

Aspetti funzionali e implementativi del Museo epigrafico digitale *EpiCUM*

Salvatore Cristofaro, Daria Spampinato

Istituto di Scienze e Tecnologie della Cognizione - CNR, Italia
{salvatore.cristofaro,daria.spampinato}@istc.cnr.it

Abstract

In questo lavoro vengono descritte le principali attività di sviluppo del progetto *EpiCUM* con particolare riferimento agli aspetti realizzativi e funzionali dei risultati conseguiti. *EpiCUM* si propone di presentare e rendere fruibile con un unico *museo digitale* tutto il corpus epigrafico del museo civico Castello Ursino di Catania, codificato in *EpiDoc*. Il progetto nasce dalla collaborazione tra l'Istituto di Scienze e Tecnologie della Cognizione del CNR e il Comune di Catania, ha visto l'interesse del progetto *I.Sicily* ed il coinvolgimento del liceo artistico M.M. Lazzaro di Catania con le attività di alternanza scuola-lavoro. La prima fase del progetto è stata dedicata soprattutto alla ricognizione delle epigrafi e all'allestimento della mostra *Voci di Pietra* che propone una selezione di epigrafi secondo modalità di esposizione innovative attraverso l'uso intelligente della tecnologia e del digitale. Nella fase successiva ci si è concentrati sull'analisi e sulla codifica nel formato aperto *EpiDoc* di tutte le iscrizioni del museo civico e sulla realizzazione del museo digitale. Tutte le informazioni sulle epigrafi sono codificate in *EpiDoc* all'interno di file *XML* e sono facilmente accessibili attraverso le varie sezioni del museo digitale in cui vengono presentate in maniera semplice e intuitiva visualizzandole all'interno di *schede epigrafiche*. Le schede epigrafiche forniscono l'accesso ai file *XML* contenenti le codifiche *EpiDoc*. Le informazioni epigrafiche possono essere interrogate opportunamente attraverso una sezione del museo digitale dedicata alle ricerche. Con questa modalità di organizzazione delle informazioni epigrafiche nel museo digitale si rende fruibile, per diversi scopi, l'intero patrimonio epigrafico del museo civico catanese sia agli studiosi del settore, ma anche a tutte le altre tipologie di possibili utenti. Si osserva che al fine di facilitare la navigazione attraverso le informazioni epigrafiche, si è pensato di includere in maniera autonoma anche le informazioni riguardanti l'ambito religioso, le collezioni di provenienza e lo status di copia delle epigrafi, codificandole all'interno dei file *XML* attraverso uno specifico set di elementi di marcatura *EpiDoc* appositamente selezionati.

This work describes the main development activities of the *EpiCUM* project with particular reference to the implementation and functional aspects of the results achieved within it. *EpiCUM* aims to present and make available within a single digital museum the whole epigraphic corpus

of the civic museum Castello Ursino of Catania, encoded in *EpiDoc*. The project stems from the collaboration between the Institute of Cognitive Sciences and Technologies of the CNR and the Municipality of Catania and has gained the interest of the *I.Sicily* project and the involvement of the M.M. Lazzaro of Catania with school-work alternation activities. The first phase of the project was mainly dedicated to the recognition of the epigraphs and to the preparation of the exhibition *Voci di Pietra* which proposes a selection of epigraphs according to innovative exposure methods through the intelligent use of technology and digital resources. The next phase focused on the analysis and coding in the open *EpiDoc* format of all the inscriptions of the civic museum and on the realization of the digital museum. All information on epigraphs are encoded in *EpiDoc* within *XML* files and are easily accessible through the various sections of the digital museum in which they are presented in a simple and intuitive way, and visualized within *epigraphic cards*. The epigraphic cards provide access to the *XML* files containing the *EpiDoc* encodings. The epigraphic information can be appropriately queried through a section of the digital museum dedicated to searches. With this method of organizing epigraphic information in the digital museum, the whole epigraphic heritage of the civic museum of Catania can be used for various purposes, both for scholars, but also for all other types of users. Note that in order to facilitate the navigation through epigraphic information, it was decided to separately include also information about the religious sphere, the provenance collections and the copy status of the epigraphs, encoding them within the *XML* files by means of a specially selected set of *EpiDoc* markup elements.

1. Introduzione

In questo lavoro vengono descritte le principali attività di sviluppo del progetto *EpiCUM* (*Epigraphs of Castello Ursino Museum*) con particolare riferimento agli aspetti realizzativo-funzionali dei risultati conseguiti. *EpiCUM* si propone di codificare in *EpiDoc*¹ e di presentare e rendere fruibile con un'unica interfaccia web tutto il corpus epigrafico del museo civico Castello Ursino di Catania. Il progetto nasce dalla collaborazione tra l'Istituto di Scienze e Tecnologie

1 *EpiDoc* (Epigraphic Documents) è un progetto internazionale open source di collaborazione tra umanisti e informatici (<https://sourceforge.net/p/epidoc/wiki/Home/>), adottato dalla comunità internazionale degli epigrafisti, con l'obiettivo di creare degli standard di codifica flessibili e scientificamente rigorosi per la realizzazione di edizioni digitali di iscrizioni antiche. Lo standard di codifica *EpiDoc* è basato su *TEI XML* (Text Encoding Initiative - <https://tei-c.org/>) e viene utilizzato in numerosi progetti di epigrafia e papirologia digitale. La pagina web <https://wiki.digitalclassicist.org/Category:EpiDoc> contiene una lista di alcuni tra i maggiori progetti che sfruttano la codifica *EpiDoc*, nonché riferimenti a varie altre risorse relative a tale codifica. (Si veda anche la pagina <http://www.stoa.org/epidoc/gl/latest/app-bibliography.html> per una bibliografia abbastanza completa su *EpiDoc*).

della Cognizione del CNR e il Comune di Catania, ha visto l'interesse del progetto *I.Sicily*² ed il coinvolgimento del liceo artistico M.M. Lazzaro di Catania con le attività di alternanza scuola-lavoro.

