▶ **Please help, I cannot find my provider**

🎯 Locate me and show nearby providers

Show providers in show all countries

DiscoJuice © UNINETT

- Cerca il nome della propria istituzione nella lista

Filter treebanks ...



Show only publicly accessible treebanks

- LANGUAGES:**
- Ancient Greek 3
 - Arabic 2
 - Basque 1
 - Bengali 1
 - Bulgarian 2
 - Catalan 3
 - Chinese 4
 - Croatian 1
 - Czech 13
 - Danish 1
 - Dutch 3
 - English 12
 - Estonian 2
 - Finnish 2
 - French 1
 - Galician 1
 - German 4
 - Gothic 1
 - Greek 1
 - Hebrew 1
 - Hindi 3
 - Hungarian 1
 - Indonesian 1
 - Irish 1
 - Italian 1
 - Japanese 2
 - Kazakh 1
 - Latin 7
 - Latvian 1
 - Norwegian 1
 - Old Church Slavonic 1
 - Persian 2
 - Polish 2
 - Portuguese 3
 - Romanian 2
 - Russian 3
 - Slovak 1
 - Slovenian 4
 - Spanish 4
 - Swedish 3
 - Tamil 2
 - Telugu 2
 - Turkish 2

- TAGS:**
- CoNLL 8
 - HamleDT 24
 - PDT 3
 - Penn Treebank 6
 - Treex 1
 - Universal Dependencies 52

Reset filter

▼ Relations ▼
Node Types ▼
Attributes ▼
Operators ▼
Functions ▼

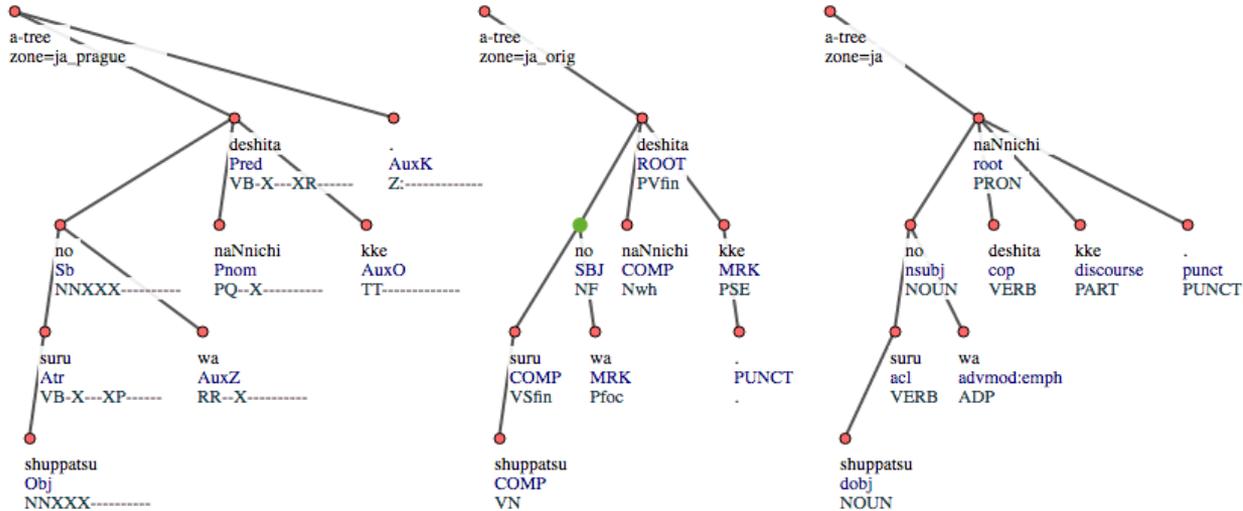
`p-node []`

[Execute query](#)

[w/o Filters](#)

← Previous
1 of 100
Next →

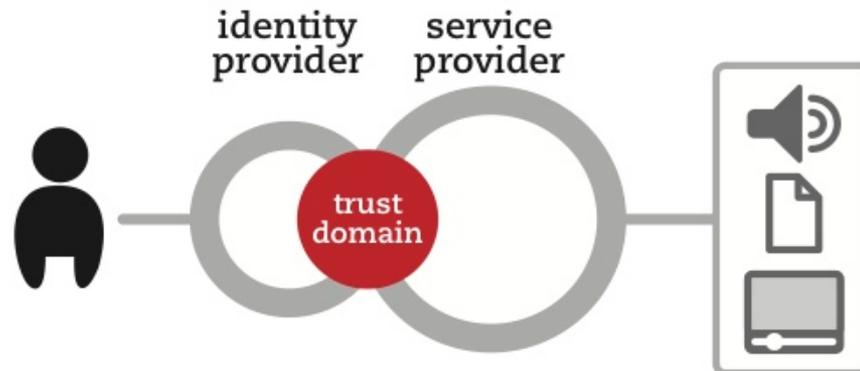
[ja_prague] shuppatsu suru no wa naNnichi deshita kke .
[ja_orig] shuppatsu suru no wa naNnichi deshita kke .
[ja] shuppatsu suru no wa naNnichi deshita kke .



