



Università
di Catania

ASSOCIAZIONE per
l'INFORMATICA UMANISTICA
e la CULTURA DIGITALE

Consiglio Nazionale
delle Ricerche

ME.TE. DIGITALI

MEDITERRANEO IN RETE TRA TESTI E CONTESTI

ATTI DEL XIII CONVEGNO ANNUALE
AIUCD 2024



28 - 30 MAGGIO
MONASTERO DEI BENEDETTINI
P.ZZA DANTE, 32 CATANIA

ISBN 978-88-942535-8-0



Copyright ©2024 AIUCD
Associazione per l'Informatica Umanistica e la Cultura Digitale



Il presente volume e tutti i contributi sono rilasciati sotto licenza Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). Ogni altro diritto rimane in capo ai singoli autori.
This volume and all contributions are released under the Creative Commons Attribution Share-Alike 4.0 International license (CC-BY-SA 4.0). All other rights retained by the legal owners.

A cura di: Di Silvestro Antonio; Spampinato Daria (2024). Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD, Catania 28-30 maggio 2024, Università di Catania.

Editing: Denise Bruno; Christian D'Agata; Laura Mazzagufò; Francesca Prado; Eliana Vitale; Alessandro Zammataro.

Ultimo accesso agli URL in data 15 maggio 2024.

Si prega di notificare all'editore ogni omissione o errore si riscontri: [aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria@aiucd.org)
Please notify the publisher of any omissions or errors found: [aiucd.segreteria \[at\] aiucd.org](mailto:aiucd.segreteria@aiucd.org)

Il programma della conferenza AIUCD 2024 è disponibile online <https://aiucd2024.unict.it/programma/>
The AIUCD 2024 Conference Program is available online <https://aiucd2024.unict.it/programma/>

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante *double-blind peer review* sotto la responsabilità del Comitato di Programma di AIUCD 2024.

All the papers published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2024 Program Committee.

Chair

Antonio Di Silvestro (Università di Catania)

Daria Spampinato (CNR Istituto di Scienze e Tecnologie della Cognizione)

Comitato di programma / Program committee

Emmanuela Carbé (Università di Siena)

Massimo Cultraro (CNR Istituto di Scienze del Patrimonio Culturale)

Christian D'Agata (Università di Catania)

Antonio Di Silvestro (Università di Catania)

Greta Franzini (Eurac Research)

Maurizio Lana (Università del Piemonte Orientale)

Cristina Marras (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee)

Marco Mazzone (Università di Catania)

Ouafae Nahli (CNR Istituto di Linguistica Computazionale "Antonio Zampolli")

Marianna Nicolosi-Asmundo (Università di Catania)

Marina Paino (Università di Catania)

Giuseppe Palazzolo (Università di Catania)

Jonathan Prag (University of Oxford Merton College)

Daria Spampinato (CNR Istituto di Scienze e Tecnologie della Cognizione)

Rachele Sprugnoli (Università di Parma)

Francesco Stella (Università di Siena)

Segreteria scientifica / Scientific Secretariat

Liborio Barbarino (Università di Catania)

Denise Bruno (Università di Catania)

Giulia Cacciatore (Università di Catania)

Giuseppe Canzoneri (Università di Catania)

Elisa Conti (Università di Catania)

Milena Giuffrida (Università di Catania)

Miryam Grasso (Università di Catania)

Francesca Prado (Università di Catania)

Emilio M. Sanfilippo (CNR Istituto di Scienze e Tecnologie della Cognizione)

Eliana Vitale (Università di Catania)

Alessandro Zammataro (Università di Catania)

Comunicazione istituzionale: Claudia Cantale (Università di Catania) e Area Per la Comunicazione dell'Università di Catania (ACOM).

Institutional communication: Claudia Cantale (University of Catania) and the Area for Communication of the University of Catania (ACOM)

Supporto tecnico: Rosario Agrò, Area della Terza Missione dell'Università di Catania, per la consulenza e la progettazione grafica dei materiali informativi del convegno.

Technical support: Rosario Agrò, Third Mission Area of the University of Catania, for advice and graphic design of the conference information materials.

Enti organizzatori / Organisers

AIUCD; Università di Catania: Dipartimento di Scienze Umanistiche; CNR Istituto di Scienze e Tecnologie della Cognizione; CINUM: Centro di Informatica Umanistica dell'Università di Catania.

Supporter

CLARIN-IT; Neperia Group; Storage; programma Piaceri 2020-2022, Linea 1; Parmalat-Sole.