La prima fase del progetto, descritta in [1], è stata dedicata soprattutto alla ricognizione delle epigrafi e all'allestimento della mostra *Voci di Pietra* che propone una selezione di epigrafi secondo modalità di esposizione innovative, attraverso l'uso intelligente della tecnologia e del digitale. La mostra è stata realizzata sviluppando un nuovo punto di vista comunicativo del patrimonio epigrafico, con la ricostruzione stilizzata di un colombario, la restituzione grafica di una tomba a "edicoletta" e tramite l'uso di materiale propriamente scenografico come la ricostruzione in plexiglass di una statuetta di *Venus Victrix*. Innovativi sono stati l'uso di video, veri e propri esempi di *storytelling* per immagini e la presenza di un chiosco multimediale *touch*, che consente di esplorare il materiale esposto e di approfondire i contenuti delle epigrafi.

Nella fase successiva del progetto ci si è concentrati sull'analisi e sulla codifica nel formato aperto *EpiDoc* di tutte le epigrafi del museo civico e la realizzazione del museo digitale.³ Le informazioni epigrafiche sono state recuperate a partire dall'edizione critica a stampa di Kalle Korhonen [2] e arricchite da ispezioni autoptiche effettuate in loco al Castello Ursino. Attraverso la codifica nel formato aperto standard *EpiDoc* delle iscrizioni e la messa online dei corrispondenti file *XML* si rende fruibile, per diversi scopi, il patrimonio epigrafico del museo civico non solo agli studiosi del settore, ma anche a tutte le altre tipologie di possibili utenti. Tutte le informazioni sulle epigrafi possono essere fruite in maniera semplice e intuitiva attraverso le interfacce grafiche delle varie sezioni del museo digitale. Ogni epigrafe è rappresentata da una scheda epigrafica *HTML*, corredata da numerose immagini fotografiche, e in cui le varie informazioni epigrafiche (materiale del supporto, lingua dell'iscrizione, testo dell'iscrizione, data dell'iscrizione, etc.) sono suddivise tra diverse sezioni navigabili attraverso l'uso di *tab menu*. Tutte le schede epigrafiche includono anche una breve descrizione riassuntiva delle epigrafi e forniscono l'accesso ai file *XML* contenenti le codifiche *EpiDoc*, visualizzabili attraverso l'interfaccia grafica e scaricabili liberamente. Si osservi che vengono visualizzati le trascrizioni dei testi complete dei segni diacritici secondo il sistema di Leida [4].

Le informazioni epigrafiche possono essere interrogate opportunamente effettuando ricerche sulle epigrafi attraverso un'apposita sezione del museo digitale. È possibile effettuare varie tipologie di ricerche semplici e avanzate; inoltre, le ricerche possono essere *combinare interattivamente* con *operatori booleani* ottenendo ricerche sempre più mirate e complesse. Queste funzionalità sono gestite da un apposito *modulo di ricerca* progettato *ad hoc* per il progetto.

Al fine di ottenere rappresentazioni delle epigrafi quanto più ricche e dettagliate, e al tempo stesso con informazioni facilmente navigabili ed estrapolabili, si è pensato di includere in maniera

2 Progetto che si pone l'obiettivo della catalogazione digitale in *EpiDoc* dell'intero patrimonio epigrafico della Sicilia antica dal VII sec. a.C. al VII sec. d.C., consultabile all'indirizzo <http://sicily.classics.ox.ac.uk/>.

3 Il numero delle epigrafi ammonta, complessivamente, a 584.

autonoma tra le informazioni epigrafiche anche l'ambito religioso, le collezioni di provenienza e lo status di copia delle epigrafi,⁴ codificandole all'interno dei file *XML* attraverso uno specifico set di elementi di marcatura *EpiDoc* appositamente selezionato.

L'articolo è organizzato come segue. Nella sezione 2 vengono illustrati alcuni aspetti di carattere generale circa la rappresentazione delle epigrafi e le modalità di creazione dei file *XML* e delle schede epigrafiche, e nella sezione 3 vengono illustrate le proposte di codifiche *EpiDoc* relative all'inclusione dell'ambito religioso, delle collezioni di provenienza e dello status di copia delle epigrafi tra le varie informazioni epigrafiche. La sezione 4 descrive la procedura di trascrizione dei testi delle iscrizioni e nella sezione 5 viene presentata l'organizzazione del museo digitale sul lato *front-end* nelle sue sezioni e sottosezioni. Nella sottosezione 5.1 si discute l'organizzazione delle numerose immagini fotografiche delle epigrafi all'interno delle schede epigrafiche del museo digitale e, successivamente, nella sottosezione 5.2 viene fornita una descrizione dettagliata delle modalità di effettuazione delle ricerche sulle epigrafi nel museo digitale e l'implementazione del modulo di ricerca e altre funzionalità connesse alle ricerche. Infine, nella sezione 6 si propongono sviluppi futuri.

2. Rappresentazioni epigrafiche

Concettualmente, ogni epigrafe viene concepita come una singola *entità astratta* costituita da un insieme di *attributi* (gli *attributi epigrafici*) che caratterizzano l'epigrafe stessa. Un attributo consiste di un pezzo di informazione (il *valore dell'attributo*) che descrive una specifica proprietà o caratteristica indipendente di un'epigrafe, assieme ad un nome (il *nome dell'attributo*) che ne identifica la caratteristica descritta. Così, ad esempio, per ogni data epigrafe si hanno nome degli attributi *materiale*, *lingua dell'iscrizione*, *testo dell'iscrizione*, *luogo di origine* e *copia* che descrivono, rispettivamente le corrispondenti proprietà delle iscrizioni. Dal punto di vista implementativo, un'epigrafe viene formalmente rappresentata come un unico file *JSON*⁵ consistente di tutte quelle coppie *chiave-valore* (K, V), dove K rappresenta il nome di un attributo dell'epigrafe e V il suo corrispondente valore. A partire dai file *JSON*, attraverso l'utilizzo di un apposito *modulo di conversione* progettato *ad hoc* e implementato nel linguaggio *JAVA*, vengono generate automaticamente le schede epigrafiche *HTML* e i file *XML* contenenti le codifiche *EpiDoc* delle epigrafi stesse. Il formato *JSON* è stato scelto in virtù della sua semplice strutturazione concettuale, facilità di *retrieval* e analisi delle informazioni in esso contenute, e della leggerezza in termini di spazio occupato in memoria. Pertanto il database delle epigrafi è costituito da un insieme di file *JSON* (uno per ogni epigrafe).