Accesso



- CLARIN si federa con le federazioni di identità nazionali (ex. IDEM-GARR) e internazionali (ex. EDUGAIN), dando priorità all'accesso dei paesi Full Member
- CLARIN ERIC ha costituito una propria federazione (SPF) per connettere i propri service providers (=centri) con le federazioni nazionali di identità (=utenti)
 - Questo evita ai singoli centri CLARIN di dover siglare accordi con ogni federazione
 - CLARIN-IT è perfettamente integrato nel sistema di accesso federato di



Participating Identity Federations

We aim at including all [European federations](#), however [CLARIN member countries](#) get the highest priority.

Country	Federation	Status	Attribute release	Opt-in
Austria	ACOnet	Fully connected.		OK (not required)
Belgium	Belnet	Fully connected.	OK (for Antwerp, Gent, Leuven)	OK (not required)
Czech Republic	eduID.cz	Fully connected.		OK (not required)
Denmark, Iceland	WAYF	Fully connected (via eduGAIN and Kalmar Union).	OK	OK (not required)
Estonia	TAAT	Fully connected (via eduGAIN).	OK	OK (not required, attribute release mandatory)
Finland	Haka	Fully connected. Also includes Kalmar Union	OK	OK (not required)
Germany	DFN-AAI	Fully connected.	Problems for some IdPs (e.g. Tübingen)	OK (not required)
Greece	GRNET Federation	Connection planned since 2015.		
Italy	IDEM GARR	Fully connected (via eduGAIN).		OK (not required)
Lithuania	LITNET fedi	Connection planned since 2015.		



Servizi protetti



- Una lista di servizi accessibili agli utenti italiani grazie all'interfederazione IDEM / CLARIN

<https://www.clarin.eu/content/easy-access-protected-resources>

Easy access to protected resources

Thanks to [federated login](#), the following applications and data sets are available to anyone with an academic account from many European countries – for a complete list, see [Participating Identity Federations](#). We are working hard to extend this to the rest of Europe. In the meanwhile, you can request [a CLARIN account](#) if you want to access these services from another country or from an institution that does not participate in these identity federations.

Resource or Tool	Description	Provided by
OpenSoNaR	over 500 million word Dutch reference corpus	INL (Instituut voor Nederlandse Lexicologie)
Corpus Hedendaags Nederlands	written corpus of contemporary Dutch	INL (Instituut voor Nederlandse Lexicologie)
VU-DNC	diachronic Dutch newspaper corpus	INL (Instituut voor Nederlandse Lexicologie)
Tündera	treebank search application	Eberhard Karls Universität Tübingen

Verso CLARIN-IT



- Il CNR, nella persona del prof Riccardo Pozzo, Dipartimento di Scienze Umane e Sociali
 - rappresentante nazionale e interfaccia con il Ministro
- Il National Coordinator ha il compito:
 - creare il consorzio nazionale CLARIN-IT
 - far da tramite tra la comunità italiana e il CLARIN-ERIC
- ILC-CNR è l'istituto esecutore con il compito di
 - creare un centro italiano per la documentazione, accesso e consultazione delle risorse linguistiche



