



Università
Ca' Foscari
Venezia



UNIVERSITÀ
DEGLI STUDI
FIRENZE



Istituto di Linguistica
Computazionale
"Antonio Zampolli"

 Consiglio Nazionale delle Ricerche

DigItAnt: A Platform for Creating and Exploring Digital Resource Ecosystems for Ancient Languages

Michele Mallia

Venezia, 20 settembre 2024 - PRIN 2017 Lingue e culture dell'Italia antica: linguistica storica e modelli digitali



DigItAnt: A Platform for Creating and Exploring Digital Resource Ecosystems for Ancient Languages

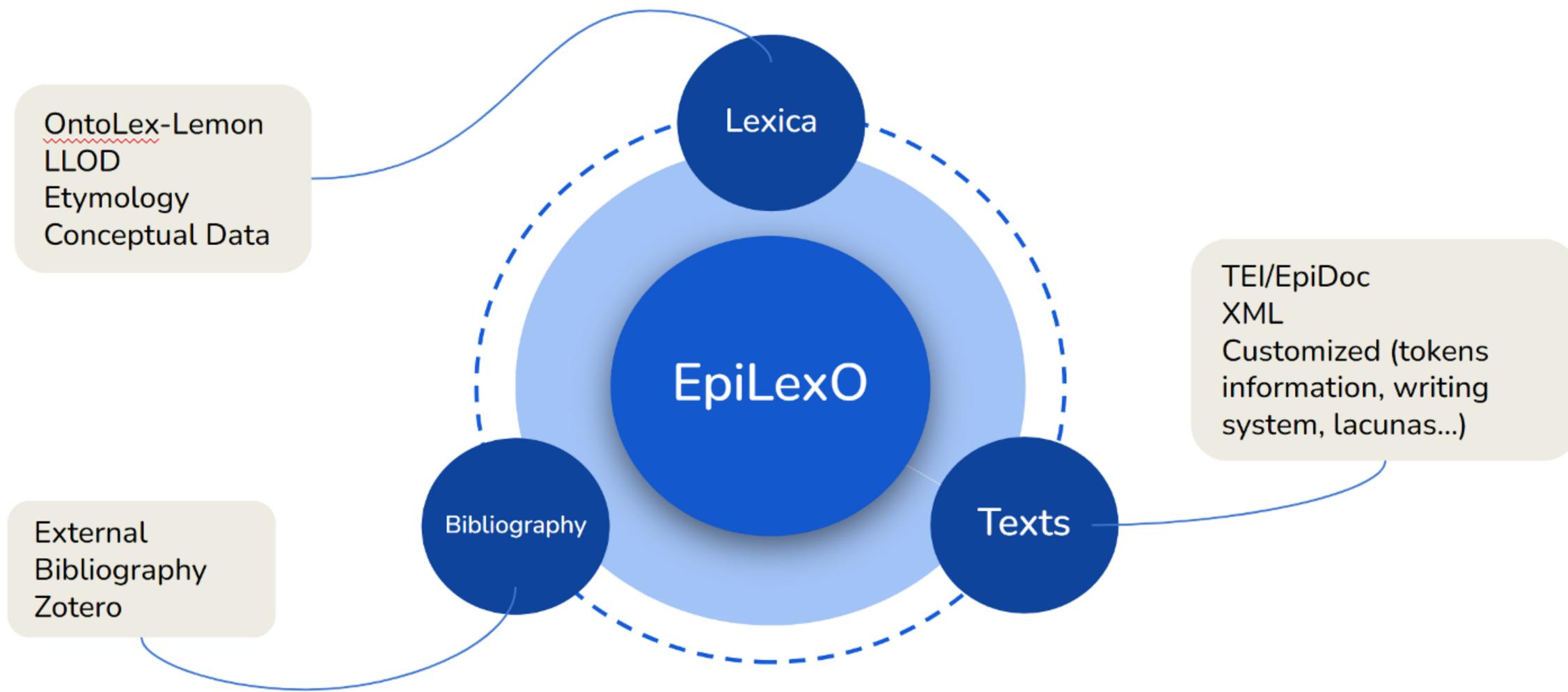
- EAGLE portal, iSicily, Cretan Institutional Inscriptions, Papyri.info...
 - Focus sull'aspetto archeologico e storico delle iscrizioni.
 - Fornire trascrizioni diplomatiche/interpretative
 - MA: (solitamente) nessun meccanismo per la creazione, l'editing o l'annotazione dei materiali e nessun o pochi altri dati interconnessi e utilizzabili.
- EFES, Recogito, INCEpTION:
 - Combinazione di testi, annotazioni e gazetteers
 - MA: nessuna facile integrazione con i dataset LOD, nessuna gestione dei formati di codifica RDF, nessuna API RESTful



