







Corso di aggiornamento sulle Digital Humanities

Giulia Pedonese, Federico Boschetti

Verona, Educandato "Agli Angeli", 22 Marzo 2024







Cosa sono le Digital Humanities

Le Digital Humanities sono un campo di studio interdisciplinare fra informatica e discipline umanistiche al fine di coniugare il metodo scientifico-sperimentale e i metodi quantitativi con il metodo storico-filologico e i metodi qualitativi





Incorporare l'analisi dei dati nelle discipline umanistiche

Utilizzare strumenti di analisi testuale computazionale per esaminare opere letterarie, documenti storici o tendenze linguistiche, permettendo agli studenti di scoprire pattern, temi ricorrenti o evoluzioni stilistiche attraverso l'uso di software e algoritmi





Incorporare l'analisi dei dati nelle discipline umanistiche







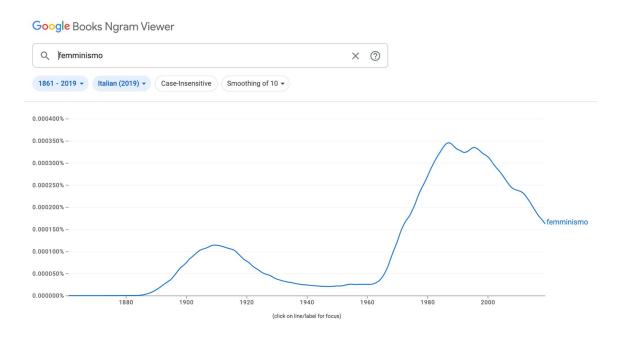
Promuovere progetti interdisciplinari

Stimolare progetti di ricerca che richiedano agli studenti di applicare metodi scientifici all'analisi di problemi umanistici, come lo studio dell'evoluzione linguistica attraverso l'analisi computazionale o l'esplorazione di tendenze culturali tramite la visualizzazione dei dati





Promuovere progetti interdisciplinari







Sviluppare competenze nel coding

Introdurre nozioni di programmazione attraverso progetti DH, come lo sviluppo di semplici siti web per esporre ricerche umanistiche o l'uso di Python per l'analisi testuale. Questo non solo migliora le competenze tecniche ma dimostra anche l'applicabilità del coding in campi non strettamente scientifici





Sviluppare competenze nel coding

```
def conta_frequenza_parole(testo):
   # Rimuovere la punteggiatura dal testo e convertire tutto in minuscole
   punteggiatura = "!\"#$%&'()*+,-./:;<=>?@[\\]^_`{|}~"
   testo_senza_punteggiatura = testo.translate(str.maketrans('', '', pun
   # Suddividere il testo in parole
   parole = testo_senza_punteggiatura.split()
   # Creare un dizionario per tenere traccia della frequenza di ogni paro
   frequenze = {}
   for parola in parole:
       if parola in frequenze:
           frequenze[parola] += 1
           frequenze[parola] = 1
   return frequenze
```





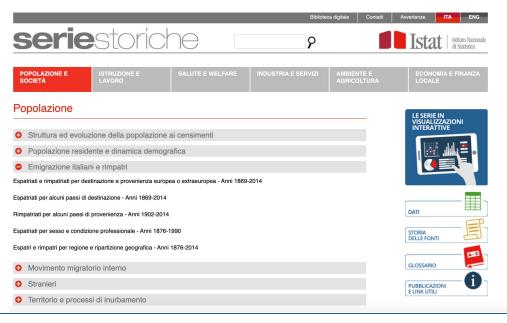
Utilizzare la visualizzazione dei dati per interpretare l'arte e la storia

Applicare strumenti di visualizzazione per analizzare e interpretare dati storici, demografici o artistici, permettendo agli studenti di scoprire connessioni e tendenze visive che arricchiscono la comprensione dei fenomeni umanistici