Le informazioni epigrafiche per popolare i file *JSON*, vengono estratte per lo più dall'edizione a stampa di riferimento [2], e in alcuni casi da archivi catalografici conservati nel museo civico e

4 Si osservi che una consistente parte della collezione epigrafica del museo civico Castello Ursino è composta proprio da falsi e da copie settecentesche di iscrizioni.

5 *JavaScript Object Notation* - <https://www.JSON.org/>

altre volte sono desunte da osservazioni autoptiche delle epigrafi, svolte in loco al Castello Ursino. Per agevolare la creazione dei file *JSON* delle epigrafi è stato sviluppato un apposito *form HTML di raccolta dati* in cui le informazioni raccolte vengono inserite in formato *plain text* all'interno di caselle di testo etichettate con il nome degli attributi epigrafici. Dopo che il *form* viene caricato con le informazioni complete di una data epigrafe, il corrispondente file *JSON* può essere generato automaticamente in maniera completamente trasparente. La Figura 1 illustra le fasi di creazione dei file *XML* e delle schede epigrafiche *HTML* visualizzate nel museo digitale a partire dalla raccolta delle informazioni sulle epigrafi.

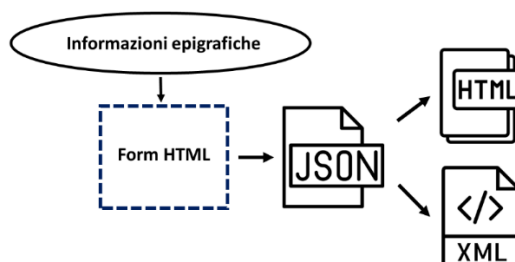


Figura 1: Fasi di creazione dei file XML e delle schede epigrafiche HTML.

3. Proposte di nuove codifiche *EpiDoc*

Come menzionato nella sezione 1, tra le varie informazioni epigrafiche sono state incluse in maniera autonoma anche l'ambito religioso, le collezioni di provenienza e lo status di copia delle epigrafi. Alla stessa stregua delle altre informazioni epigrafiche, tali informazioni vengono anch'esse concettualizzate, a livello astratto, come proprietà indipendenti delle epigrafi, descritte da corrispondenti attributi epigrafici. Dal lato della rappresentazione in *EpiDoc*, sono state proposte le seguenti codifiche.

- **Ambito religioso.** Negli schemi di codifica dei corpora epigrafici in *EpiDoc* finora disponibili, l'informazione riguardante il contenuto religioso del testo delle iscrizioni viene conservato in maniera implicita all'interno dei testi stessi; al fine di esplicitare tale informazione, in accordo con le linee guida *TEI*,⁶ si è scelto di utilizzare l'elemento di marcatura *TEI* `<domain>`, con attributo `type="rel"`, collocandolo immediatamente all'interno dell'elemento `<textDesc>`. Così, ad esempio, la proprietà che l'ambito religioso del testo di una data epigrafe è cristiano, viene codificata formalmente come segue:

```
<domain type="rel">epigrafe cristiana</domain>
```

⁶ *Text Encoding Initiative* (cfr. nota 1).

- **Collezioni di provenienza.** Per quanto riguarda la codifica dell'informazione circa le collezioni di provenienza delle epigrafi è stato usato l'elemento <provenance type="transferred" subtype="given"> all'interno dell'elemento <history>.
- **Copie.** Per la codifica dello status di copia di un'epigrafe, è stato proposto l'elemento di marcatura *TEI* <derivation> (all'interno di <textDesc>), con attributo type="copy"; secondo le specifiche della codifica *TEI*, l'elemento <derivation> consente infatti di descrivere la natura ed il grado di originalità del testo.

Per quello che ci risulta, le scelte indicate sopra per la codifica dell'ambito religioso, delle collezioni di provenienza e per le copie non sono mai state esplicitate nei diversi progetti di epigrafia digitale. In particolare, relativamente alle copie, menzioniamo il recente progetto PRIN 2015 «*False testimonianze. Copie, contraffazioni, manipolazioni e abusi del documento epigrafico antico*» che rende accessibile all'indirizzo <http://edf.unive.it> un corposo database di falsi epigrafici e copie. Le schede ricercabili delle epigrafi non dispongono, tuttavia, di link per scaricare e/o visualizzare file *XML*, né viene indicata alcun tipo di codifica eventualmente adoperata per rappresentare le epigrafi stesse.

4. Trascrizione dei testi epigrafici

Particolare attenzione è stata rivolta alla trascrizione (in *EpiDoc*) dei testi delle iscrizioni presenti sulle epigrafi e alla loro rappresentazione (in *plain text*) all'interno dei file *JSON* delle epigrafi stesse. Nella maggior parte dei casi, i testi da trascrivere sono stati prelevati dalle schede delle epigrafi incluse in [2], attraverso un processo di estrazione semi-automatizzato, a partire dalla versione *pdf* del volume. In Figura 2 vengono rappresentate l'immagine frontale di un'epigrafe (inv. n.235) con il testo e la corrispondente trascrizione presente in [2].



Δαίμοσι
Χθονίοις.
Καιλία Ἐορτή
ἐβίωσε ἔτη δ'·
5 Βλάστη θυγάτηρ
μητρὶ εὐσεβέσ-
τάτῃ ἔποισε.

Figura 2: Immagine e trascrizione testo epigrafico (inv. n.235).

Nei rari casi in cui la scheda dell'epigrafe non era presente in [2], la trascrizione è avvenuta in maniera completamente manuale, ispezionando visivamente il reperto epigrafico. Per agevolare l'operazione di trascrizione dei testi delle epigrafi prelevandoli da [2], ci si è avvalsi di tool open