Chair di area/ Track chair

Le culture digitali nel Mediterraneo

Cristina Marras (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee)

Paola Moscati (CNR Istituto di Scienze del Patrimonio Culturale)

Archivi ed edizioni digitali

Christian D'Agata (Università di Catania)

Greta Franzini (Eurac Research)

Analisi computazionale dei testi

Angelo Mario Del Grosso (CNR Istituto di Linguistica Computazionale "Antonio Zampolli")

Simone Reborra (Università di Verona)

Ontologie e Semantic Web

Marianna Nicolosi Asmundo (Università di Catania)

Francesca Tomasi (Università di Bologna)

Preservazione della memoria e del patrimonio digitale

Fabio Ciraci (Università del Salento)

Anna Maria Marras (Università di Torino)

Lista dei revisori /List of reviewers

Maristella Agosti (Università di Padova), **Stefano Allegrezza** (Università di Bologna), **Chiara Alzetta** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Liborio Barbarino** (Università di Catania), **Nicola Barbuti** (Università di Bari Aldo Moro), **Stefano Bazzaco** (Università di Verona), **Benedetta Bessi** (Università Ca' Foscari di Venezia), **Andrea Bolioli** (ricercatore indipendente), **Paolo Bonora** (Università di Bologna), **Federico Boschetti** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Dominique Brunato** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Flavia Bruni** (Università Gabriele D'Annunzio di Chieti-Pescara), **Marina Buzzoni** (Università Ca' Foscari di Venezia), **Alberto Campagnolo** (Université Catholique de Louvain/KULeuven), **Anna Cappellotto** (Università di Verona), **Emmanuela Carbé** (Università di Siena), **Vittore Casarosa** (CNR Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – Università di Pisa), **Fabio Ciotti** (Università di Roma "Tor Vergata"), **Fabio Ciraci** (Università del Salento), **Elisa Conti** (Università di Catania), **Salvatore Cristofaro** (CNR Istituto per il Lessico Intellettuale Europeo e Storia delle Idee), **Christian D'Agata** (Università di Catania), **Elisa D'Argenio** (HUN-REN Hungarian Research Centre for Linguistics), **Mauro De Bari** (Università di Bari Aldo Moro), **Riccardo Del Gratta** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Angelo Mario Del Grosso** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Saulo Delle Donne** (Università del Salento), **Giorgio Maria Di Nunzio** (Università di Padova), **Antonio Di Silvestro** (Università di Catania), **Filippo Diara** (Università di Torino), **Giulia Fabbris** (Università Ca' Foscari di Venezia), **Riccardo Fedriga** (Università di Bologna), **Franz Fischer** (Università Ca' Foscari di Venezia), **Greta Franzini** (Eurac Research), **Francesca Frontini** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Daniele Fusi** (Stuttgart University & VeDPH – Università Ca' Foscari di Venezia), **Carola Gatto** (Università del Salento), **Lucia Giagnolini** (Università di Bologna), **Emiliano Giovannetti** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Milena Giuffrida** (Università di Catania), **Edmondo Grassi** (Università San Raffaele di Roma), **Miryam Grasso** (Università di Catania), **Alessandro Iannella** (Università di Cagliari - Università di Pisa – Università di Torino), **Paola Italia** (Università di Bologna), **Maurizio Lana** (Università del Piemonte Orientale), **Pietro Maria Liuzzo** (Bibliotheca Hertziana), **Dominique Longrée** (Université de Liège), **Francesco Mambrini** (Università Cattolica del Sacro Cuore di Milano), **Tiziana Mancinelli** (Istituto Italiano di Studi Germanici), **Anna Maria Marras** (Università di Torino), **Cristina Marras** (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee), **Federico Meschini** (Università della Toscana), **Alessio Miaschi** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Paolo Monella** (Università Sapienza di Roma), **Ouafae Nahli** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Marianna Nicolosi-Asmundo** (Università di Catania), **Giuseppe Palazzolo** (Università di Catania), **Valentina Pasqual** (Università di Bologna), **Gianluca Pavan** (Università di Roma "Tor Vergata"), **Giulia Pedonese** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Jonathan Prag** (University of Oxford Merton College), **Simone Reborra** (Università di Verona), **Giulia Renda** (Università di Bologna), **Roberto Rosselli Del Turco** (Università di Torino), **Enrica Salvatori** (Università di Pisa), **Emilio M. Sanfilippo** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Eva Sassolini** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Pietro Sichera** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Daniele Silvi** (Università di Roma "Tor Vergata"), **Elena Spadini** (University of Basel), **Daria Spampinato** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Linda Spinazzè** (Università Ca' Foscari di Venezia), **Rachele Sprugnoli** (Università di Parma), **Francesco Stella** (Università di Siena), **Cecilia Tamagnini** (Università di Bologna), **Timothy Tambassi** (Università Ca' Foscari di Venezia), **Francesca Tomasi** (Università di Bologna), **Marco Venuti** (Università di Catania), **Fabio Vitali** (Università di Bologna).