Utilizzare la visualizzazione dei dati per interpretare l'arte e la storia







Esplorare la modellazione 3D, la realtà virtuale e la realtà aumentata

Ricreare virtualmente siti archeologici, edifici storici o oggetti d'arte per studi più approfonditi, collegando la tecnologia alla comprensione del patrimonio culturale e storico





Esplorare la modellazione 3D, la realtà virtuale e la realtà aumentata











Analisi e critica dei media digitali

Incoraggiare gli studenti a esaminare criticamente i media digitali, come i social media, i videogiochi o i siti web, per comprendere il loro impatto sulla società e sulla cultura, utilizzando al tempo stesso strumenti digitali per condurre tali analisi





Analisi e critica dei media digitali



Home > When Photos Lie



This fake photo has been around since at least 2011. The real shot of the shark was taken off the coast of South Africa in 2005.

THOMAS P. PESCHAK/National Geographic Creative (real); via Twitter (fake)





Progetti collaborativi online

Sfruttare piattaforme collaborative per permettere agli studenti di lavorare insieme su progetti di ricerca interdisciplinari, sviluppando al contempo competenze di lavoro di gruppo e di comunicazione digitale





Progetti collaborativi online







CLARIN, CLARIN-IT e H2IOSC per le Digital Humanities





Fonti

Questa presentazione è il risultato dell'adattamento e della modifica delle seguenti fonti:

- van der Lek, Iulianna; Fišer, Darja. (2023). Introduction to Language Data: Standards and Repositories. In <u>UPSKILLS</u> Learning Content. https://upskillsproject.eu/project/standards_repositories/. CC BY 4.0.
- van der Lek, I., Fišer, D., Samardzic, T., Simonovic, M., Assimakopoulos, S., Bernardini, S., Milicevic Petrovic, M., & Puskas, G. (2023). Integrating research infrastructures into teaching:
 Recommendations and best practices (Versione 2) https://doi.org/10.5281/zenodo.8114407. CC BY 4.0.
- CLARIN ERIC Official Website: https://www.clarin.eu/ . CC BY 2.0





Piano della presentazione

- 1. Breve introduzione a CLARIN, CLARIN-IT e il progetto H2IOSC
- 2. Accesso ai servizi
- 3. Panoramica dei servizi di deposito
- 4. Uso dei CLARIN core services per le lingue di interesse: Latino, Greco Antico, Ebraico, Aramaico, Arabo, Copto, Armeno, Etiopico e Siriaco





Cos'è CLARIN

CLARIN è acronimo di Common Language Resources and Technology Infrastructure

CLARIN è un'infrastruttura digitale distribuita che fornisce un accesso facile e sostenibile a un'ampia gamma di dati e strumenti linguistici a sostegno della ricerca nelle scienze umane e sociali e non solo. CLARIN fornisce l'accesso a dati linguistici digitali multimodali (testo, audio, video) e a strumenti avanzati con cui esplorare, analizzare o combinare questi insiemi di dati.





CLARIN...

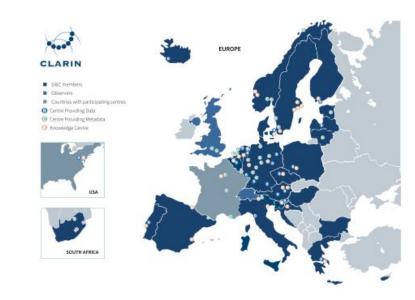
- Appartiene al cluster Social Sciences and Humanities (SSH) ed è parte integrante dello European Open Science Cloud
- Ha lo status di ERIC dal 2012, ESFRI Landmark dal 2016
- Fa da ecosistema per la condivisione della conoscenza e la formazione
- Fornisce un accesso facile e sostenibile ai dati linguistici digitali; a strumenti avanzati per scoprirli, esplorarli, sfruttarli, annotarli, analizzarli o combinarli attraverso un ambiente accessibile tramite single sign-on