source per il "riconoscimento ottico dei caratteri" (*OCR*), specialmente nel caso di testi greci o comunque testi contenenti simboli non corrispondenti a caratteri latini. Più specificatamente, partendo dalla versione *pdf* di [2], i testi da trascrivere, dopo essere stati selezionati e catturati in screenshot, sono stati convertiti tramite i tool *OCR* in formato *plain text* con codifica UTF-8,⁷ raggruppandoli temporaneamente all'interno di file di testo (.txt) in modo da essere ispezionati visivamente al fine di verificare la correttezza della conversione. E difatti, la conversione dei caratteri in UTF-8 operata dagli *OCR* utilizzati non si è dimostrata, purtroppo, particolarmente affidabile, risultando alcune volte in caratteri convertiti in maniera errata (come ad esempio nel caso di caratteri corrispondenti a lettere greche con segni diacritici). Per questo motivo si è reso necessario procedere all'inserimento manuale dei caratteri corretti in sostituzione di quelli errati, prelevandoli da una lista di caratteri UTF-8. Il testo prodotto dagli *OCR*, dopo essere stato eventualmente corretto, viene inserito nel form di raccolta dati *HTML*. Qui viene ulteriormente revisionato e completato inserendo degli appositi *metacaratteri* che identificano i vari *fenomeni testuali* (interruzioni di riga, soprallineature, caratteri ambigui, restauri editoriali, caratteri cancellati, etc.) riportati nelle trascrizioni dei testi presenti in [2] e ivi rappresentati dalle particolari convenzioni editoriali adottate (*Panciera* [4]).⁸ Il testo revisionato e completato con l'introduzione dei metacaratteri viene infine memorizzato nei file *JSON* delle epigrafi durante il processo di creazione di questi ultimi tramite il form *HTML*. La Figura 3 riporta schematicamente il processo di trascrizione dei testi delle iscrizioni descritto sopra.

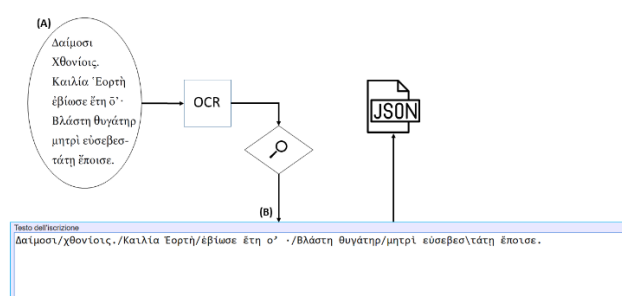


Figura 3: Trascrizione del testo di una iscrizione. (A) Screenshot del testo dell'iscrizione estratto da [2]. (B) Inserimento del testo prodotto dall'OCR all'interno del form di raccolta dati HTML.

⁷ <https://home.unicode.org/>.

⁸ Ad esempio, i metacaratteri ' / ' e \ ' , vengono utilizzati per identificare l'inizio di una nuova riga di testo nel duplice caso in cui (a) la parola che comincia la riga non è continuazione dell'ultima parola sulla riga precedente (*riga senza continuazione*), e (b) la parola che comincia la riga è continuazione dell'ultima parola sulla riga precedente (*riga con continuazione*), rispettivamente. Altri metacaratteri si riferiscono principalmente a quei fenomeni testuali rappresentati (in [2]) mediante notazioni bidimensionali (non fedelmente riproducibili quindi nel formato *plain text*) quali segni affissi sopra o sotto le lettere del testo, come lettere sottopuntate (es. α̇), che rappresentano caratteri ambigui; lettere sottolineate (es. ὕ̅), che rappresentano restauri editoriali da copie parallele; lettere soprallineate (es. m̄), etc.

Si osservi che durante la fase di creazione dei file *XML* delle epigrafi, i metacaratteri presenti nei testi completi delle iscrizioni memorizzati all'interno dei file *JSON* vengono convertiti, attraverso un processo di *riconoscimento sintattico* operato da *parser* creati *ad hoc*, nelle codifiche *EpiDoc* dei fenomeni testuali da essi individuati.⁹ L'introduzione dei metacaratteri nasce infatti dall'idea di semplificare il processo di codifica in *EpiDoc* dei testi delle iscrizioni rendendolo accessibile anche a personale non esperto.¹⁰

5. Organizzazione del museo digitale

Al fine di soddisfare l'obiettivo di promuovere online il patrimonio epigrafico del museo civico Castello Ursino di Catania sia agli utenti specialisti, sia ai visitatori occasionali e non esperti del settore, sono state realizzate diverse modalità di visualizzazione e ricerca delle iscrizioni.

Il museo digitale è stato progettato come sistema multicanale, utilizzando appropriate tecniche di responsive web design (RWD) e con opportuni fogli di stile e sono stati adottati gli standard *HTML 5*, *CSS 3* e *JavaScript*.¹¹

Sul lato *front-ent*, ogni pagina del museo digitale presenta un unico *layout* che caratterizza il museo digitale stesso, con loghi, font, menu, collegamenti ipertestuali ed altri elementi grafici utili alla navigazione e alla fruizione dei contenuti (Figura 4).



Figura 4: Front-end del museo digitale.

9 Ad esempio, i metacaratteri *'* e ** vengono rispettivamente convertiti in `<lb>` e `<lb break="no">`.

10 Si noti che i metacaratteri vengono anche usati per riprodurre graficamente i fenomeni testuali nelle schede epigrafiche attraverso un processo di riconoscimento e conversione in codifica *HTML* simile a quanto indicato prima nel caso della codifica *EpiDoc*.

11 Il museo digitale è accessibile online attraverso il link <http://epicum.istc.cnr.it/>.

L'interfaccia si compone di diverse sezioni:

EPIGRAFI con le sottosezioni **ELENCO**, **BIBLIOGRAFIA** e **RICERCA**.

- Tramite **ELENCO** si accede a una lista di tutte le epigrafi, suddivisa per pagine, con le foto, il testo dell'iscrizione e i numeri identificativi di ogni singola epigrafe. A partire dalle righe della lista si accede alle schede epigrafiche contenenti le descrizioni dettagliate delle epigrafi (Figura 5).
- In **BIBLIOGRAFIA** sono presenti i corpora e i repertori autorevoli delle iscrizioni (CIL, IG, ...) e alcuni testi in cui sono analizzate le iscrizioni del museo civico catanese.
- **RICERCA** permette di effettuare ricerche semplici ed avanzate sulle epigrafi custodite nel Castello Ursino. È possibile filtrare i risultati delle ricerche *intersecandoli* tra loro, creando l'insieme *complementare* ed estraendo, ad esempio, tutte e sole le iscrizioni che corrispondono a copie, ovvero le iscrizioni appartenenti a ben determinati periodi storici, mediante l'utilizzo, in quest'ultimo caso, di una *timeline* che raggruppa le epigrafi in intervalli temporali di un secolo. (Nella sezione 5.2 vengono descritte in dettaglio le varie funzionalità e gli aspetti tecnico-implementativi delle ricerche sulle epigrafi.)