Funzioni e sostenibilità di una piattaforma digitale per le lingue arcaiche

Michele Mallia¹, Riccardo Del Gratta², Valeria Quochi³

¹CNR Istituto di Linguistica Computazionale "A. Zampolli", Italia - michele.mallia@ilc.cnr.it

²CNR Istituto di Linguistica Computazionale "A. Zampolli", Italia - riccardo.delgratta@ilc.cnr.it

³CNR Istituto di Linguistica Computazionale "A. Zampolli", Italia - valeria.quochi@ilc.cnr.it

ABSTRACT

Questo contributo, basato sull'esperienza acquisita in un progetto di ricerca triennale dedicato allo sviluppo di tecnologie e risorse digitali nel campo degli studi linguistico-storici su lingue epigrafiche frammentarie, riflette sulla sostenibilità a lungo termine dei risultati digitali ottenuti da piccoli gruppi di ricerca nelle Digital Humanities (DH). In particolare, l'analisi si concentra sulla possibilità di integrare questi risultati nelle infrastrutture di ricerca virtuali, distribuite e federate, come CLARIN(-IT) e la recente iniziativa di federazione delle infrastrutture di ricerca italiane per le Scienze Umane e il Patrimonio Culturale, denominata 'Humanities and Heritage Italian Open Cloud' (H2IOSC).

PAROLE CHIAVE

Epigrafia Digitale; Lessicografia Computazionale; OntoLex Lemon.

1. INTRODUZIONE

Negli anni recenti, la sostenibilità è emersa come un concetto cruciale e onnicomprensivo che influenza vari ambiti, dalla salvaguardia ambientale all'uso sostenibile di dati e strumenti nella ricerca scientifica e umanistica. In particolare, si evidenzia l'importanza della "FAIRificazione" dei dati e l'adozione di modelli, formati e metodologie che aderiscono a questo paradigma. Tra questi, il paradigma dei linked open data si sta diffondendo anche in ambito umanistico e storico, offrendo formati e modelli essenziali per la condivisione dei dati. Diversi studi, come quelli indicati in [2] e [3], confermano i numerosi vantaggi nell'adottare questi approcci, soprattutto nel campo delle digital humanities. Questi approcci mirano a garantire l'accessibilità e il riuso dei dati in maniera etica e trasparente, e favoriscono l'interconnessione tra le varie risorse, facilitando così la creazione di un ecosistema di conoscenze più ampio.

Nel settore specifico dell'epigrafia digitale, si assiste a una rapida evoluzione delle tecnologie con lo sviluppo di metodi, protocolli e progetti digitali per trattare anche le lingue poco documentate, storiche e arcaiche, come quelle trattate in [11]. Un esempio significativo è il progetto EAGLE¹ (Electronic Archive of Greek and Latin Epigraphy), che unisce diverse banche dati epigrafiche e stabilisce linee guida per produrre contenuti digitali standardizzati, favorendo l'interoperabilità e il riutilizzo dei dati. Altri progetti rilevanti e più recenti includono iSicily² [8], un corpus digitale di iscrizioni su pietra dal VII secolo a.C. al VII secolo d.C., e il corpus Cretan Institutional Inscriptions³ [12], focalizzato sulle iscrizioni istituzionali di Creta. Entrambi adottano il modello TEI-EpiDoc per le iscrizioni, uno standard XML per codificare le iscrizioni, ma non integrano risorse lessico-concettuali e non mettono a disposizione dello studioso tecnologie integrate per la creazione o modifica di dati interconnessi. Il progetto LiLa (Linking Latin) [7], invece, rappresenta un'evoluzione tecnologica, integrando risorse linguistiche latine, sia lessico-concettuali sia testuali, secondo i principi FAIR e Linked Open Data, promuovendo l'interoperabilità e facilitando la connessione di risorse esterne. A differenza di iSicily o LiLa, molti altri progetti sono di breve-media durata e/o hanno scarso supporto infrastrutturale, e questo pone rischi di sostenibilità a lungo termine per i loro prodotti scientifici (specialmente software) in termini di manutenzione, distribuzione, e consultabilità pubblica.

In questo contributo, utilizziamo l'esperienza acquisita nel progetto PRIN "Lingue e Culture dell'Italia Antica. Linguistica Storica e modelli digitali" (ItAnt) per riflettere sulla sostenibilità a lungo termine dei risultati tecnologici ottenuti da piccoli gruppi di ricerca nel contesto di progetti di breve-media durata nel campo delle Digital Humanities (DH). Attraverso l'analisi della piattaforma DigItAnt, questo lavoro mira a delineare un quadro chiaro delle potenzialità e delle sfide legate all'integrazione delle tecnologie digitali negli studi antichistici, evidenziando come tali strumenti possano arricchire la comprensione e l'interpretazione dei dati storico-linguistici. Infine, esaminiamo in particolare le modalità con cui questi risultati possono essere mantenuti accessibili grazie alle infrastrutture di ricerca virtuali, distribuite e federate, come

¹ <https://www.eagle-network.eu/>

² <https://sicily.classics.ox.ac.uk/>

³ <https://ilc4clarin.ilc.cnr.it/cretaninscriptions/en/>

CLARIN e la recente iniziativa Humanities and Heritage Italian Open Cloud (H2IOSC), la quale mira a federare quattro infrastrutture italiane in un cluster dedicato alle DH.

L'articolo è così strutturato: nella prima parte descriviamo la piattaforma DigItAnt, dedicata alla creazione e interrogazione di un insieme di dati interconnessi: lessici collegati a edizioni critiche di iscrizioni arcaiche, riferimenti bibliografici, vocabolari controllati). Ci concentreremo in particolare sull'applicazione di fruizione, un componente che non è stato ancora descritto.