- About
- Governance
- Centri
- Partecipa
- Eventi
- Notizie

L'Infrastruttura Comune Italiana delle Risorse e delle Tecnologie Linguistiche



Italiano

IN EVIDENZA

Conferenza Annuale CLARIN 2016
Aix-en-Provence
26/10/2016 - 28/10/2016

NOTIZIE RECENTI

Bando per Proposte di CLARIN
Workshop
Scadenza Invio: 01/09/2016

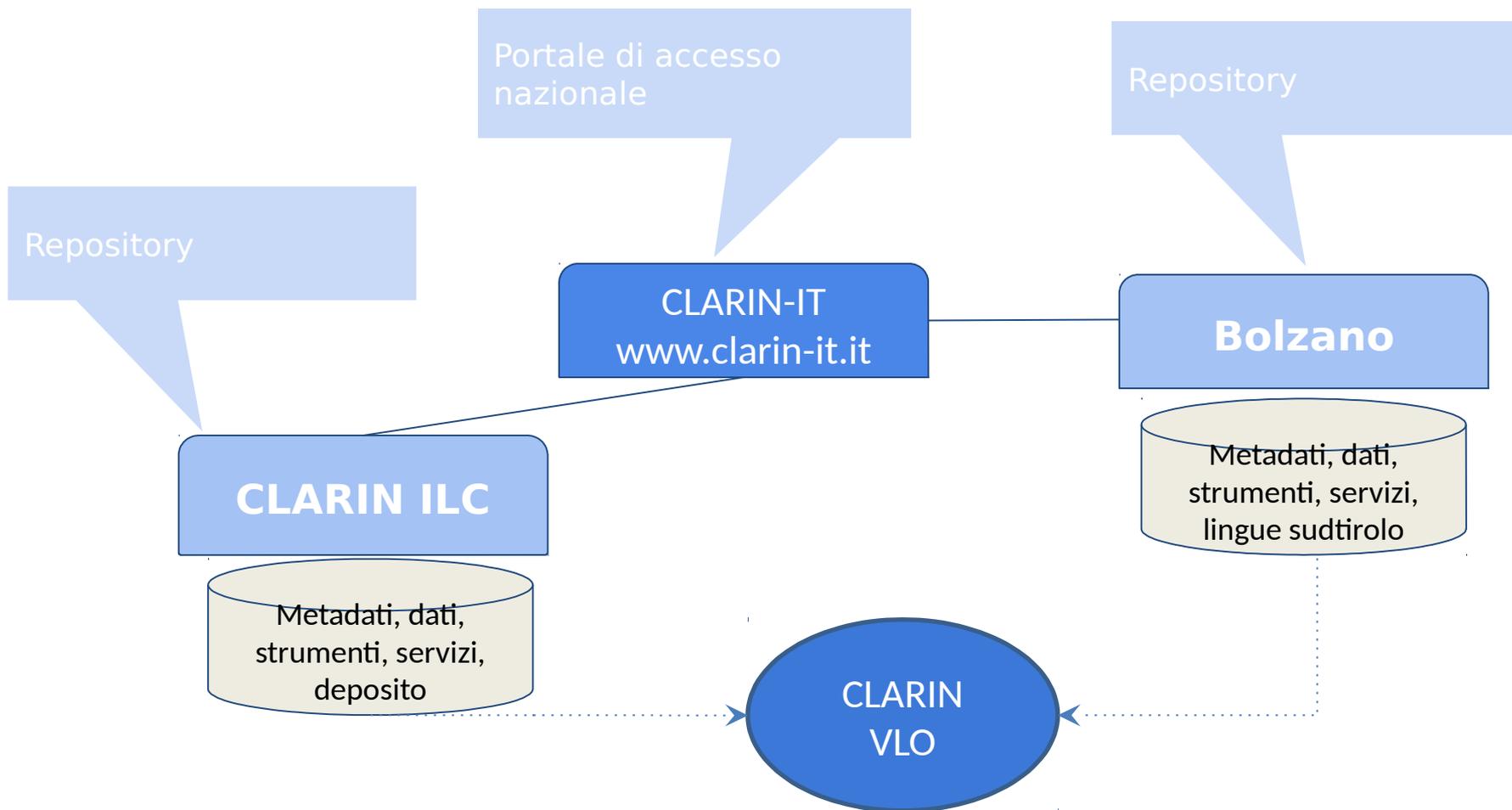
[Visualizza tutte le notizie](#)

CLARIN-IT: Consorzio



- Collaborazione con **IDEM-GARR** per l'accesso federato
- Le istituzioni che hanno mostrato espressione di interesse:
 - **Università di Siena** (collaborazione aperta ad AIUCD15)
 - **EURAC Bolzano**
 - **Università di Venezia**
 - **Università di Pisa; Scuola Normale Superiore**
 - **Università di Parma**
 - **FBK Trento**
 - **Università Cattolica Milano**
 - **Università di Roma**

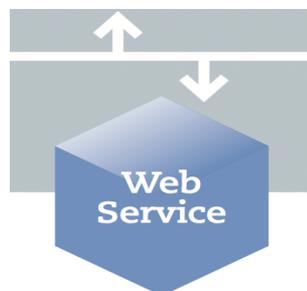
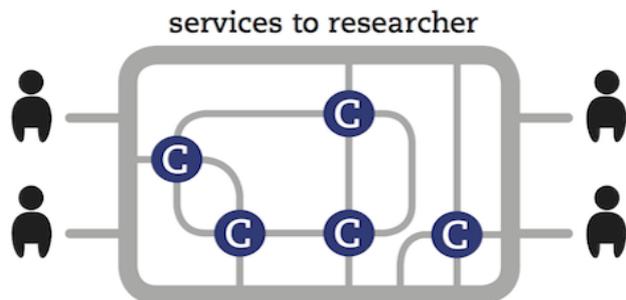
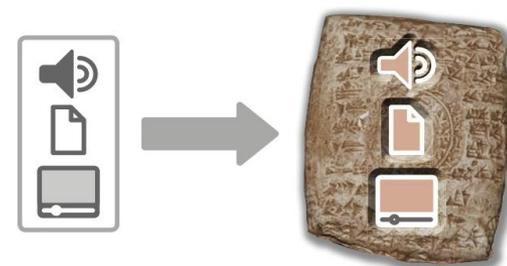
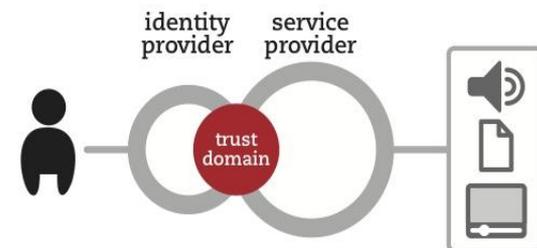
CLARIN-IT