PERCORSI permette di navigare attraverso le epigrafi secondo i seguenti percorsi tematici: Collezione di provenienza, Lingua, Luogo di provenienza, Ambito religioso, Periodo, Copie, Onomastica, Bibliografia. Ad esempio in Ambito religioso si possono visualizzare tutte le epigrafi cristiane, pagane, ebraiche oppure non identificabili. Mentre Bibliografia consente una ricerca delle iscrizioni a partire dalla bibliografia presente nella scheda epigrafica.

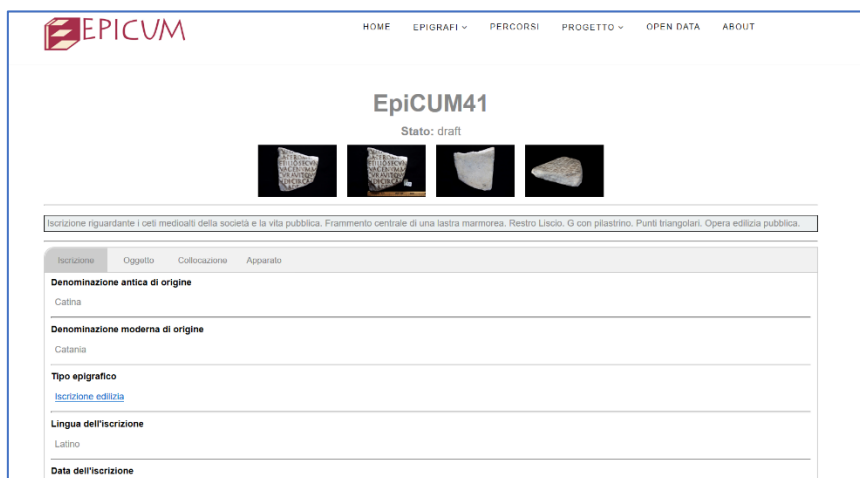


Figura 5: Esempio di scheda epigrafica.

PROGETTO con le sottovoci **DESCRIZIONE**, **PARTNER**, **DOCUMENTAZIONE**, **MOSTRA** e **CONVEGNO**.

- Nelle pagine **DESCRIZIONE**, **PARTNER** e **DOCUMENTAZIONE** viene presentato il progetto *EpiCUM* con la descrizione degli enti coinvolti e le pubblicazioni e la rassegna stampa sul progetto e sulla mostra *Voci di Pietra*.
- **MOSTRA** descrive i contenuti della mostra *Voci di Pietra* fornendo anche l'accesso alle pagine web del *chiosco multimediale* appositamente creato per la mostra al fine di favorirne l'approfondimento della visita attraverso la navigazione tra le informazioni raccolte nelle schede, nei video e nelle immagini digitalizzate, sia in italiano che in inglese.
- **CONVEGNO** visualizza una pagina con il programma del convegno scientifico internazionale "*Voci di Pietra*". *Pluralismo culturale e integrazione nella Sicilia antica e tardo-antica*, organizzato all'interno delle attività del progetto *EpiCUM*.

5.1 Immagini delle epigrafi

L'interfaccia grafica del museo digitale include svariate immagini fotografiche digitali ad alta risoluzione che ritraggono diversi particolari delle epigrafi e delle quali forniscono una rappresentazione grafica ricca e dettagliata a corredo e completamento delle descrizioni testuali presenti nelle schede epigrafiche. (Così, ad esempio, ci sono immagini che ritraggono la parte frontale (*front*) di un'epigrafe che contiene l'iscrizione, la parte posteriore (*back*) con il retro dell'epigrafe, immagini a tre-quarti (*three-quarter*), etc. Si veda la Figura 6).¹²

Complessivamente si contano 1284 diverse immagini fotografiche, corrispondenti ad altrettanti file in formato *JPEG*, per un totale di 6.62 GB di spazio occupato.



Figura 6: Immagini fotografiche ritraenti diversi particolari delle epigrafi. Da sinistra verso destra: immagine frontale; immagine posteriore; immagine a tre-quarti.

Queste immagini possono essere visualizzate attraverso il viewer *OpenSeadragon*¹³ che consente di ispezionarne i dettagli mediante operazioni combinate di *zooming* e *scrolling* (si veda la Figura 7).

12 Si osservi che la corrispondenza tra le epigrafi e le immagini fotografiche che le ritraggono non è una corrispondenza *one-to-one*, nel senso che una stessa immagine può rappresentare epigrafi differenti. Questo è il caso di due iscrizioni collocate su differenti facciate di uno stesso supporto epigrafico; a tali iscrizioni corrispondono quindi due epigrafi differenti, relazionate comunque dal fatto che l'una è il retro dell'altra. Quindi l'immagine che ritrae il front di una delle due epigrafi corrisponde all'immagine che ritrae il retro dell'altra, e viceversa.

13 <https://openseadragon.github.io/>.



Figura 7: Ingrandimento di un'immagine fotografica tramite il viewer OpenSeadragon.

Si osservi che gli *URL* dei file *JPEG* delle immagini fotografiche vengono memorizzati all'interno dei file *JSON* delle epigrafi: durante il processo di generazione automatica delle schede epigrafiche (cfr. sezione 2), essi vengono infatti estratti dai file *JSON* e quindi utilizzati per includere le immagini all'interno delle schede epigrafiche. Facciamo notare che, dato il numero elevato di immagini fotografiche da gestire, la memorizzazione degli *URL* dei file *JPEG* nei file *JSON* non è stata effettuata manualmente, bensì attraverso un processo semi-automatico a partire da una suddivisione dei file *JPEG* in differenti folder, dove l'*i*-esimo folder contiene i file *JEPG* corrispondenti all'*i*-esima epigrafe. Tramite l'esecuzione di un apposito programma *JAVA* creato ad hoc, i folder vengono ispezionati in sequenza, uno dopo l'altro, realizzando il processo di memorizzazione degli *URL*. Si osservi che durante la fase di ispezione dei folder, vengono estratte, e memorizzate all'interno dei file *JSON*, anche le informazioni relative alle dimensioni (*width*, *height*) dei file *JPEG*. Queste informazioni sono utilizzate per codificare la sezione *facsimile* nei file *XML* delle epigrafi.