Nella seconda parte riflettiamo sulla sostenibilità della piattaforma in ottica di migrazione verso l'infrastruttura di ricerca CLARIN-IT, al fine di garantire la sua disponibilità a lungo termine e offrire così un servizio affidabile alla comunità scientifica.

2. LA PIATTAFORMA DIGITANT

La piattaforma descritta in questo lavoro offre agli studiosi uno strumento avanzato per lo studio delle lingue arcaiche dell'Italia Antica⁴, quali Osco, Falisco, Venetico, Celtico ecc., basato sulle loro evidenze testuali, consentendo loro di creare ecosistemi di dati collegati e di renderli accessibili e interrogabili a un pubblico più ampio. Il presente lavoro si inserisce nel contesto delle ricerche condotte sullo sviluppo di strumenti grafici interattivi per la creazione e l'interconnessione di dataset linguistici. In particolare, esso prende ispirazione da alcune esperienze precedenti, quali ad esempio EFES⁵, che consente di creare interfacce per la consultazione di dati codificati in TEI-EpiDoc, ed estende le funzionalità di altri lavori, come ad esempio [1], per permettere di codificare lessici direttamente compatibili con il Web Semantico. Come illustrato in [9], questa applicazione consente di integrare voci lessicali con informazioni codificate in risorse indipendenti, in particolare edizioni di iscrizioni in TEI-EpiDoc, e di stabilire collegamenti azionabili con altri dataset esterni. L'applicazione mira anche a rendere questi dati fruibili e interrogabili attraverso un'interfaccia grafica intuitiva, arricchendo l'esperienza utente con un approccio che rispetta al contempo le prassi condivise delle discipline informatico-umanistiche e dei Linguistic Linked Open Data.

La progettazione della piattaforma DigItAnt è centrata su un'architettura orientata ai servizi (SOA) in cui le funzionalità di elaborazione, potenzialmente generiche, sono separate dalle interfacce grafiche rivolte all'utente e dunque specifiche per il caso d'uso particolare. DigItAnt espone due interfacce utente, corrispondenti a due modalità di utilizzo: un'applicazione è dedicata alla creazione e/o revisione di lessici storici (conformi al modello OntoLex-Lemon) collegati ai testi epigrafici (in XML secondo TEI-EpiDoc), alla bibliografia (gestita con Zotero) e ad altre risorse rilevanti disponibili come Linguistic Linked Open Data (LOD); l'altra applicazione è dedicata alla consultazione e interrogazione incrociata di questi dati. Questa integrazione mira a democratizzare l'accesso e la comprensione dei dati epigrafici e lessicografici, rendendoli fruibili da un vasto spettro di utenti, inclusi accademici, studenti e appassionati senza la necessità di competenze tecniche pregresse.

2.1. Architettura funzionale

La piattaforma DigItAnt adotta uno stile architetturale REST, caratterizzato da un'implementazione indipendente dei client e dei server. Nella Fig. 1 si vede un diagramma funzionale che descrive la struttura e le interazioni tra i vari componenti dello stack applicativo. All'interno di questo schema possiamo notare come la piattaforma di fruizione si colloca all'esterno dello stack, in quanto non deve passare attraverso altri servizi di autenticazione o altre componenti interne.

Le fonti di dati principali su cui si basa la piattaforma sono due: un servizio dedicato alle iscrizioni, rappresentato da CASH, che si occupa della gestione dei materiali epigrafici in TEI/XML, e un servizio per i lessici, rappresentato da LexO [1], un back-end che si occupa della gestione dei lessici modellati tramite lo standard OntoLex-Lemon. Questi software di back-end espongono API che operano sul protocollo HTTP e scambiano dati in formato JSON, conformandosi alle specifiche OpenAPI⁶, un insieme di file d'interfaccia leggibili da macchine che descrivono, producono, consumano e visualizzano servizi REST.

Le funzionalità offerte dai servizi vengono attivate tramite due interfacce grafiche sviluppate in Angular, concepite come applicazioni web composte da diversi componenti. Ogni componente è dedicato a un insieme di diverse funzioni per la codifica o il recupero di vari aspetti delle voci lessicali e il loro collegamento con altre risorse dati sia interne che esterne, via URI, come voci di lessico, porzioni di testo, riferimenti bibliografici, vocabolari controllati e altre risorse LOD.

