15 ottobre 2026

Finanziato dall'Unione europea NextGenerationEU

MINISTERO DELLA CULTURA

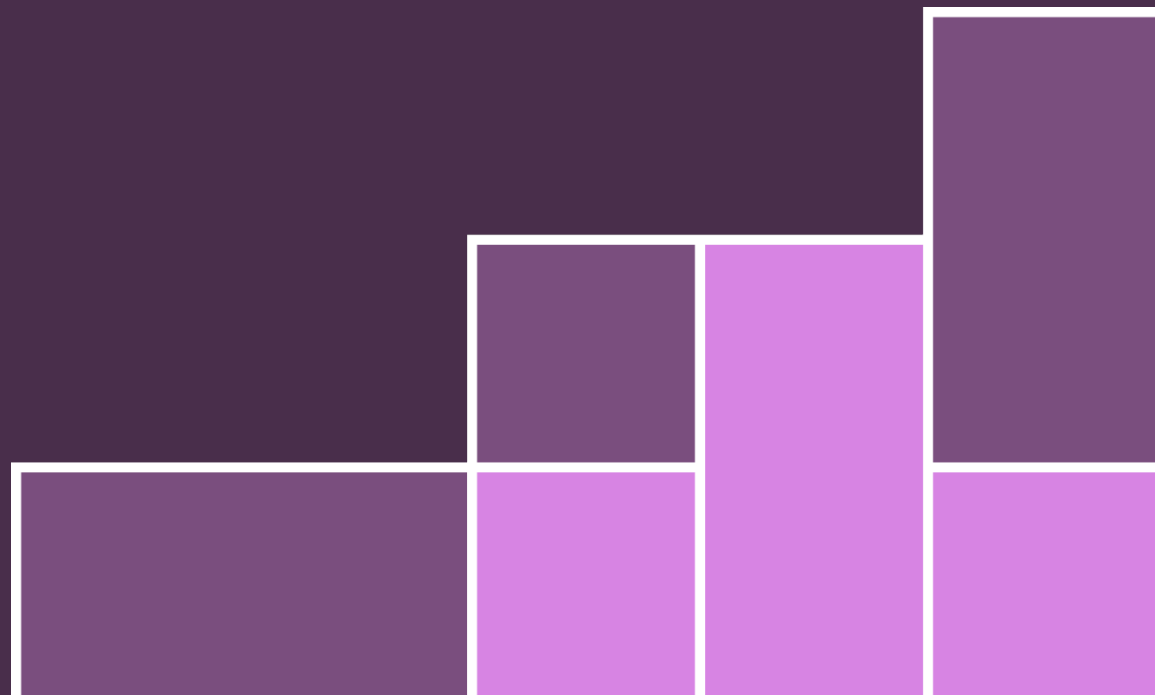
DIGITAL LIBRARY
Istituto centrale per la digitalizzazione del patrimonio culturale

Ecomic ecosistema digitale per la cultura

DPaaS Piattaforma data product as a service

Corredi funebri aragonesi

SABAP Napoli



Scopo del Data product

Rendere accessibile e fruibile un corpus di schede narrative sui personaggi sepolti nelle arche funebri aragonesi della basilica di San Domenico Maggiore a Napoli, attraverso schede digitali strutturate e arricchite, utilizzate anche per alimentare l'app di gamification urbana "Le trame degli aragonesi".

Le fonti

Materiale fornito dai responsabili della SABAP Napoli

- Corpus di 8 schede narrative dedicate ai personaggi sepolti nelle arche aragonesi della basilica di San Domenico Maggiore
- 138 immagini fotografiche a corredo (indumenti, manufatti tessili, dettagli dei corredi funebri)
- Categorie principali: schede narrative dei personaggi (contesto storico, descrizione della foggia, dei materiali e dello stato di conservazione) e schede catalografiche degli oggetti del corredo
- 49 schede catalografiche di oggetti del corredo (mantelli, camicie, berrette, calzature, brache e altri manufatti)
- Dati tecnici strutturati per ciascun oggetto (materiale, misure, tecnica, modulo decorativo, elementi accessori, note critiche)

Ingestione e Normalizzazione

- Schede personaggio: titolo, descrizione narrativa, immagini, dati extra strutturati
- Indicizzazione da manifest json delle schede (sabap_napoli_mic.json)
- Descrizioni personaggio segmentate in blocchi testuali coerenti (allineati con le traduzioni e l'audio)
- Scheda catalografica con metadati strutturati per ciascun oggetto del corredo
- Immagini normalizzate per l'inferenza (max 1000×1000 px)
- Il testo italiano originale resta la fonte originale

Generazione dei Contenuti arricchiti

Testo (Qwen2):

- Traduzione di titolo e descrizione → EN, FR, DE, ES
- Traduzione dei metadati strutturati

Immagini (Qwen2 multimodale):

- OCR con ruolo del testo (didascalia, iscrizione...)
- Riconoscimento entità visive (NER visuale)
- Descrizione visiva multilingue

Audio:

- Kokoro TTS: sintesi vocale nelle 5 lingue, per blocco testuale

Ciclo di Vita e Approvazione

Ogni contenuto generato automaticamente (una traduzione, una descrizione, un audio) passa attraverso un controllo con Governance Human-in-the-Loop prima di essere pubblicato: prima viene creato dal sistema, poi controllato, e solo dopo l'ok di una persona diventa visibile al pubblico.

- Niente viene pubblicato in automatico dopo l'elaborazione: **serve sempre l'approvazione di un operatore**
- Il controllo è puntuale: ad esempio si può pubblicare il titolo e le foto di una scheda anche se l'audio non è ancora stato revisionato

Il Dataset in Numeri

- **8 schede personaggio**
- **138 immagini**
- **5 lingue supportate** (IT + traduzioni EN/FR/DE/ES/PT)
- **49 oggetti del corredo catalogati**

Finanziato dall'Unione europea
NextGenerationEU

MINISTERO DELLA CULTURA

DIGITAL LIBRARY
Istituto centrale per la digitalizzazione del patrimonio culturale

Ecomic
ecosistema digitale per la cultura

DPaaS
Piattaforma data product as a service

Grazie