Goals

- Obiettivo principale: creazione di un ecosistema di risorse linguistiche per le antiche lingue “frammentarie” (osco, neo-falisco, venetico, celtico)
 - Numero molto ridotto di testi
 - Mancanza di conoscenze
 - Necessità di codificare digitalmente sia i dati primari che quelli critici in formati condivisi
- La pubblicazione di risorse linguistiche per le lingue antiche sul Web semantico è ancora in una fase iniziale



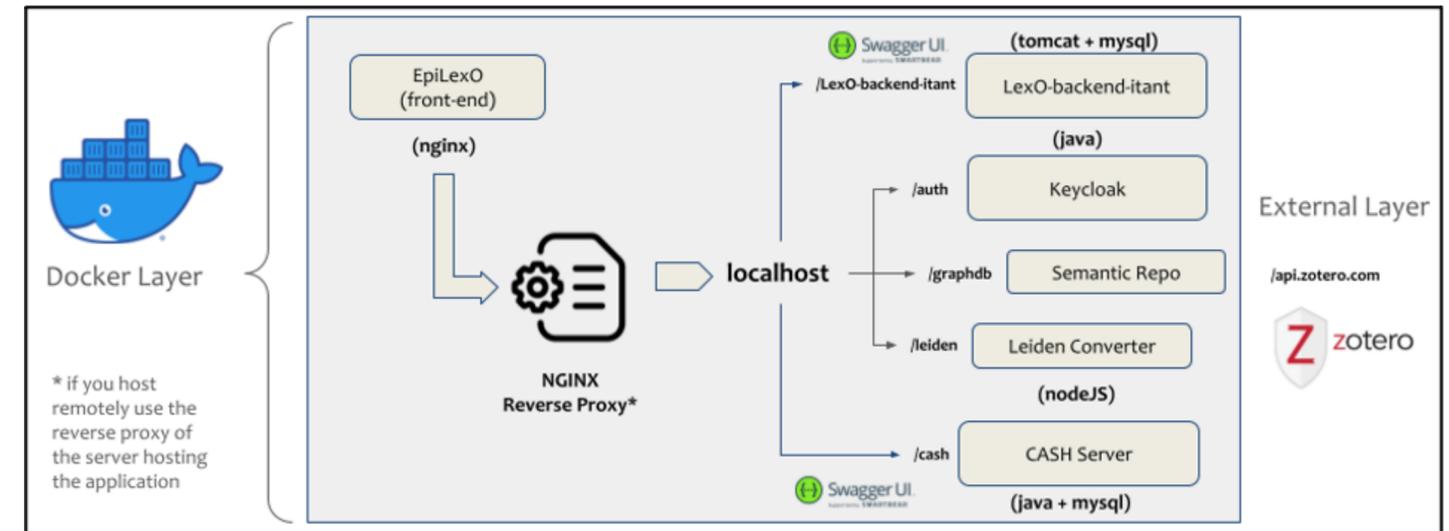


ItAnt ecosystem



DigItAnt web platform

- **Architettura:**
- Architettura Service Oriented (OpenAPI, REST)
- Due servizi principali:
 - LexO-server (A. Bellandi 2022)
 - dati lessicali, ontologie, riferimenti bibliografici...
 - CASH-server (C. Zavattari & A. Tommasi 2023 in preparazione)
 - gestione di documenti, metadati, testi & annotazioni
- Due interfacce scritte in Angular: EpiLexo & DigItAnt-search
- Manutenzione architetturale: Docker
- Progettato per facilitare la collaborazione tra studiosi
 - lavoro simultaneo sugli stessi set di dati
 - tracciare il contributo dell'utente