⁴ Si tratta di lingue attestatae tra VIII° secolo A.C e I° secolo D.C, risalente al periodo precedente alla romanizzazione della penisola italiana [5]. Per maggiori informazioni sulle lingue trattate si veda <https://www.prin-italia-antica.unifi.it/p161.html>

⁵ <https://github.com/EpiDoc/EFES>

⁶ <https://www.openapis.org/>

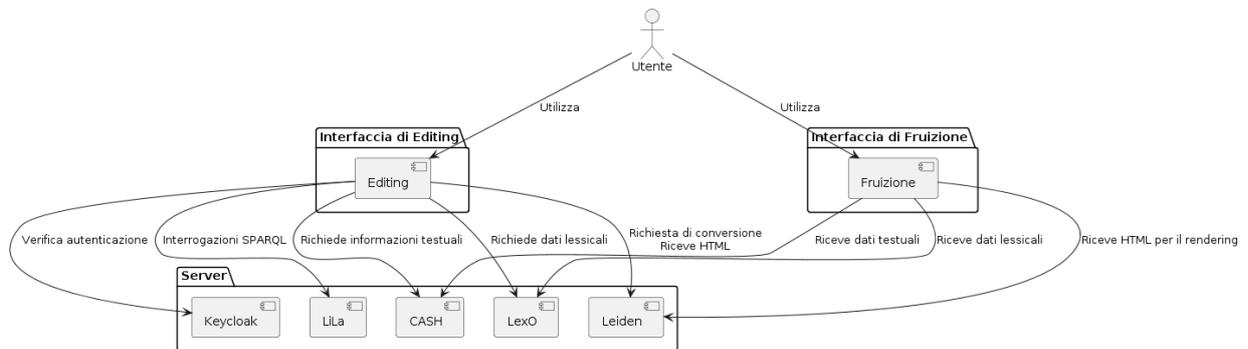


Figura 1. Schema funzionale dell'infrastruttura DigItAnt

2.2. Interfaccia di editing

L'interfaccia di editing consente di effettuare due tipologie di attività: la prima permette la creazione di un lessico conforme agli standard del Web Semantico basandosi su un corpus di testi epigrafici, mentre la seconda facilita il collegamento di elementi lessicali a token o porzioni di testo epigrafico. Nel primo caso, l'epigrafista importa un corpus di testi epigrafici in formato EpiDoc XML (già metadatoato, ricostruito e opportunamente annotato) e inizia a compilare e codificare le voci lessicali attestate nel corpus, collegandole alle rispettive forme ricostruite nel testo, alla bibliografia pertinente e, ove possibile, a risorse esterne significative. Nel secondo caso, partendo da un lessico già esistente per le lingue di interesse, l'epigrafista importa un corpus e codifica i collegamenti ai vari dati, sia interni che esterni. Inoltre può modificare e arricchire il lessico aggiungendo nuovi elementi e nuove proprietà, come ad esempio le informazioni sull'etimologia di una parola grazie ad un'integrazione nel modello OntoLex-Lemon [4]. Come si può vedere nella figura 2, il testo dell'iscrizione appare in un componente presente nella sezione centrale dell'interfaccia, il quale è denominato come "Linker", che facilita l'associazione di porzioni di testo a voci del lessico attraverso l'uso dei servizi del server CASH.

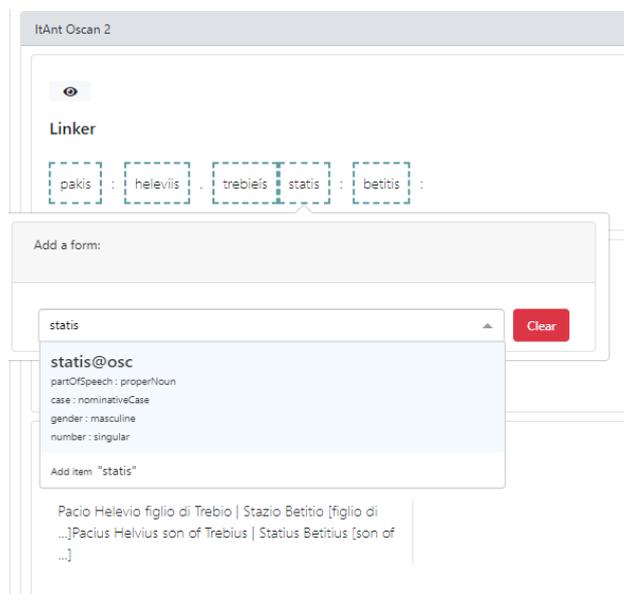


Figura 2. Componente del Linker per collegare un elemento del testo ad una forma del lessico

Il processo di collegamento tra una porzione di testo e una forma lessicale corrisponde effettivamente alla creazione di una attestazione per la forma data.

Non entriamo qui nei dettagli di questa interfaccia e dei dati perché già descritti altrove, rispettivamente in [9, 10] e [6].

3. INTERFACCIA DI FRUIZIONE

Analogamente all'interfaccia di editing, quella di fruizione è concepita come un'applicazione web avanzata, progettata per visualizzare e interagire con i dati presenti negli endpoint di back-end, precedentemente creati o importanti. Questa applicazione si distingue per la sua capacità di integrare e presentare in modo coeso dati eterogenei, anche provenienti da fonti diverse, in particolar modo dati testuali rappresentati in XML con dati linguistici-lessicali in RDF (ospitati su database a grafo). L'interfaccia di front-end funge dunque anche da orchestratore delle interazioni tra le diverse sorgenti.