5.2 Ricerca

Il museo digitale dispone di un modulo di ricerca (avanzata) che consente di effettuare sofisticate operazioni di ricerca sulle epigrafi, risultando un utile strumento di studio e di analisi della Collezione Epigrafica del Castello Ursino. È possibile effettuare un'ampia tipologia di ricerche sui diversi attributi che caratterizzano le epigrafi, ossia ricerche per *materiale*, *tipo di oggetto*, *tipo di iscrizione*, *tecnica di esecuzione*, *luogo di origine*, *collezione di provenienza*, *lingua dell'iscrizione*, etc. La Figura 8 mostra l'interfaccia grafica del modulo di ricerca del museo digitale.

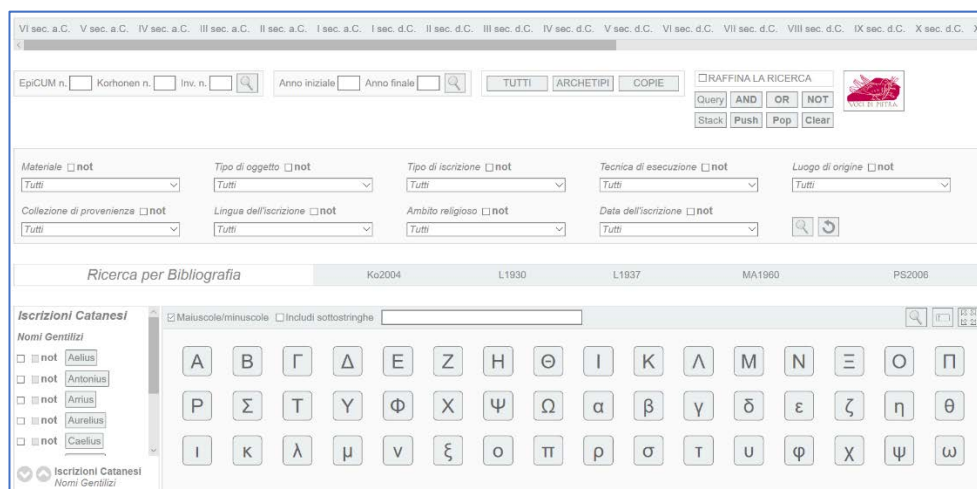


Figura 8: Interfaccia grafica del modulo di ricerca.

Ogni operazione di ricerca risulta dalla combinazione, secondo diversi *operatori booleani*, di una o più *ricerche elementari*, cioè ricerche su singoli attributi epigrafici (si veda sotto). Durante una successione di operazioni di ricerca, le singole operazioni possono essere temporaneamente memorizzate all'interno di uno *stack* dedicato,¹⁴ dal quale possono venir successivamente estratte e combinate *interattivamente* con gli operatori booleani, determinando nuove ricerche. Questa modalità interattiva di combinare le ricerche con gli operatori booleani conferisce al modulo di ricerca del museo digitale una certa flessibilità che consente, anche all'utente non particolarmente esperto, di effettuare con facilità ricerche complesse. Si osservi che è possibile visualizzare ogni volta i termini di ricerca relativi alle combinazioni di ricerche che vengono via via determinate in modo da avere un quadro dettagliato delle operazioni effettuate.

Sul lato prettamente tecnico, dal punto di vista computazionale il modulo di ricerca del museo digitale risulta essere particolarmente efficiente. La sua implementazione si basa infatti su un uso combinato di *hashmap* e *insiemi (linearmente) ordinati* che consente di ottenere prestazioni molto buone in termini specialmente di velocità di esecuzione (*time-complexity*) delle operazioni di ricerca nonché di spazio utilizzato (*space-complexity*) (si veda sotto).

Il risultato di un'operazione di ricerca è sempre un insieme (possibilmente vuoto) di epigrafi (*result-set*), ossia, l'insieme delle epigrafi soddisfacenti una determinata condizione (*query di ricerca*) stabilita in input dall'utente. Le *query di ricerca* possono essere combinate con gli *operatori booleani AND (congiunzione)*, *OR (disgiunzione)* e *NOT (negazione)* che, in termini di *result-set*, corrispondono, rispettivamente, alle operazioni insiemistiche di *intersezione*, *unione* e *complementare*. Più precisamente, date due *query* Q_1 e Q_2 , il *result-set* corrispondente alla loro congiunzione (Q_1 **AND** Q_2) (risp., disgiunzione (Q_1 **OR** Q_2)) è uguale all'intersezione (risp.

¹⁴ Si veda [2] per una descrizione dettagliata delle varie strutture dati e, più in generale, dei vari concetti e notazioni proprie del campo dell'algoritmica che vengono menzionati e utilizzati in questo lavoro.

unione) dei *result-set* corrispondenti a Q_1 e Q_2 . Similmente, il *result-set* relativo alla negazione **NOT** Q di un query Q coincide con il complementare, rispetto all'universo di tutte le epigrafi, del *result-set* corrispondente a Q . L'implementazione del modulo di ricerca garantisce che il *result-set* corrispondente ad ogni data *query* di ricerca sia sempre un insieme strettamente ordinato di epigrafi.¹⁵ Questa proprietà consente di operare in maniera particolarmente efficiente ogniquale volta i *result-set* delle *query* di ricerca devono essere combinati tra loro a formare il *result-set* di *query* più complesse ottenute attraverso combinazioni con gli operatori *booleani* come indicato sopra; cioè quando si deve computare l'intersezione, l'unione o il complementare dei *result set*.

Partendo da ricerche elementari, cioè corrispondenti a *query* di ricerca relative a singoli attributi epigrafici, la possibilità di combinare le *query* di ricerca con gli operatori booleani **AND**, **OR** e **NOT**, consente di effettuare operazioni di ricerca su attributi multipli sempre più complesse e articolate. Il modulo di ricerca del museo digitale prevede tre differenti tipologie di ricerche elementari: la *ricerca per singolo valore*, la *ricerca per range*, e la *ricerca testuale*. Il risultato di una ricerca per singolo valore consiste dell'insieme di tutte (e sole) le epigrafi per le quali il valore di un dato attributo epigrafico (di nome) A eguaglia un determinato parametro V fornito in input dall'utente. Tale tipologia di ricerca corrisponde alla query ($A = V$). Così, ad esempio, una ricerca per singolo valore può coinvolgere le epigrafi in lingua latina (*lingua dell'iscrizione* = "latino"), oppure le epigrafi aventi supporto di marmo (*materiale* = "marmo") e così via. In una ricerca per singolo valore il parametro input V può essere specificato mediante selezione da una lista precompilata di possibili valori (come nel caso delle ricerche per *materiale*, *tipo di oggetto*, *tipo di iscrizione*, *tecnica di esecuzione*, *luogo di origine*, *collezione di provenienza*, *lingua dell'iscrizione*, *ambito religioso* e *data dell'iscrizione*;¹⁶ si veda la Figura 9) oppure può essere liberamente inserito in forma di stringa all'interno di un'apposita casella di testo (come nel caso delle ricerche per numero *EpiCUM*, numero di Korhonen e numero di Inventario).