Centro ILC-CLARIN



- accesso a risorse e servizi dell'Istituto di Linguistica Computazionale
- servizi di deposito di risorse di terze parti, come collettore di dati e strumenti a livello nazionale.
- forte accento sull'italiano e a risorse per le lingue classiche
- preservazione a lungo termine
- single sign-on



Deposito & documentazione



- Gli utenti possono già ora documentare ed eventualmente depositare le proprie risorse sul repository ILC-CLARIN
- Questo garantisce :
 - Reperibilità
 - Citabilità
 - Usabilità (scaricamento)

CENTRO ILC: dspace-clarin-it.ilc.cnr.it



Repository

About



Login

Deposit Free and Safe

License of your Choice (Open licenses encouraged)

Easy to Find

Easy to Cite



[Advanced Search](#)

Author	Subject	Language (ISO)
Bertagna, Francesca (1)	800 (1)	Italian (2)
Boschetti, Federico (1)	900 (1)	
Del Gratta, Riccardo (1)	Letteratura Italiana (1)	
Enea, Alessandro (1)	lexicon (1)	
Manzoni, Alessandro (1)	synsets (1)	
... View More	... View More	

What's New

Corpus

OPEN

[I promessi sposi](#)

Author(s):

Manzoni, Alessandro

Description:

A TEI P5 encoding of the I Promessi Sposi

[This item contains 1 file \(1.31 MB\).](#)

Publicly Available

LexicalConceptualResource

OPEN

[ItalWordNet](#)

Author(s):

Roventini, Adriana ; Marinelli, Rita ; Enea, Alessandro ; Bertagna, Francesca

What can you do?



Browse

> All of the Repository

My Account

Login

General Information

Deposit

Cite

Submission Lifecycle

CENTRO ILC: dspace-clarin-it.ilc.cnr.it



Repository

About



Login

Deposit Free and Safe
License of your Choice (Open licenses encouraged)
Easy to Find
Easy to Cite



[Advanced Search](#)

Author	Subject	Language (ISO)
Bertagna, Francesca (1)	800 (1)	Italian (2)
Boschetti, Federico (1)	900 (1)	
Del Gratta, Riccardo (1)	Letteratura Italiana (1)	
Enea, Alessandro (1)	lexicon (1)	
Manzoni, Alessandro (1)	synsets (1)	
... View More	... View More	

What's New

Corpus

OPEN

I promessi sposi

Author(s):

Manzoni, Alessandro

Description:

A TEI P5 encoding of the I Promessi Sposi

[This item contains 1 file \(1.31 MB\).](#)

Publicly Available

LexicalConceptualResource

OPEN

ItalWordNet

Author(s):

Roventini, Adriana ; Marinelli, Rita ; Enea, Alessandro ; Bertagna, Francesca

What can you do?



Browse

> All of the Repository

My Account

Login

General Information

Deposit

Cite

Submission Lifecycle

CENTRO ILC: dspace-clarin-it.ilc.cnr.it



Repository

About



Login

Deposit Free and Safe

License of your Choice (Open licenses encouraged)

Easy to Find

Easy to Cite



[Advanced Search](#)

Author	Subject	Language (ISO)
Bertagna, Francesca (1)	800 (1)	Italian (2)
Boschetti, Federico (1)	900 (1)	
Del Gratta, Riccardo (1)	Letteratura Italiana (1)	
Enea, Alessandro (1)	lexicon (1)	
Manzoni, Alessandro (1)	synsets (1)	
... View More	... View More	

What's New

Corpus

OPEN

I promessi sposi

Author(s):

Manzoni, Alessandro

Description:

A TEI P5 encoding of the I Promessi Sposi

[This item contains 1 file \(1.31 MB\).](#)

Publicly Available

LexicalConceptualResource

OPEN

ItalWordNet

Author(s):

Roventini, Adriana ; Marinelli, Rita ; Enea, Alessandro ; Bertagna, Francesca

What can you do?

DEPOSIT **CITE**

Browse

- > All of the Repository

My Account

- Login

General Information

- Deposit
- Cite
- Submission Lifecycle

Step 3: Select type of your submission

You have initiated a new workflow item. In the next few steps you will provide the details about the item and upload content files. First select the type of the resource you are about to submit.

The screenshot shows a progress bar for 'Item submission' with 8 steps: 1. Basic Info (active), 2. Who's involved, 3. Describe, 4. Upload, 5. License, 6. Note, 7. Review, 8. Complete. Below the progress bar is a 'Submission Info' section with four buttons: 'Corpus', 'Lexical conceptual', 'Language description', and 'Technology / Tool / Service'. An information icon (i) is followed by a text box containing the following text: 'Type of the resource: "Corpus" refers to text, speech and multimodal corpora. "Lexical Conceptual Resource" includes lexica, ontologies, dictionaries, word lists etc. "Language Description" covers language models and grammars. "Technology / Tool / Service" is used for tools, systems, system components etc.' Below this is a 'Title' label and an empty text input field. A blue information icon (i) is followed by the text: 'Enter the main title of the item in English.'