EpiLexO. Editor motivations

- Obiettivo principale: facilitare i linguisti storici nel loro lavoro quotidiano di descrizione e codifica dei lessici delle lingue arcaiche.
 - Collegare le fonti primarie e secondarie alle risorse disponibili e pertinenti
 - Preparazione di risorse lessicali conformi al LOD
- Necessità di un'applicazione che si occupi sia di lessici che di testi:
 - Collegare le forme lessicali alle loro attestazioni (frac, Chiarcos et al. 2022)
 - Rappresentare le informazioni etimologiche (lemonEty, Khan 2018),
 - Sfruttare le risorse esistenti per evitare la duplicazione degli sforzi.
- ==> Ecosistema di risorse collegate per facilitare la ricerca e la collaborazione



Back-end – Dati testuali

The image shows a Swagger UI interface for an API definition. At the top left is the Swagger logo with the text "Supported by SMARTBEAR". To the right of the logo is a search bar containing the text "/cash_itant/v3/api-docs" and an "Explore" button. Below the search bar, the main heading is "OpenAPI definition" with a "v0" badge and an "OAS3" badge. Underneath this heading is the URL "/cash_itant/v3/api-docs".

A "Servers" section contains a dropdown menu with the selected value "http://digitant.ilc.cnr.it - Generated server url".

The main content area is titled "annotation-controller" and contains a list of API endpoints:

- PUT** /api/annotation
- POST** /api/annotation
- POST** /api/unstructured
- POST** /api/token

Below the endpoints is a "Parameters" section with a "Try it out" button. The parameters table has two columns: "Name" and "Description".

Name	Description
requestUUID * required	requestUUID

Back-end – Dati lessicali

Swagger
powered by SMARTBEAR

https://digitant.lic.cnr.it/LexO-backend-itant_itant/service/swagger.json [Explore](#)

LexO-server API documentation ^{V0.1}

[Base URL: /LexO-backend-itant/service]
https://digitant.lic.cnr.it/LexO-backend-itant_itant/service/swagger.json

OntoLex-Lemon lexica manager
LexO-backend

Schemes
HTTPS

Lexicon creation

- GET /lexicon/creation/component Multiword component creation
- GET /lexicon/creation/etymologicalLink Etimological link creation
- GET /lexicon/creation/lexicoSemanticRelation Creation of a reified Lexico-Semantic Relation
- GET /lexicon/creation/translationSet Creation of a translation set
- GET /lexicon/creation/conceptSet Concept Set creation
- GET /lexicon/creation/lexicalEntry Lexical entry creation
- GET /lexicon/creation/lexicalSense Lexical sense creation
- GET /lexicon/creation/etymology Etimology creation
- GET /lexicon/creation/lexicalConcept Lexical Concept creation
- POST /lexicon/creation/bibliography Bibliographic reference creation
- GET /lexicon/creation/form Form creation



```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="https://www.stoa.org/epidoc/schema/9.2/tei-epidoc.rng"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="https://www.stoa.org/epidoc/schema/9.2/tei-epidoc.rng"
schematypens="http://purl.oclc.org/dsdl/schematron"?>
<tei:TEI xmlns:dcr="http://www.isocat.org/ns/dcr" xml:id="ItAnt_Oscan_2"
xmlns:tei="http://www.tei-c.org/ns/1.0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <tei:teiHeader>
    <tei:fileDesc>
      <tei:titleStmnt>
        <tei:title type="ItAnt_ID">Inscription ItAnt Oscan 2</tei:title>
      </tei:titleStmnt>
      <tei:editionStmnt>
        <tei:edition>
          <tei:idno>ItAnt Oscan 2</tei:idno>
        </tei:edition>
        <tei:editor>
          <tei:persName>Francesca Murano</tei:persName>
        </tei:editor>
      </tei:editionStmnt>
      <tei:publicationStmnt>
        <tei:authority>PRIN 2017 "Lingue e culture dell'Italia antica: linguistica storica e modelli digitali | Languages and Cultures of Ancient Italy. Historical Linguistics and Digital Models"</tei:authority>
        <tei:date>2023</tei:date>
        <tei:idno type="URI" subtype="handle"></tei:idno>
        <tei:availability>
          <tei:licence target="https://creativecommons.org/licenses/by-nc-sa/4.0/">This file is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license - (CC BY-NC-SA 4.0).
        </tei:licence>