Una caratteristica distintiva di questa piattaforma di fruizione, illustrata in Fig. 3, è la sua strutturata divisione in aree funzionali, ciascuna dedicata a un aspetto specifico. La sezione dedicata ai testi e alle iscrizioni permette agli utenti di accedere a dettagliate trascrizioni testuali e a informazioni contestuali, come note storiche e geografiche, fornendo così una visione completa delle iscrizioni. Questa sezione si avvale di un componente che consente di visualizzare in maniera integrata il testo epigrafico, renderizzato secondo le convenzioni di Leida, insieme alle voci lessicali codificate nel dizionario computazionale collegato. Questo componente rappresenta un'innovazione rispetto a progetti di epigrafia digitale attualmente disponibili.

Parallelamente, la sezione dei lessici si rivela uno strumento importante per gli studi in prospettiva linguistica, permettendo di navigare tra voci e forme lessicali e di accedere immediatamente a informazioni approfondite su ogni elemento selezionato. Le informazioni lessicali sono anch'esse arricchite dalle informazioni testuali e bibliografiche codificate nelle risorse collegate e, quando disponibili e codificati, da collegamenti a risorse lessicali LOD esterne, come [7]. Questa focalizzazione sugli aspetti lessicali rappresenta un'altra innovazione nel panorama delle piattaforme oggi disponibili per l'epigrafia digitale e arricchisce la comprensione delle lingue arcaiche.

La bibliografia agisce da complemento essenziale, aggregando materiali bibliografici correlati e consentendo un ulteriore approfondimento degli studi.

Infine, la funzione di ricerca avanzata, che consente interrogazioni complesse e incrociate tra i diversi dataset, rappresenta un punto di forza della piattaforma.

The screenshot shows the EpiLexO Search interface for the inscription 'ItAnt Oscan 2'. The page is titled 'ItAnt Oscan 2' and 'Curse tablet from Monte Vairano'. The text is displayed as '1 pakis : heleviis . tre(bieis) 1 statis : betitis : [---]'. A tooltip for the word 'pakis@osc' is shown, providing its morphology: 'partOfSpeech properNoun', 'case nominativeCase', and 'gender masculine'. The page also includes a map of the location and a list of metadata: Object type: tablet, Material: lead, Dimensions: Width: 1,5 cm, Height: 6,7 cm, Layout notes: Inscribed on both sides, with a line on each. Palaeographic notes: Words are separated by double dots; the second word-break on face A is a single dot. Condition: fragmentary. A detailed note states: 'The tablet was found rolled up and flattened. It broke into five fragments during the unrolling process. The inscription it is exhibited at the Museo Sannitico in Campobasso. Compared to the photographs published by , the last fragment of the tablet is now missing.'

Figura 3. Vista dell'interfaccia di fruizione, sezione testi

La combinazione armoniosa di informazioni TEI/XML e LOD all'interno di questa interfaccia non solo migliora la qualità e la profondità delle informazioni presentate, ma rappresenta anche un salto qualitativo nell'esperienza utente, che può così avvalersi di una visione integrata e multiforme delle iscrizioni e dei lessici. L'integrazione dei LOD, in particolare, mette in risalto la capacità della piattaforma di collegare le informazioni lessicali e epigrafiche a una rete più ampia di conoscenze pre-esistenti, arricchendo l'esperienza di ricerca e di studio.

4. SOSTENIBILITÀ E INTEGRAZIONE INFRASTRUTTURALE

Poiché la piattaforma è stata sviluppata nel contesto di un progetto competitivo triennale, fin dalla sua progettazione è stato previsto un collegamento con l'infrastruttura CLARIN-IT, almeno per il deposito dei risultati digitali del progetto (dataset e software). Questo non solo a tutela dell'integrità del lavoro svolto, ma anche per facilitarne il riuso. CLARIN-IT offre infatti un servizio di deposito in un archivio digitale certificato⁷ che garantisce, oltre alla conservazione e accessibilità a lungo termine, il versionamento e la citabilità delle risorse grazie all'assegnazione di identificativi univoci persistenti e alla catalogazione secondo metadati standardizzati altamente condivisi nella comunità scientifica.

Durante il progetto, si è deciso poi di trasferire anche la gestione della piattaforma DigItAnt a CLARIN-IT che potrà offrirla come servizio alla comunità, in modo da assicurare accessibilità e riutilizzo non solo dei dati, ma anche (del software) della piattaforma, anche per scopi diversi da quelli originari.

In questo senso DigItAnt rappresenta un caso di studio interessante, dal punto di vista della gestione tecnica dell'infrastruttura, per valutare la sostenibilità di questo tipo di servizi.

In questo paragrafo presentiamo quindi un'analisi preliminare, descrivendo le scelte infrastrutturali adottate e come queste possono essere migliorate in ottica di sostenibilità. La piattaforma, attualmente un prototipo operativo, risponde bene alle esigenze del progetto grazie a soluzioni software e hardware adeguate, con un accesso limitato ai collaboratori per le funzionalità di editing.