Fig4. Submission info

Click on one of the type buttons e.g. Corpus. Proceed with filling the basic information such as the title. Click Next to continue the following step.

Step 4: Describe your item

In the following two steps you will provide more details for your item. First describe the people, organization and projects involved with the item.

The screenshot shows a progress bar for 'Item submission' with 8 steps: 1. Basic Info (checked), 2. Who's involved (active), 3. Describe (checked), 4. Upload (checked), 5. License (checked), 6. Note (checked), 7. Review (checked), 8. Complete. Below the progress bar is a 'People and organizations involved' section with a sub-section 'Authors'. It contains two text input fields: 'Last name, e.g. Smith' and 'First name(s) + "Jr", e.g. Donald Jr'. Below the input fields is a blue 'Add' button.

CENTRO ILC: dspace-clarin-it.ilc.cnr.it

I promessi sposi

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

BIBTEX CMDI

Manzoni, Alessandro, 2014, *I promessi sposi*, Digital Repository for the CLARIN Research Infrastructure provided by ILC-CNR, <http://hdl.handle.net/20.500.11752/OPEN.47>.

Share:   

OPEN

 Authors	Manzoni, Alessandro
 Date issued	2014
 Type	corpus
 Size	1 file
 Language(s)	Italian
 Description	A TEI P5 encoding of the I Promessi Sposi
 Publisher	Institute for Computational Linguistics
 Subject(s)	Letteratura Italiana 800 900
 Collection(s)	Collection_1

[Show full item record](#)

Files in this item

This item is **Publicly Available** and licensed under:
Creative Commons - Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Name	Alessandro_Manzoni_IPS.XML
Size	1.31 MB
Format	XML
Description	Alessandro_Manzoni_TEI_P5.XML

[Download file](#)



Alessandro Enea | [Logout](#)

What can you do?



Browse

> All of the Repository

My Account

[Logout](#)

[Profile](#)

[Submissions](#)

Context

> Edit this item

> Export Item

> Export Metadata

Administrative

[Control Panel](#)

[Access Control](#)

> Content Administration

[Registries](#)

[Collections & Communities](#)

[Statistics](#)

[Curation Tasks](#)

[Licenses](#)

[Handles](#)

> External Handles

[Statistics](#)

CENTRO ILC: dspace-clarin-it.ilc.cnr.it

I promessi sposi

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

Manzoni, Alessandro, 2014, *I promessi sposi*, Digital Repository for the CLARIN Research Infrastructure provided by ILC-CNR, <http://hdl.handle.net/20.500.11752/OPEN.47>.

Share:    

✍ Authors	Manzoni, Alessandro
📅 Date issued	2014
📁 Type	corpus
📏 Size	1 file
🗣 Language(s)	Italian
📄 Description	A TEI P5 encoding of the I Promessi Sposi
🏢 Publisher	Institute for Computational Linguistics
🏷 Subject(s)	Letteratura Italiana 800 900
🗑 Collection(s)	Collection_1

[Show full item record](#)

Files in this item

This item is **Publicly Available** and licensed under:
Creative Commons - Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Name	Alessandro_Manzoni_IPS.XML
Size	1.31 MB
Format	XML
Description	Alessandro_Manzoni_TEI_P5.XML

[Download file](#)



Alessandro Enea | [Logout](#)

What can you do?



Browse

> All of the Repository

My Account

[Logout](#)

[Profile](#)

[Submissions](#)

Context

> Edit this item

> Export Item

> Export Metadata

Administrative

[Control Panel](#)

[Access Control](#)

> Content Administration

[Registries](#)

[Collections & Communities](#)

[Statistics](#)

[Curation Tasks](#)

[Licenses](#)

[Handles](#)

> External Handles

[Statistics](#)

CENTRO ILC: dspace-clarin-it.ilc.cnr.it

I promessi sposi

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

BIBTEX CMDI

Manzoni, Alessandro, 2014, *I promessi sposi*, Digital Repository for the CLARIN Research Infrastructure provided by ILC-CNR, <http://hdl.handle.net/20.500.11752/OPEN.47>.

Share:   

OPEN

 Authors	Manzoni, Alessandro
 Date issued	2014
 Type	corpus
 Size	1 file
 Language(s)	Italian
 Description	A TEI P5 encoding of the I Promessi Sposi
 Publisher	Institute for Computational Linguistics
 Subject(s)	Letteratura Italiana 800 900
 Collection(s)	Collection_1

[Show full item record](#)

Files in this item

This item is **Publicly Available** and licensed under:
Creative Commons - Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

Name	Alessandro_Manzoni_IPS.XML
Size	1.31 MB
Format	XML
Description	Alessandro_Manzoni_TEI_P5.XML

 [Download file](#)



Alessandro Enea | [Logout](#)

What can you do?