```

Title	Creator	D.	Item Type
Ombriens, Sabins, Picéniens, peuples sabelliques des Abruzzes : un...	Aberson and Wa...	2...	Journal Article
Entre archéologie et histoire : dialogues sur divers peuples de l'Itali...	Aberson et al.	2...	Rivista di Epigrafia Italica
L'Italia centrale e la creazione di una koiné culturale? I percorsi dell...	Aberson et al.	2...	Author De Benedittis, Gianfranco
«Nos sumus Romani qui fuimus ante...» Memory of ancient Italy	Aberson et al.	2...	Publication Studi Etruschi
CRMtex. An Ontological Model for Ancient Textual Entities (Poster)	Achille and Franc...	2...	Volume 49
A Passage of Varro, "De Lingva Latina" and an Oscan Fragment of ...	Adams	2...	Pages 292-295
Bilingualism and the Latin Language	Adams	2...	Date 1981
Social Variation and the Latin Language	Adams	2...	Date Added 10/18/2021, 2:51:38 PM
The Regional Diversification of Latin. 200 BC - AD 600	Adams	2...	Date Modified 10/18/2021, 2:57:23 PM
The ending of the second person plural in Osco-Umbrian and Toch...	Adams	1...	Added By edoardomiddei
Adaptaciones del alfabeto griego	Adiego	2...	
Osco Tribuf Plifriks	Adiego	2...	
Ancora sul sostrato sudpiceno nei dialetti oschi settentrionali	Adiego	2...	
Osco central y meridional frente a osco del norte (¿ o más bien est...	Adiego Lajara	2...	
Sudpiceno vepses	Adiego Lajara	1...	
Consideraciones sobre peligno anacet(h)a (y demás variantes)	Adiego Lajara	2...	
Protosabello, osco-umbro, sudpiceno	Adiego Lajara	1...	



f="https://sws.geonames. na="https://sabapmolise. el Molise de del del Mol seumID">3974</tei:idno> tablet from Monte Vairan type="trismegistos"> ce="www.trismegistos.org

Sistema ItAnt - EpiLexO Modalità editor

- Editing/creazione lessici
- Linking al testo == associazione elementi testuali - uri di elementi lessicali;
- linking elementi lessicali/testuali a record bibliografici



Editor (UX) Design

- Tre sezioni principali:

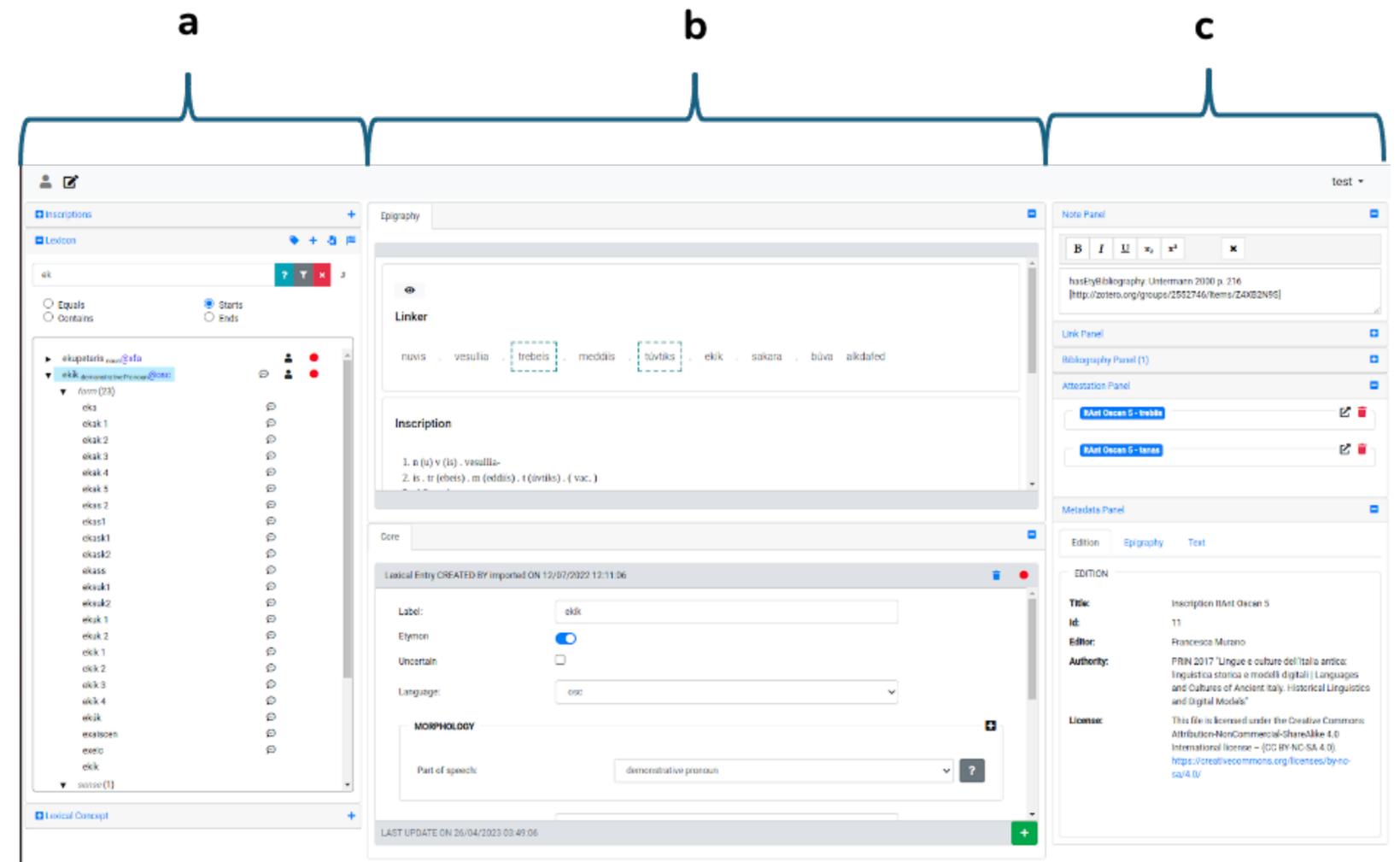
A. Navigation Trees (Texts, Lexicon and Lexical Concepts)

B. Editor Area:

- Epigraphy
- Lexicon (pivot)

C. Contextual and additional information

- Related resources
- Bibliography
- Inscriptions metadata (from Epidoc XML docs)
- Attestations



Editor

- **Caratteristiche principali:**
 - Creazione e gestione di lessici
 - Import di dati sui testi (ora iscrizioni da documenti XML Epidoc)
 - Ricerca di informazioni bibliografiche da collegare a testi e lessici
 - Interrogazione SPARQL a LiLa (per Cognates e Etymon)
 - Modifica dello stato delle voci lessicali
 - Monitoraggio dei contributi degli utenti / Log
 - Gestione della tassonomia e dei concetti (SKOS)
 - Gestione degli utenti

Lexical Entry CREATED BY Imported ON 12/07/2022 12:11:06

Label:

Etymon:

Uncertain:

Language:

MORPHOLOGY

Part of speech:

Stem type:

EVOKES

COGNATES

<input type="text" value="hīc"/>	<input type="text" value="x"/>	<input type="checkbox"/>	<input type="button" value="↗"/>	<input type="button" value="↶"/>	<input type="button" value="🗑"/>
<input type="text" value="ecuc"/>	<input type="text" value="x"/>	<input type="checkbox"/>	<input type="button" value="↗"/>	<input type="button" value="↶"/>	<input type="button" value="🗑"/>
<input type="text" value="ecan"/>	<input type="text" value="x"/>	<input type="checkbox"/>	<input type="button" value="↗"/>	<input type="button" value="↶"/>	<input type="button" value="🗑"/>
<input type="text" value="eso"/>	<input type="text" value="x"/>	<input type="checkbox"/>	<input type="button" value="↗"/>	<input type="button" value="↶"/>	<input type="button" value="🗑"/>