CLARIN Oggi

- CLARIN è una rete di 72 centri partecipanti in Europa e oltre fra università, centri di ricerca, biblioteche e archivi pubblici, vd. la Mappa interattiva
- 24 membri: AT, BE, BG, CY, CZ, DK, EE,
 ES, FI, GR, HR, HU, IS, IT, LT, LV, NL,
 NO, PL, PT, SE, SI, ZA
- 2 osservatori: CH, UK
- Terza parte: Università Carnegie Mellon,
 USA







CLARIN-IT

CLARIN-IT è il nodo italiano di CLARIN

L'Italia è diventata il 16° membro a pieno titolo di CLARIN ERIC nel 2015. Il membro fondatore del Consorzio nazionale è l'Istituto di Linguistica Computazionale "Antonio Zampolli" del Consiglio Nazionale delle Ricerche (CNR-ILC). CLARIN è coinvolto nel progetto H2IOSC attraverso il suo consorzio nazionale italiano CLARIN-IT





CLARIN-IT

Il consorzio Italiano di CLARIN comprende I seguenti centri:

- <u>ILC4CLARIN</u> presso l'Istituto di Linguistica Computazionale "Antonio Zampolli" di Pisa, riconosciuto centro fornitore di servizi CLARIN (centro <u>Type B</u>)
- <u>EURAC Research CLARIN Centre</u> (ERCC) press l'EURAC Research Association di Bolzano, un centro di fornitura metadati (centro <u>Type C</u>)
- The Knowledge Centre for Computer-Mediated Communication and Social Media Corpora
 (CKCMC), un centro di tipo K che fornisce conoscenza e competenze su aspetti del dominio
 disciplinare compreso dall'infrastruttura CLARIN
- <u>The CLARIN Knowledge Centre for Digital and Public Textual Schholarship</u> (DiPText), centro virtuale distribuito fra due partner: L'Università Ca'Foscari di Venezia e CNR-ILC





II Progetto H2IOSC

Il progetto H2IOSC mira a creare un **cluster federato e inclusivo di RI** nel dominio ESFRI dell'innovazione sociale e culturale per consentire ai ricercatori di varie discipline nei settori delle scienze umane, delle tecnologie linguistiche e dei beni culturali di collaborare nella ricerca ad alta intensità di dati e calcoli. **Comprende i nodi italiani di quattro Infrastrutture** di Ricerca:

- CLARIN, Common Language Resource and Technology Infrastructure
- DARIAH, <u>Digital Research Infrastructure for the Arts and Humanities</u>
- E-RHIS, <u>European Research Infrastructure for Heritage Science</u>
- OPERAS, Open Scholarly Communication in the European Research Area for SSH





CLARIN per H2IOSC

WP8 - Training, Capacity Building

- Engagement
- Infrastruttura di training
- Metodologia condivisa
- Materiali didattici FAIR
- Prospettiva Train the Trainers



I servizi di CLARIN

- <u>Servizi di deposito</u> per garantire che le risorse linguistiche possano essere archiviate e rese disponibili alla comunità in modo affidabile e per aiutare i ricercatori a conservare le loro risorse in modo sostenibile
- <u>Virtual Language Observatory</u> fornisce un'interfaccia di facile utilizzo, che consente un processo di ricerca e scoperta di risorse provenienti da un'ampia varietà di domini
- <u>Federated Content Search</u> un motore di ricerca che si collega alle raccolte di dati locali disponibili nei centri.
- <u>Language Resource Switchboard</u> aiuta gli utenti a trovare un'applicazione web per elaborare i propri dati linguistici
- <u>Virtual Collection Registry</u> fornisce un registro in cui gli utenti possono creare e pubblicare le loro collezioni virtuali





Come accedere ai servizi CLARIN

Tutti gli utenti possono esplorare liberamente i servizi centrali di CLARIN per cercare risorse e competenze linguistiche. A causa di restrizioni di licenza, alcune risorse sono disponibili solo per gli utenti accademici e il login è richiesto utilizzando le credenziali istituzionali o le credenziali CLARIN.