The screenshot shows a search interface with several dropdown menus and checkboxes. The 'Materiale' dropdown is open, showing options: Tutti, Calceolare, Calceolare arancione, Gesso, Marmo (highlighted), Marmo bianco, Pietra lavica, Terracotta, and Altro. Other dropdowns include 'Tipo di oggetto', 'Tipo di iscrizione', 'Tecnica di esecuzione', and 'Luogo di origine', all set to 'Tutti'. There are also checkboxes for 'not' and 'not' for 'Lingua dell'iscrizione', 'Ambito religioso', and 'Data dell'iscrizione'. Below the dropdowns is a search bar containing 'bliografia' and a table of results with columns for 'Kc2004', 'L1830', 'L1937', 'MA1960', and 'PS2008'.

Figura 9: Selezione dei valori degli attributi in una ricerca per singolo valore.

- 15 Più formalmente, si osservi che le epigrafi vengono indicizzate con i numeri interi da 1 a N , con cui vengono identificate, dove N è il numero complessivo delle epigrafi stesse, e che ogni insieme di epigrafi è rappresentato mediante l'array degli indici ad esse corrispondenti, ordinati in senso crescente. Pertanto, dal punto di vista concettuale, un *result-set* corrisponde ad un insieme di numeri interi ordinati in senso crescente.
- 16 Si noti che relativamente alle ricerche per *materiale*, *tipo di oggetto*, *tipo di iscrizione*, *tecnica di esecuzione* e *lingua dell'iscrizione*, tali liste formano sottoinsiemi dei vocabolari controllati definiti nel progetto *EAGLE* (Electronic Archive of Greek and Latin Epigraphy), accessibili all'indirizzo web <https://www.eagle-network.eu/resources/vocabularies/>.

La ricerca per *range* riguarda esclusivamente la collocazione temporale delle epigrafi. In questo caso viene specificato come parametro input un *range* R di possibili valori di date compresi tra una data iniziale I e una data finale F e si è interessati ad estrarre le epigrafi la cui data dell'iscrizione ricade in R . Il *range* R può essere specificato secondo due differenti modalità: (a) inserendo manualmente gli anni I (anno iniziale) ed F (anno finale); (b) attraverso una *timeline* suddivisa in periodi temporali della durata di un secolo ciascuno (Figura 10).



Figura 10: Timeline per la selezione del range di date.

La ricerca testuale coinvolge le epigrafi il cui testo contiene una determinata parola w specificata (liberamente) in input dall'utente. Si osservi che sono previste tre diverse modalità di ricerca testuale: *ricerca case sensitive* (default); *ricerca case insensitive* e *ricerca per sottostringhe*. Inoltre, al fine di semplificare l'inserimento della parola w da ricercare, l'interfaccia grafica dispone di una tastiera virtuale con cui è possibile inserire i diversi caratteri greci. Tale tastiera virtuale può essere espansa, a discrezione, includendo i caratteri greci con segni diacritici.

Dal punto di vista operativo, le operazioni di ricerca elementare sono state implementate come segue. Partiamo dal caso della ricerca per singolo valore. Preliminarmente, attraverso la scansione automatizzata dei file *JSON* delle epigrafi, per ogni (nome di) attributo epigrafico A , è stato estratto l'insieme S_A dei possibili valori V che può assumere A su tutte le epigrafi. Per ogni valore $V \in S_A$ è stata quindi costruita la lista ordinata $LIST_V$ (degli indici) di tutte (e sole) le epigrafi aventi V come valore dell'attributo A . Tutte le coppie $(A, \{(V, LIST_V) : V \in S_A\})$, per A (nome di) attributo epigrafico, sono state quindi raggruppate a formare un unico *oggetto JavaScript* $SvObj$, definito una volta per tutti, che rappresenta la struttura dati di base su cui si fonda il meccanismo di ricerca elementare per singolo valore. Quando si devono estrarre le epigrafi il cui attributo A possiede un dato valore V fornito come input, si effettua quindi un *look-up* all'interno dell'oggetto $SvObj$, estraendo dapprima l'insieme

$$Obj[A] = \{(V, LIST_V) : V \in S_A\},$$

usando A come chiave,¹⁷ e, successivamente, a partire da questo insieme, viene estratta la lista (possibilmente vuota) di epigrafi $SvObj[A][V] = LIST_V$, usando V come chiave. Poiché un oggetto *JavaScript* è effettivamente implementato come una *hashmap*, l'operazione complessiva di estrazione della lista $SvObj[A][V]$ delle epigrafi interessanti impiega sostanzialmente tempo costante. Una simile strategia viene usata anche nel caso della ricerca testuale riguardante l'estrazione delle epigrafi il cui testo contiene una data parola w fornita come input. Preliminarmente, tramite la scansione dei file *JSON* delle epigrafi, è stato costruito l'insieme $Words$ delle parole contenute nei testi delle epigrafi. Per ogni parola $w \in Words$ è stata quindi costruita la lista $LIST_w$ delle epigrafi il cui testo contiene w . Tutte le coppie $(w, LIST_w)$ sono poi state raggruppate in un unico oggetto *JavaScript* $WordObj$. Pertanto, la ricerca riguardante la

¹⁷ Si osservi che per un oggetto *JavaScript* Obj consistente dell'insieme di coppie chiave-valore $\{(k_1, v_1), \dots, (k_n, v_n)\}$, la notazione $Obj[k_i]$ indica il valore v_i corrispondente alla chiave k_i , per $i = 1, 2, \dots, n$.