Browse

> All of the Repository

My Account

[Logout](#)

[Profile](#)

[Submissions](#)

Context

> Edit this item

> Export Item

> Export Metadata

Administrative

[Control Panel](#)

[Access Control](#)

> Content Administration

[Registries](#)

[Collections & Communities](#)

[Statistics](#)

[Curation Tasks](#)

[Licenses](#)

[Handles](#)

> External Handles

[Statistics](#)

Harvesting



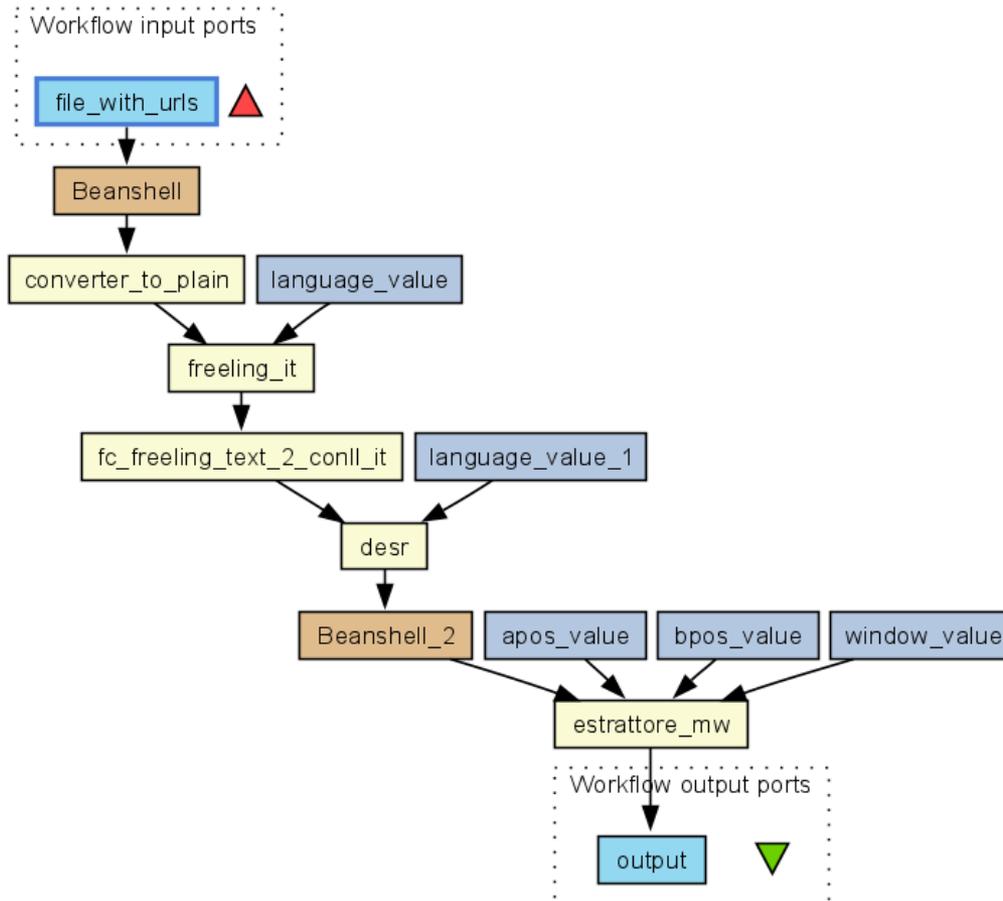
- I metadati di ILC-CLARIN, come di tutti gli altri CENTRI europei, sono raccolti dal catalogo centralizzato dell'ERIC
- Le risorse italiane divengono quindi cercabili nel Virtual Language Observatory di CLARIN

Concretamente....(2)



- Un italianista vuole studiare le caratteristiche del lessico letterario
- In CLARIN-IT può trovare:
 - testi di opere letterarie
 - strumenti per l'analisi morfo sintattica dei testi,
 - strumenti per l'estrazione di informazioni lessicali;
 - strumenti per l'estrazione di relazioni;
 - ...

Ad esempio:



freeling_it (WSDL)

Freeling executable for Italian language only. Up to tagged

Run service

Inputs

input

as URL

direct data or local file

Sfoglia... **piccoloMondoAntico.txt**

language it.cfg

mandatory

optional

multiword yes no

ner none

output_format --Use Default--

- token
- splitted
- tagged

Reset fields

Report

Tagger = analisi
morfosintattica = analisi
grammaticale

freeling_it (WSDL)

Freeling executable for Italian language only. Up to tagged



Run service

RUNNING

Status updated every 7 seconds

Result **Size** **Type**

 report 254 text

Update now

Terminate

Inputs

Report

input

as URL

direct data or local file

Sfoggia... piccoloMondoAntico.txt

language

it.cfg

mandatory

optional

multiword

yes no

ner

none

output_format

tagged

Reset fields

freeling_it (WSDL)

Freeling executable for Italian language only. Up to tagged



Run service

COMPLETED

Result	Size	Type
std_output	35	text
report	394	text
detailed_status	1	unknown

Remove

Inputs

input

as URL

direct data or local file

Sfoggia... piccoloMondoAntico.txt

language

it.cfg

mandatory

optional

multiword

yes no

ner

none

output_format

tagged

Reset fields

Report

Summary:

Completed: Successfully
Termination status: 0
Started: 2016-Aug-30 17:20:34 (CEST)
Ended: 2016-Aug-30 17:20:54 (CEST)
Duration: 0:00:20.052

Report:

Name: panacea.freeling_it
Job ID: [panacea.freeling_it]12d53536.15474e6e299._7fe5
Program and parameters:
/application/panacea/script/myfreeling.sh
-i
i_input
-f
it.cfg
--outf
tagged
--ner
none
--- end of parameters
Exit: 0

desr (WSDL)

DeSR is a shift-reduce dependency parser for Italian Language



Run service

COMPLETED

Result	Size	Type
output	35	text
report	467	text
detailed_status	1	unknown

[Remove](#)

Parser = analisi sintattica
funzionale

Inputs

as URL

direct data
or local file

[Sfogliala...](#) Nessun file selezionato.

language

[Reset fields](#)

Report

Summary:
Completed: Maybe
Termination status: 0
Started: 2016-Aug-30 17:22:56 (CEST)
Ended: 2016-Aug-30 20:20:50 (CEST)
Duration: 2:57:54.017

Report:
Some error messages were reported.

Name: panacea.desr
Job ID: [panacea.desr]12d53536.15474e6e299._7fe3
Program and parameters:
/application/panacea/script/desr.sh
-i
i_input
-l
it.cfg
--- end of parameters
Exit: 0

Standard error stream:
LAS: 0.16 % (158/97953)
UAS: 4.59 % (4495/97953)
Sentences: 3841

extractor_mw (WSDL)

Multiword Extractor language independent.

Run service

Inputs

as URL

direct data or local file

Sfoglia... Nessun file selezionato.

Report

```
Summary:
Completed: Successfully
Termination status: 0
Started: 2016-Aug-31 08:34:32 (CEST)
Ended: 2016-Aug-31 08:34:35 (CEST)
Duration: 0:00:02.317

Report:
Name: panacea.extractor_mw
Job ID: [panacea.extractor_mw]12d53536.15474e6e299._7fe0
Program and parameters:
/application/panacea/scripts/LexicalExtractor/ExtractorMW/estrattoreMW.sh
-c
-i_input
-a
S
-b
S
-w
5
-f
overmean
-p
averagef
-o
frelative
-out
tsv
-d
literature
--- end of parameters
Exit: 0
```

apos

bpos

domain

filtering_type

max_entry_num

order_by

output_type

prefiltering_type

as URL

direct data or local file

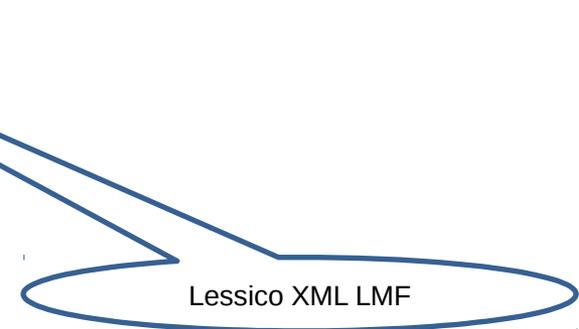
Sfoglia... Nessun file selezionato.

window

Reset fields

Estrazione candidate
Polirematiche, collocazioni,
combinazioni lessicali
tipiche


```
SampleOutput.TSV.txt | SampleOutput.LMF.xml
1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE LexicalResource SYSTEM "http://www.tagmatica.fr/lmf/DTD_LMF_REV_16.dtd">
3 <LexicalResource dtdVersion="16">
4   <GlobalInformation>
13   <Lexicon>
14     <feat att="type" val="Panacea_MWE_Lexicon"/>
15     <feat att="language" val="Italian"/>
16     <LexicalEntry id="le_0f3c15d0cdaa8cc69b14db8fe32888c5">
17       <feat att="entryType" val="Multiword"/>
18       <feat att="MWEPattern" val="s+ea+s"/>
19       <feat att="absoluteFrequency" val="3"/>
20       <feat att="logLikelihood" val="6.050826946349334e-4"/>
21       <feat att="writtenForm" val="scalinata del pomodoro"/>
22       <feat att="lemmaPair" val="scalinata-pomodoro"/>
23       <feat att="domain" val="fiction"/>
24     <Lemma>
25     </Lemma>
26     <ListOfComponents>
27       <Component entry="le_9a0cf9fb3aa296e903ef11fde65bd8c4">
28         <feat att="rank" val="0"/>
29         <feat att="pos" val="s"/>
30         <feat att="lemma" val="scalinata"/>
31         <feat att="writtenForm" val="scalinata"/>
32         <feat att="function" val="head"/>
33       </Component>
34       <Component entry="le_e19232a3a4f95cba523f09df6e1c2b73">
35         <feat att="rank" val="1"/>
36         <feat att="pos" val="ea"/>
37         <feat att="lemma" val="del"/>
38         <feat att="writtenForm" val="del"/>
39       </Component>
40       <Component entry="le_f6d677cb2e6cd36f1cf7c8249e8167ee">
41         <feat att="rank" val="2"/>
42         <feat att="pos" val="s"/>
43         <feat att="lemma" val="pomodoro"/>
44         <feat att="writtenForm" val="pomodoro"/>
45       </Component>
46     </ListOfComponents>
47   </LexicalEntry>
48   <LexicalEntry id="le_bac9ab8c65c2284225104f85b8a000fc">
49     <feat att="entryType" val="Multiword"/>
50     <feat att="MWEPattern" val="s+sp"/>
51     <feat att="absoluteFrequency" val="47"/>
52     <feat att="logLikelihood" val="0.009479628882613958"/>
53     <feat att="writtenForm" val="zio piero"/>
54     <feat att="lemmaPair" val="zio-piero"/>
55     <feat att="domain" val="fiction"/>
56   <Lemma>
57   </Lemma>
58   <ListOfComponents>
59     <Component entry="le_de7a6562a25b91e311978c3e7b878151">
60       <feat att="rank" val="0"/>
61       <feat att="pos" val="s"/>
62       <feat att="lemma" val="zio"/>
63       <feat att="writtenForm" val="zio"/>
64       <feat att="function" val="head"/>
65     </Component>
```