Per facilitare una transizione efficace verso un'infrastruttura più robusta, si è scelta l'adozione di Docker, che permetterà una migrazione fluida e scalabile attraverso la "containerizzazione" del software, garantendo la portabilità e l'indipendenza dal sistema operativo sottostante. Questo approccio facilita la gestione e il deployment delle applicazioni preparando la piattaforma per un aumento dell'utenza e delle richieste di accesso ai database⁸. Prevediamo che queste modifiche permetteranno alla piattaforma di sostenere un maggior numero di utenti e di gestire un volume molto più elevato di interrogazioni, consolidando la sua posizione come risorsa importante per lo studio delle lingue arcaiche dell'Italia pre-romana.

Per quanto riguarda l'utilizzo di dati, le attuali politiche di backup dell'applicazione di editing generano 12 MB di dati al giorno, con un lessico RDF che contiene 72.000 triple. Tali politiche non sono state (ancora) ottimizzate, ma potrebbero in caso aumento significativo di magnitudo. In ogni caso, la piattaforma, che occupa al momento 1 GB di spazio disco e richiede circa 4 GB di RAM per operare efficacemente, mostra una sostenibilità intrinseca dato che l'impatto delle risorse utilizzate non è eccessivo per l'infrastruttura ospitante. Anche se il progetto è quasi concluso, si prevede che la creazione e l'aggiunta di nuovi dati continuerà anche oltre il termine anche se con un ritmo diverso. In ogni caso, data la natura epigrafica frammentaria dei materiali primari e della forte specialità dell'ambito disciplinare, la quantità rimarrà naturalmente contenuta. L'impegno richiesto per mantenere la piattaforma web di fruizione è quindi gestibile, rendendo la manutenzione dell'infrastruttura praticabile ed economicamente sostenibile. Con i parametri di spazio attuali e con le quantità di dati attualmente disponibili⁹, pensiamo che si possano ospitare molte altre applicazioni che non contengano dati pesanti, come ad esempio immagini ad alta risoluzione o eventualmente altri dati multimediali che occupano un certo spazio; in quest'ultimo caso, andrebbero adottate delle strategie diverse. In questo scenario, l'elevata efficienza operativa si traduce in vantaggi sia tecnologici che economici, assicurando una gestione dei dati duratura e affidabile. Sebbene la strategia di versionamento e conservazione assicuri la disponibilità e l'integrità dei dati nel tempo, è importante sottolineare che la correlazione diretta con l'accessibilità dei dati attraverso la piattaforma può non essere immediata. I dati, conservati e versionati nel repository, richiedono un'interfaccia separata per l'accesso e l'utilizzo effettivo. Pertanto, mentre il versionamento garantisce la preservazione dei dati, l'accessibilità e l'utilizzo pratico necessitano di soluzioni aggiuntive per garantire la piena efficacia della piattaforma, riutilizzabili e conformi ai principi FAIR, favorendo così una maggiore sostenibilità e utilità a lungo termine del progetto.

5. CONCLUSIONI

L'articolo ha esplorato l'implementazione e l'integrazione di una piattaforma di creazione e fruizione sviluppata nel contesto di un progetto di ricerca triennale, sottolineando l'importanza di collegamenti strategici con infrastrutture esistenti come CLARIN-IT. Questo non solo garantisce la conservazione e la citabilità a lungo termine dei dati raccolti, ma supporta anche la loro disseminazione e il riuso all'interno della comunità scientifica.

⁷ <https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/>

⁸ Finora, la piattaforma non ha affrontato un uso massivo. Un aumento significativo dell'uso potrebbe richiedere adattamenti e ottimizzazioni. Tuttavia, l'accesso a quest'ultima limitato ai soli studiosi autorizzati, e la natura di nicchia del progetto riducono il rischio di sovraccarico. Le robuste tecnologie di back-end mitigano ulteriori rischi per l'interfaccia di fruizione, rendendo improbabili problemi legati a un uso massimo simultaneo.

⁹ Si prevede che entro la fine del progetto i dati aumenteranno, ma l'ordine di grandezza non si sposterà di molto.

Per quanto riguarda le prospettive future, la piattaforma DigItAnt continuerà a operare sotto l'egida di CLARIN-IT come una "vetrina" stabile e affidabile per i dati digitali delle lingue dell'Italia Antica. Rimarrà attiva anche nella modalità di editing per gli utenti autorizzati, permettendo l'aggiornamento e l'espansione dei dataset, in particolare dei corpora e dei lessici relativi anche ad altre lingue epigrafiche frammentarie.

Al momento non sono previste innovazioni o cambiamenti sostanziali alla piattaforma, il cui valore intrinseco risiede nella capacità di presentare un'interfaccia coerente e intuitiva per la gestione e interrogazione di dati linguistici collegati. In futuro, gli sviluppi potrebbero includere l'ottimizzazione delle funzionalità esistenti e l'ampliamento delle capacità di integrazione dei Linked Open Data, per migliorare l'interoperabilità e arricchire ulteriormente le connessioni tra diversi set di dati. Questo rafforzamento potrebbe facilitare una comprensione più profonda e una maggiore fruibilità delle informazioni in ambito epigrafico e linguistico delle lingue antiche.