Gli utenti accademici di tutti i Paesi partecipanti possono accedere e utilizzare le risorse linguistiche disponibili nei centri dati CLARIN con un accesso single sign-on attraverso la <u>CLARIN Service Provider Federation</u> utilizzando le proprie credenziali istituzionali.





Servizi di Deposito

Molti centri CLARIN offrono un servizio di deposito. Questo garantisce molti vantaggi:

- Archiviazione e cura a lungo termine delle risorse linguistiche, degli insiemi di dati e degli strumenti
- Descrizione delle risorse con metadati specifici e attribuzione di identificatori persistenti (PID), e.g. Handle, che ne consentono una facile ricerca e citazione
- Le risorse protette da password possono essere rese disponibili tramite un login istituzionale
- Una volta integrate nell'infrastruttura CLARIN, le risorse possono essere analizzate e arricchite più facilmente con vari strumenti linguistici.





Servizi di Deposito

Depositare una risorsa in CLARIN:

- Cercare un repository che si occupa dei dati di interesse nel catalogo dei <u>Depositing</u>
 <u>Services</u>
- Controllare la compatibilità dei dati nel <u>CLARIN Standards Information System</u>
- Contattare il centro di tipo B che ospita il repository per la preparazione dei dati
- Seguire la procedura descritta nelle linee guida del repository





ILC4CLARIN

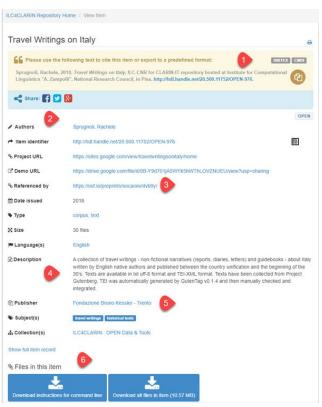
ILC4CLARIN è il CLARIN-IT B-centre istituito presso l'<u>Istituto di Linguistica Computazionale "Antonio Zampolli" di Pisa, CNR</u>. Offre servizi di deposito di dataset linguistici e strumenti per la ricerca, in particolare per l'italiano e le lingue classiche, attraverso il suo repository.

Il <u>repository istituzionale ILC4CLARIN</u> è un repository disciplinare certificato <u>CoreTrust Seal</u> e offre servizi avanzati per l'esplorazione delle risorse linguistiche e dei loro metadati (e.g. VLO, Switchboard ecc.)





Esempio di un Corpus in ILC4CLARIN



- Informazioni sulla citazione tramite handle
- 2. Campi di metadati che descrivono il corpus
- 3. Il corpus è citato in una rivista
- Il corpus è descritto, ad esempio i testi sono disponibili in formato UTF-8 e TEI-XML
- 5. Informazioni sull'editore
- 6. Istruzioni per il download



Trovare risorse linguistiche pubblicate

Virtual Language Observatory

- Un catalogo che raccoglie metadati sulle risorse linguistiche disponibili in archivi distribuiti
- Non contiene risorse linguistiche, ma aiuta a localizzarle tramite identificatori persistenti
- Anche se una risorsa ha un accesso limitato, i metadati sono sempre liberamente accessibili
- Utilizza la ricerca a faccette per restringere le ricerche





Ricerca tra le collezioni di testi

Federated Content Search

- Un motore di ricerca per individuare modelli linguistici specifici in diverse raccolte di testi
- I dati stessi rimangono presso il Centro in cui sono ospitati
- Può essere utilizzato come primo passo per scoprire dove sono ospitate risorse interessanti
- Non è possibile classificare i risultati della ricerca, ma è possibile scaricarli in vari formati ed eseguire ulteriori analisi con altri strumenti





Federated Content Search

Come funziona:

- Selezionare la lingua e il linguaggio di interrogazione (CQL)
- Specificare le raccolte e il numero di risultati da visualizzare
- Visualizzare i risultati della ricerca come Parola chiave nel contesto
- Scaricare i risultati in diversi formati di file





Trovare e Interrogare Corpora

Resource Families

- Corpora ben curati e risorse lessicali organizzate per tipo di dati e lingua
- Strumenti NLP
- Informazioni sulle licenze
- Possono essere scaricati direttamente
- Alcuni corpora sono disponibili tramite concordancers, e.g. Korp, Corpuscle o KonText





Trovare Strumenti di Analisi Linguistica

Language Resource Switchboard

- Un servizio centrale che consente di caricare un testo o un URL e di trovare uno strumento web corrispondente per analizzarlo, visualizzarlo o tradurlo
- Vi si può accedere da altri archivi e servizi di dati di ricerca, ad esempio VLO, Virtual
 Collection Registry, ARCHE, DARIAH-DE Repository, PARTHENOS VRE e TextGrid.





Utilizzare Risorse e Strumenti Linguistici

Virtual Collection Registry

- Permette di raccogliere i dataset scoperti (tramite il VLO) in una collezione virtuale
- Una collezione virtuale può contenere link a dataset, tutorial, documenti di riferimento su un argomento
- È possibile citare e condividere le raccolte virtuali con altri utenti
- Le raccolte di testi possono essere scaricate o elaborate con uno strumento Switchboard corrispondente





CLARIN Knowledge Infrastructure

Knowledge infrastructure: Un insieme di misure e strutture volte a garantire un continuo trasferimento di conoscenze tra tutti gli attori coinvolti nella costruzione, nel funzionamento e nell'utilizzo dell'infrastruttura

- Ricerca sui <u>Knowledge Centres</u> in ordine alfabetico o per specializzazione (alcune categorie sono già evidenziate
- possibilità di accedere al <u>Tour de CLARIN</u> per approfondimento su iniziative di UI e attività del network





DiPText-KC



PARTNERS PEOPLE

KNOWLEDGE

HELPDESK

SURVEYS

NEXT CLARIN CAFÉ

> CLARIN Survey 'AI in Higher Education'

CONTACT

About

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.

- · 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
- · organizing training activities (for instance webinars, workshops and summer schools).

Audiences Targeted

- Digital Philologists;
- · Computational Philologists;
- Scholars in Literary Studies:
- Art Historians;

Types of Services Offered

- Helpdesk;
- · Frequently Asked Questions;
- User Assistance;
- Training.

Languages Covered

- · Italian;
- · Ancient Greek:
- · Latin;
- Arabic;

F. Boschetti

· English;

The CLARIN CAFÉ is an informal and interactive space where the CLARIN Ambassadors hold discussions with other experts in the CLARIN community on various





Areas of Competence

DiPText-KC: Partners

Partners

DiPText-KC operates as a distributed virtual centre constituted by two partners.

[ve]dph

Ca' Foscari University of Venice – Department of Humanities – Venice Centre of Digital and Public Humanities

The main activities of the <u>Venice Centre of Digital</u> <u>and Public Humanities</u> of the Department of Humanities of the Ca' Foscari University of Venice (UniVE-DSU VeDPH) are focused on digital textual scholarship, digital archaeology, digital art and public history.

Since its foundation in 2019, UniVE-DSU VeDPH has organized seminars, webinars and summer schools in Digital and Public Humanities and has provided counseling for European, national and in-kind DH projects.



National Research Council of Italy – Institute for Computational Linguistics "Antonio Zampolli"

The Institute for Computational Linguistics
"Antonio Zampolli" of the National Research
Council of Italy (CNR-ILC) participates in DiPText-KC
through the Collaborative and Cooperative
Philology Lab (CNR-ILC CoPhiLab), focused on the
modeling and development of resources for the
digital scholarly editing and the processing of
historical languages, in particular Italian (in its
diachronic perspective), Ancient Greek, Latin and
Arabic.