Networking



- Oltre alle attività del centro, CLARIN svolge attività di
 - coinvolgimento degli utenti (user involvement)
 - facilitazione delle collaborazioni e degli scambi di competenze
 - disseminazione della ricerca
 - educazione all'uso delle risorse e dei servizi
- Organizzazione di eventi e seminari, promozione di scambi

Networking



- Conferenza CLARIN annuale
- Seminari CLARIN PLUS
- Finanziamenti alla mobilità
- Partecipazione in progetti di ricerca
- Interazione con altre infrastrutture e network (DARIAH...)

CLARIN-IT: success story



- Università di Siena e progetto Gra.fo (archivi orali)
- Adesione di Silvia Calamai a CLARIN-IT
 - Partecipazione al seminario CLARIN-plus su Oral History
 - Coinvolgimento della comunità degli oralisti in CLARIN
 - Futura visita presso centro CLARIN di Monaco di Baviera

CLARIN-IT Progetti futuri



- Coinvolgere i produttori di risorse per venire incontro alle esigenze di long term preservation
- Coinvolgere gli utenti per far conoscere le nuove opportunità
- Organizzare eventi a livello nazionale

Questionario



- Per far conoscere CLARIN e raccogliere le esigenze della comunità delle DH.
- Vai al [questionario](#)

VLO, la ricerca



Virtual Language Observatory

Search

Help

CLARIN

↩!🔍?

Showing all 907344 records

Results per page: 10

⚙️ Hide facets and search options

Facets

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

Language



Collection



Resource type



Modality



Format



Keyword



VLO, le faccette



Language

Collect

Resource

Type

Text (1)

TextAn

Televi

Sound

Physic

Data P

Biosco

Stillm

Dialog

Sessio

more..

Modali

Forma

Keywo

Search

Resource type

All values in this facet

Search: corp

Sort by Name Only show values that occur at least 2 times

- AddressesAnthologiesLinguistic corpora (1)
- AnnotatedTextCorpus (7)
- AnthologiesLinguistic corpora (2)
- Bible storiesLinguistic corpora (1)
- Corpus (814)
- Linguistic corpora (60)
- Linguistic corporaAnthologies (3)
- Linguistic corporaLetters (1)
- Linguistic corporaLettersAnthologies (1)
- Linguistic corporaNewspapersAnthologies (1)
- Linguistic corporaPeriodicals -- Italy -- 20th Century (1)
- Linguistic corporaSpeech--Research (1)
- MultimodalCorpus (2)
- NewspapersAnthologiesLinguistic corpora (1)
- NewspapersLinguistic corpora (1)
- NovelsAnthologiesLinguistic corpora (1)
- OralCorpus (7)
- PlaysAnthologiesLinguistic corpora (1)
- RomancesLinguistic corpora (1)
- SongsAnthologiesLinguistic corpora (1)
- SpeechCorpus (5)
- Written Corpus (36)
- WrittenCorpus (7)

CLARIN-IT: contatti



- www.clarin-it.it
- Twitter @CLARIN_IT
- coordination@clarin-it.it

Grazie per la vostra attenzione!!!!!!

Short summary



- Italy joined Clarin
- My message today was about the aims, the functioning of Clarin, the potentialities, what is already in place, what planned, how it can help advancing the sector;
 - appropriately document your resources
 - give them international visibility
 - find resources, combine them with tools for further analysis
 - be in touch with scholars in the field,
 - be informed about relevant events,
 - apply for mobility grant to visit colleagues for specific joined research
 - be invited at clarin related events on strategic themes
 - take part to the Annual Clarin Conference
- Now the challenge is in front of us who have to build the center, create the network and develop services for the community
- For you, DH community, is an opportunity to let Clarin-it grow based on your needs and expectations, give us the research directions.