parola w procede semplicemente mediante un look-up all'interno di *WordObj* con estrazione della $LIST_w$ usando appunto w come chiave. Anche in questo caso, il tempo impiegato è ovviamente costante $O(1)$. Per quanto riguarda infine l'implementazione della ricerca per range relativa alle date delle iscrizioni, si osservi che la data dell'iscrizione delle epigrafi viene effettivamente rappresentata come un intervallo di date $[I, F]$, con I che precede F .¹⁸ In maniera analoga ai casi precedenti, attraverso la scansione automatizzata dei file *JSON* delle epigrafi, viene costruito l'array *JavaScript DateArray* contenente le coppie $(D, LIST_D)$, dove D è un intervallo di date e $LIST_D$ è la più lunga lista ordinata di epigrafi con datazione in D . Fornito quindi in input un range R di date da ricercare, l'operazione di ricerca per range relativa alle epigrafi che si collocano temporalmente all'interno di R viene effettuata come segue. Partendo da un insieme S inizialmente vuoto di epigrafi, si confronta R con ogni coppia $(D, LIST_D)$ presente in *DateArray*: se R e D si sovrappongono, le epigrafi di $LIST_D$ vengono aggiunte a S ; in caso contrario si passa alla successiva coppia di *DateArray*. Al termine di questa procedura l'insieme S contiene le epigrafi cercate.

Dopo aver effettuato un'operazione di ricerca è sempre possibile visualizzare la *query attiva*, ossia la query a cui corrispondono gli attuali risultati della ricerca appena effettuata (si veda la Figura 11). Grazie alla specifica notazione formale adottata per rappresentare graficamente le query di ricerca (cfr. Figura 11), questa funzionalità risulta particolarmente utile per avere un riscontro dettagliato degli attributi epigrafici coinvolti nelle varie ricerche effettuate e delle condizioni imposte su di essi, specialmente nel caso di ricerche combinate complesse.

Per quanto riguarda infine le modalità di presentazione nell'interfaccia grafica dei risultati delle ricerche sulle epigrafi, si osservi che essi vengono visualizzati, in forma di elenco, all'interno di una tabella dinamica *HTML* (implementata attraverso il *plug-in* open source *DataTables* per *jQuery*¹⁹) in cui ogni riga corrisponde ad un'epigrafe del *result-set* della ricerca, e in cui vengono riportati (da sinistra verso destra): un'immagine che ritrae la parte frontale dell'epigrafe, il testo dell'iscrizione, il numero *EpiCUM* dell'epigrafe, il suo numero di Korhonen e il numero d'inventario. Cliccando sull'immagine dell'epigrafe ne viene aperta la relativa scheda epigrafica *HTML* contenente i vari dettagli descrittivi (cfr. Figura 5).

¹⁸ I valori particolari di I e F vengono desunti da uno studio delle epigrafi.

¹⁹ <https://datatables.net/>

Figura 11: Visualizzazione/rappresentazione delle query di ricerca.

6. Considerazioni finali e sviluppi futuri

Come si evince dalle descrizioni delle precedenti sezioni, l'approccio seguito nell'implementazione delle funzionalità del museo epigrafico digitale *EpiCUM*, specialmente la parte peculiare relativa alle ricerche epigrafiche, è un approccio sostanzialmente di tipo *self-contained*, che può essere inquadrato in un contesto “minimalista” relativamente all'utilizzo di strumenti di terze parti: in particolare, non è stato utilizzato nessun database strutturato preconstituito per mantenere (e interrogare) le informazioni sulle epigrafi. Le motivazioni di tali scelte risiedono principalmente nell'intenzione di collezionare e mantenere le informazioni epigrafiche in un formato aperto, leggero e facilmente accessibile, quale il formato *JSON plain text*, evitando così le varie problematiche connesse alla gestione presente e futura di database strutturati di terze parti. Vista anche la quantità non eccessiva di dati sulle epigrafi da mantenere ed interrogare, e il formato nativo scelto per rappresentare questi ultimi (cioè il formato *JSON plain text*), si è pensato di implementare algoritmi di ricerca ad hoc, specifici, appunto, per le rappresentazioni dei dati utilizzati, e per la loro mole, nell'ottica della maggiore efficienza computazionale e riduzione complessiva di *overhead*. Si preme evidenziare inoltre che le tecnologie usate per l'implementazione delle funzionalità del museo digitale (principalmente costruite nel linguaggio *JavaScript*) sono universalmente adottate e condivise e si ritiene sopravvivano per lungo tempo supportate dai browser.

Per quanto riguarda la manutenzione dei dati sulle epigrafi, si sta studiando la possibilità di raccogliere gli stessi all'interno di appropriati *repository* condivisibili che ne possano garantire la persistenza negli anni, e si stanno sviluppando degli appositi moduli software che consentono di aggiornare/modificare questi dati online attraverso una sezione dedicata del museo epigrafico digitale.

Inoltre, coerentemente con le nuove tendenze di promozione del patrimonio culturale attraverso il digitale e le tecnologie informatiche, è prevista nella sezione OPEN DATA del museo digitale

la rappresentazione semantica, secondo il paradigma dei *Linked Open Data*, dei *concetti epigrafici*, organizzandola in un'unica ontologia del patrimonio epigrafico del museo, corredata da uno *SPARQL endpoint* per interrogare i *dataset*. Lo schema ontologico completo dell'ontologia è già stato sviluppato (e incluso nella relativa sezione del museo digitale), ma si stanno ancora realizzando gli opportuni *mapping* per popolare l'ontologia stessa con i dati dei file *JSON*.

Ringraziamenti

Il progetto *EpiCUM* è stato finanziato dal Patto per Catania a valere sul Fondo Sviluppo e Coesione 2014-2020: Piano per il Mezzogiorno. Si ringrazia Jonathan Prag, responsabile del progetto *I.Sicily*, per la preziosa collaborazione e per la supervisione dell'edizione digitale della collezione.

References

- [1] Agodi, S., Cristofaro, S., Noto, V., Prag, J., Spampinato, D. 2018. “Una collaborazione tra museo, enti di ricerca e scuola: l'epigrafia digitale e l'alternanza scuola-lavoro.” *Umanistica Digitale* 2: 207-224.
- [2] Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C. 2001. *Introduction To Algorithms*. Cambridge: MIT Press.
- [3] Korhonen, K. 2004. *Le iscrizioni del Museo Civico di Catania: Storia delle collezioni - Cultura epigrafica - Edizione*. Helsinki: Societas Scientiarum Fennica.
- [4] Panciera, S. 1991. “Struttura dei supplementi e segni diacritici dieci anni dopo.” *Supplementa Italica* 8: 9-21.