6. RINGRAZIAMENTI

Questo lavoro è stato svolto nel contesto dei seguenti progetti: H2IOSC - Humanities and cultural Heritage Italian Open Science Cloud finanziato dall'Unione europea NextGenerationEU - Piano Nazionale di Ripresa e Resilienza (PNRR) - Missione 4 "Istruzione e Ricerca" Componente 2 "Dalla ricerca all'impresa" Linea di Investimento 3.1 "Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione" Azione 3.1.1 "Creazione di nuove IR o potenziamento di quelle esistenti che concorrono agli obiettivi di Eccellenza Scientifica di Horizon Europe e costituzione di reti" - Codice progetto IR0000029 - CUP B63C22000730005. Soggetto attuatore CNR; e PRIN 2017XJLE8J "Lingue e Culture dell'Italia antica. Linguistica Storica e Modelli Digitali".

BIBLIOGRAFIA

- [1] Bellandi, Andrea. «LexO: an open-source system for managing OntoLex-Lemon resources». *Language Resources and Evaluation* 55 (2021): 1093–1126. <https://doi.org/10.1007/s10579-021-09546-4>.
- [2] Da Sylva, Lyne. «Towards Linked Data: Some Consequences for Researchers in the Social Sciences and Humanities». In *Proceedings of the Association for Information Science and Technology*, a cura di Luanne Freund, 94–103. Hoboken, NJ: Wiley, 2018.
- [3] Hawkins, Ashleigh. «Archives, Linked Data and the Digital Humanities: Increasing Access to Digitised and Born-Digital Archives via the Semantic Web». *Archival Science* 22, fasc. 3 (settembre 2022): 319–44. <https://doi.org/10.1007/s10502-021-09381-0>.
- [4] Khan, Anas F. «Towards the Representation of Etymological Data on the Semantic Web». *Information* 9, fasc. 12: 304 (2018): 1–17. <https://doi:10.3390/info9120304>.
- [5] Marinetti, Anna. «Scritture e lingue dell'Italia antica». In *Le grandi vie della civiltà. Relazioni e scambi fra Mediterraneo e il Centro Europa dalla preistoria alla romanità*, a cura di Franco Marzatico, 385–91. Trento: Castello del Buonconsiglio. Monumenti e collezioni provinciali, 2011.
- [6] Murano, Francesca, Valeria Quochi, Angelo Mario Del Grosso, Luca Rigobianco, e Mariarosaria Zinzi. «Describing Inscriptions of Ancient Italy. The ItAnt Project and Its Information Encoding Process». *Journal on Computing and Cultural Heritage* 16.3, fasc. 1 (2023).
- [7] Passarotti, Marco C., e Francesco Mambrini. «Linking Latin: Interoperable Lexical Resources in the LiLa Project». In *Building new resources for historical linguistics*, a cura di Erica Biagetti, Chiara Zanchi, e Silvia Luraghi, 103–24. Pavia: Pavia University Press, 2021. <https://hdl.handle.net/10807/194955>.
- [8] Prag, Jonathan R.W., James Chartrand, e James Cummings. «I. Sicily: an EpiDoc corpus for ancient Sicily». In *Digital and Traditional Epigraphy in Context: The Proceedings of the Second EAGLE International Conference*, a cura di Silvia Orlandi, Raffaella Santucci, Francesco Mambrini, e Pietro Maria Liuzzo. Newcastle: Newcastle University, 2016.
- [9] Quochi, Valeria, Andrea Bellandi, Anas F. Khan, Michele Mallia, Francesca Murano, Silvia Piccini, Luca Rigobianco, Alessandro Tommasi, e Cesare Zavattari. «From Inscriptions to Lexica and Back: A Platform for Editing and Linking the Languages of Ancient Italy». In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, a cura di Rachele Sprugnoli e Marco Passarotti, 59–67. Marseille: European Language Resources Association, 2022. <https://aclanthology.org/2022.lt4hala-1.9>.
- [10] Quochi, Valeria, Andrea Bellandi, Michele Mallia, Alessandro Tommasi, e Cesare Zavattari. «Supporting Ancient Historical Linguistics and Cultural Studies with EpiLexO». In *CLARIN Annual Conference Proceedings*, a cura di Tomáš Erjavec e Maria Eskevich, 39–43. Prague: Czechia, 2022.
- [11] Rigobianco, Luca. «La linguistica delle lingue di attestazione frammentaria». In *Metodi e prospettive della ricerca linguistica*, a cura di Chiara Meluzzi e Nicholas Nese, 29:83–94. Ledizioni, 2022. <https://iris.unive.it/handle/10278/3762809>.
- [12] Vagionakis, Irene. «Cretan Institutional Inscriptions: A New EpiDoc Database». *Journal of the Text Encoding Initiative*, 2021, 1–21. <https://journals.openedition.org/jtei/3570>.