LAST UPDATE ON 26/04/2023 03:49:06

Editor

- **Linking:**
 - Tre tipi di “linking”:
 - **Collegamento interno ai lessici**
 - collegamento di voci lessicali all'interno della stessa risorsa
 - **Collegamento (voci lessicali) a risorse lessicali esterne**
 - es: Lila
 - **Text-lexicon linking**
 - collegamento di spazi testuali a forme lessicali per la codifica di attestazioni
- Tre meccanismi per il collegamento:
 - Copia-incolla
 - Digitazione e ricerca su risorse interne
 - Digitazione e ricerca su endpoint remoto



Editor

- **Attestation**

Attestation Panel

ItAnt Oscan 5 - meddíss

author: test

uncertain:

externalRef: No info

form: *meddíss*

leiden: *m (eddiís)*

note: No info

bibliography: *No bibliography* +

```
{
  "layer": "attestation",
  "value": "http://lexica/mylexicon#meddíss_noun_osc_meddíss_form",
  "imported": false,
  "attributes": {
    "author": "test",
    "creator": "test",
    "note": "",
    "confidence": 1,
    "timestamp": "1683551330862",
    "bibliography": [],
    "leiden": " m (eddiís) ",
    "validity": "",
    "externalRef": "",
    "node_id": 112,
    "label": "meddíss",
    "lexicalEntry": "http://lexica/mylexicon#meddíss_noun_osc_entry"
  },
  "spans": [
    {
      "start": 23,
      "end": 30
    }
  ],
  "id": 1168
}
```




- [Project](#)

- [About us](#)

- [Contact](#)

- [How to cite?](#)

Benvenuti su DigItAnt – La piattaforma per le lingue dell'Italia antica.

DigItAnt è una piattaforma innovativa progettata per supportare studiosi e ricercatori nella creazione, gestione e consultazione di risorse linguistiche digitali per le lingue arcaiche frammentarie dell'Italia antica.

DigItAnt integra e connette diverse fonti di informazione e consente la loro esplorazione in un ambiente unificato e facilmente accessibile.

Lo sviluppo di DigItAnt è stato finanziato dal Ministero dell'Università e della Ricerca nell'ambito del programma Progetti di ricerca di Rilevante Interesse Nazionale (PRIN) 2017.

Interfaccia di fruizione

- Una piattaforma di esplorazione
- Diverse visualizzazioni: edizioni erudite di iscrizioni, lessico, bibliografia, concordanze, ...
- Componente Leiden personalizzato → unire i dati di testo e i dati del lessico per il collegamento incrociato e la visualizzazione interattiva
- Ricerca avanzata → esecuzione di interrogazioni trasversali complesse sull'intero ecosistema di insiemi di dati interconnessi

The screenshot displays a web interface for the 'ItAnt Oscan 6' inscription. At the top, navigation tabs include 'Home', 'Inscriptions', 'Lexicon', 'Bibliography', 'Concordances', and 'Advanced Search'. A sidebar on the left offers options: 'Show XML', 'Print/Save PDF', and 'Show/Hide restorations'. Below the sidebar is a 'Map data' section with a map showing the location of the inscription. The main content area features a drawing of the inscription, a description: 'Inscription ItAnt Oscan 6 (drawing by Francesca Murano)', and a Creative Commons license link. A 'Text Data' section provides the following details:

IDs:	TM 170818 ST Sa 5 ImIt_Teruentum 11 Ve 153
Language:	Oscan
Alphabet:	Oscan National alphabet
Word division type:	punctuation
Inscription type:	building inscription
Ductus:	sinistrorse
Execution:	engraved
Date:	<i>approx.</i> 175-90 BC
Opisthography:	no