CNR-ILC is the Executing Institution of CLARIN-IT, the Italian node of CLARIN, and set up ILC4CLARIN, the first CLARIN B-Centre of CLARIN-IT.





DiPText-KC: Workshops



8 MARCH 2024

Workshop on Digital Humanities

DiPText-KC is pleased to announce the Workshop
"A digital turn in the Humanities: Methodologies
and issues", which will be held in Venice, at
Palazzo Cosulich and Palazzo Vendramin, on 12
and 13 March 2024 and will also be broadcast via
Zoom. The event, organized by the Department
of Asian and North African Studies of the Ca'
Foscari University of Venice [...]

2 Paola Baroni 🗀 Workshops 🕦 48 sec read



7 FEBRUARY 2024

Workshop on Digital Scientific Editions

DiPText-KC is pleased to announce the Workshop on Accessibility, Sustainability and Reuse in the Ecosystem of Digital Scientific Editions, organized by the Italian Institute of Germanic Studies (IISG) under the patronage of UniVE, CNR-ILESI, DARIAH-IT, CNR-ILC, DiPText-KC, H2IOSC, University of Rome 'Tor Vergata' (UniRoma2) and University of Verona (UniVR). The Workshop will focus on the Digital Scholarly Editions Platform and [...]

△ Paola Baroni □ Workshops ③ 39 sec read





DiPText-KC: Helpdesk

| ame * | |
|-----------------|--|
| rst Last | NEXT CLARIN CAFÉ |
| Mail * | |
| ffiliation | CLARIN Café |
| ddressee * | |
| VeDPH vibject * | |
| opic | |
| | The CLARIN CAFÉ is an informal and |
| essage * | interactive space where the CLARIN Ambassadors hold discussions with other |
| | experts in the CLARIN community on variou topics of current interest. Next Session |
| | CLARIN Café - Creating pedagogical corpora with annotation of sensitive content and |
| uthorization * | offensive language - the CrowLL project 04/04/2024 14:00-16:00 Registration |

https://diptext-kc.clarin-it.it/helpdesk







Iniziative della comunità CLARIN

CLARIN Newsflash

aggiornamento mensile sulle attività dei consorzi nazionali

CLARIN Café

webinar informali e interattivi per discutere su temi di interesse





L'Associazione di Informatica **Umanistica e Cultura Digitale (AIUCD)**



CHI SIAMO ▼

PUBBLICAZIONI V

DIDATTICA

CENTRI

PROGETTI

CONVEGNI

INIZIATIVE V

NEWS

Q







Il convegno annuale dell'AIUCD



La conferenza AIUCD 2024, che si terrà dal 28 al 30 maggio 2024 a Catania, è organizzata dal CNR Istituto di Scienze e Tecnologie della Cognizione, dal DISUM (Dipartimento di Scienze Umanistiche) e dal CINUM (Centro di Informatica Umanistica) dell'Università di Catania, con il supporto di **CLARIN-IT**. Il tema di quest'anno è *MeTe digitali. Mediterraneo in rete tra testi e contesti*.

Il Mediterraneo è inteso nella sua accezione più inclusiva di spazio di elaborazione culturale, letteraria, artistica, filosofica e scientifica.





Strumenti per ...

- acquisire il testo tramite OCR e HTR (eScriptorium)
- analizzare linguisticamente i testi (UDPipe)
- elaborare e visualizzare dati testuali per il *distant reading* (Voyant Tools)
- interrogare dati testuali con apparato critico (MQDQ)
- ricercare temi e motivi (Memorata poetis)
- annotare i testi (Euporia)
- esplorare testi di interesse storico (Voci della Grande Guerra, Memorie di Guerra)





Grazie per l'attenzione!

Per qualsiasi dubbio o curiosità non esitate a contattarci

https://diptext-kc.clarin-it.it/helpdesk

giulia.pedonese@ilc.cnr.it federico.boschetti@ilc.cnr.it