Interfaccia di fruizione

- Quattro sezioni:
 - texts
 - lexicon
 - bibliography
 - concordances
- Ogni sezione sarà divisa in due parti: una parte di browse e un simple search
- Consentire ai “non addetti ai lavori” di poter sfogliare i documenti/testi

Text

Intepretative

1 sten[is] [c. 10 - 12] meddiss The gap in l. 1 contained gentilicium and patronymic of the magistrate

2 túvt[ik]s . úpsannam

3 inim . prúfatted

Translation

Stenis [-], meddix tuticus (quest'opera) Stenis [-], meddix tuticus, gave to be built (i.e. commissioned) and approved (this work/monument)

Support

Object type: **stone slab**

Material: **travertine**

Dimensions: **Width:** 82 cm
Height: 55.5 cm
Depth: 19 cm

Layout notes: The text is engraved on the upper part of the block, near the top.

Palaeographic notes: Words are separated by dashes.

Condition: fragmentary

Origin and Conservation

Modern find spot name: **Pietrabbondante (Isernia)**
Found in close proximity to the façade the temple A

Discovery year: 1857

meddiss@osc
Go to lexicon

Type: **lexicalForm**

Morphology:
partOfSpeech **noun**
case **nominativeCase**
gender **masculine**
number **singular**

Distribuzione e accessibilità

- Dati lessicali esportabili/esportati in RDF LOD standard formats --> disponibili in varie modalità e licenza (CC-BY-NC-SA 4.0):
 - deposito in ILC₄CLARIN per conservazione a lungo termine, e download integrale
 - via GUI DigItAnt-search
 - via API Rest di LexO_ItAnt server
 - via SPARQL End Point (--> H2IOSC LLOD Platform)
- edizioni digitali iscrizioni (CC-BY-NC-SA 4.0):
- corpus EpiDoc originale: deposito in ILC₄CLARIN, GitHub, GUI DigItAnt-search



ItAnt & CLARIN-IT / H2IOSC

- **H2IOSC LLOD Platform:**

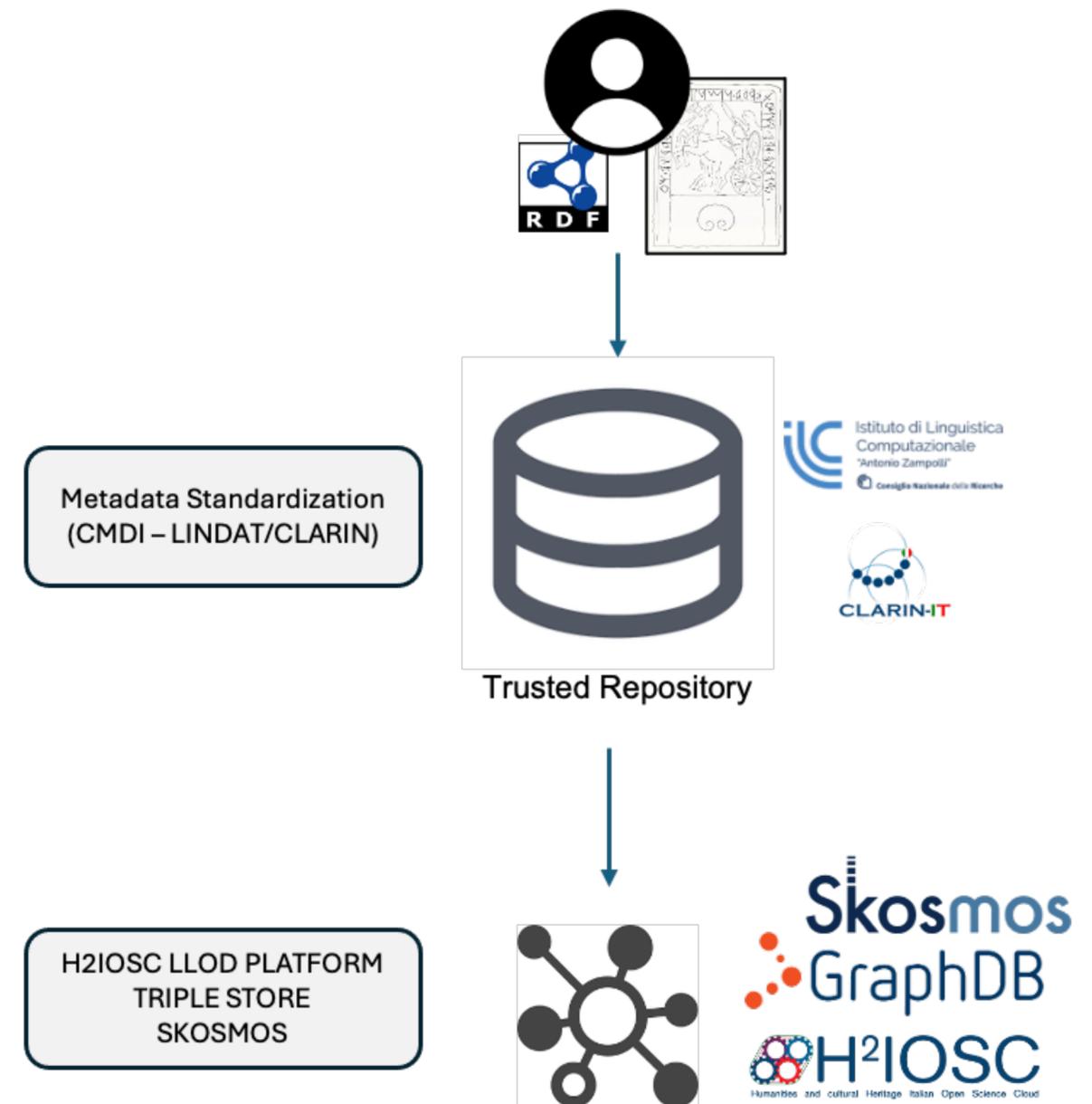
- •Un «hub» dedicato alla raccolta di risorse e strumenti per LLOD
- •I dati verranno importati e resi disponibili
- •Lessici e campi semantici pubblicati su vocabulary platform (SKOSMOS)

- **DigitAnt search & EpiLexO_ItAnt:**

- •Mantenimento ed erogazione offerti da CLARIN-IT

- **EpiLexO software stack:**

- •offerto come servizio su richiesta da CLARIN-IT in H2IOSC, anche per TNA



Bibliografia

- Bandini, Michela & Quochi, Valeria. [2024](#). EpiLexO - User Manual (Version 1).
- Boschetti, Federico, Rigobianco, Luca, & Quochi, Valeria. [2023](#). «Domain-Specific Languages for Epigraphy: The Case of ItAnt». In CLARIN Annual Conference Proceedings 2023, 80–84.
- Mallia, Michele, Michela Bandini, e Valeria Quochi. [2024](#). «An interface for linking ancient languages». In Proceedings of the CLaDA-BG 2023 Conference: Language Technologies and Digital Humanities: Resources and Applications (LTaDH-RA), Sofia, Bulgaria, 10-12 May 2023, a cura di K. Simov e P. Osenova, 1–12. Aachen: CEUR-WS. org.
- Mallia, Michele & Quochi, Valeria. [2023](#). «Epilexo: uno strumento per la creazione, modifica e interlinking di lessici antichi».
- Quochi, Valeria, Andrea Bellandi, Fahad Khan, Michele Mallia, Francesca Murano, Silvia Piccini, Luca Rigobianco, Alessandro Tommasi, e Cesare Zavattari. [2022](#). «From Inscriptions to Lexicon and Back: A Platform for Editing and Linking the Languages of Ancient Italy». In LREC 2022 Workshop Language Resources and Evaluation Conference. Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022). Proceedings, 59–67. Marseille: European Language Resources Association.
- Quochi, Valeria, Andrea Bellandi, Michele Mallia, Alessandro Tommasi, e Cesare Zavattari. [2022](#). «Supporting Ancient Historical Linguistics and Cultural Studies with EpiLexO». In CLARIN Annual Conference Proceedings, a cura di Tomaž Erjavec e Maria Eskevich, 39–43. Prague: CLARIN.

