15 ottobre 2026

Finanziato dall'Unione europea NextGenerationEU

MINISTERO DELLA CULTURA

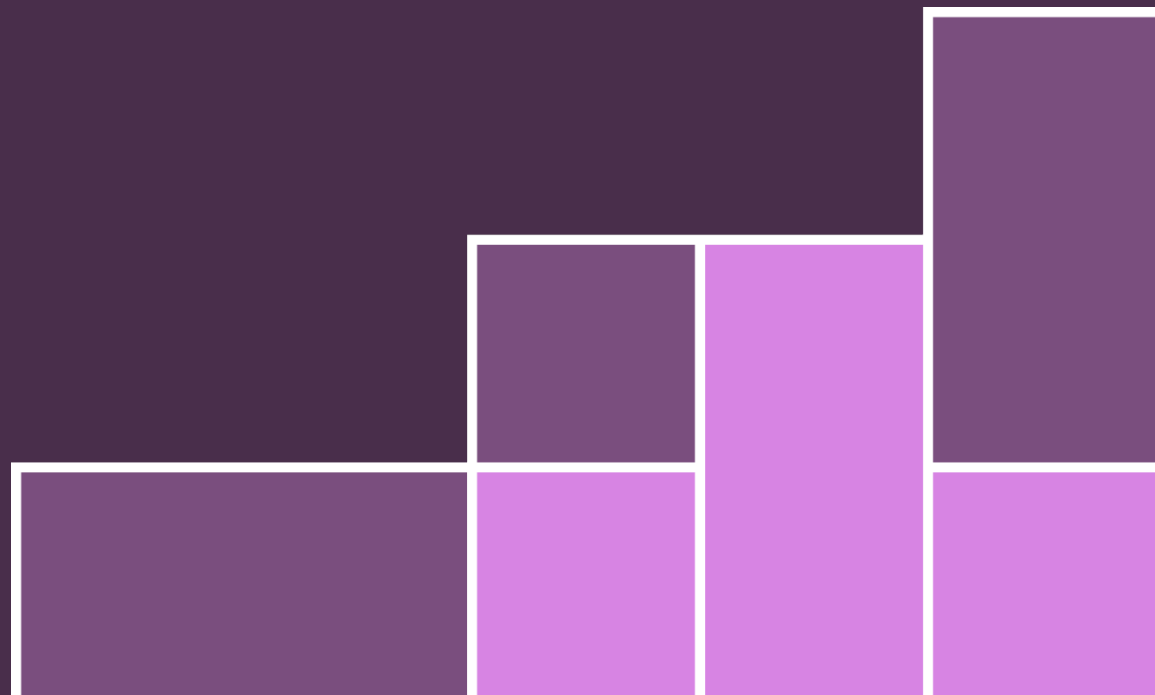
DIGITAL LIBRARY
Istituto centrale per la digitalizzazione del patrimonio culturale

Ecomic ecosistema digitale per la cultura

DPaaS Piattaforma data product as a service

Storia della scrittura e delle tecniche di stampa

Biblioteca Nazionale di Cosenza



Scopo del Data product

Rendere accessibile e fruibile un corpus di schede informative sulla storia della scrittura e delle tecniche di stampa, utilizzate anche per alimentare l'applicazione di gamification immersiva “Tra Inchiostro e Segreti: L’Enigma della Biblioteca” realizzato per la Biblioteca Nazionale di Cosenza.

Le fonti

Materiale fornito dai responsabili della Biblioteca Nazionale di Cosenza

- Corpus di 20 schede informative sulla storia della scrittura e delle tecniche di stampa
- 19 immagini a corredo delle schede (fotografie e riproduzioni)
- Ampio arco cronologico: dall'età dei Brettii (IV–III sec. a.C.) al Novecento
- Temi principali: storia della scrittura (inchiostri, pergamena e carta, scritture onciale/carolina/gotica), storia della stampa (xilografia, incunaboli, carattere mobile, torchio di Gutenberg), storia locale di Cosenza (reperti brettii, epoca romana, trasformazioni urbane, dopoguerra)
- A differenza di altri Data Product, le schede non includono un blocco di metadati strutturati: titolo, descrizione e immagini sono gli unici elementi in ingresso

Ingestione e Normalizzazione

- Schede informative: titolo, descrizione, immagini (nessun blocco di metadati strutturati aggiuntivo)
- Indicizzazione da manifest json delle schede
- Descrizioni segmentate in blocchi testuali (allineati con le traduzioni e audio dei blocchi di testo)
- 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, con protezione di codici e numeri d'inventario

Immagini (Qwen2 multimodale):

- 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: **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

- **20 schede informative**
- **19 immagini**
- **5 lingue supportate** (IT + traduzioni EN/FR/DE/ES)
- **Contenuto interamente narrativo** (nessun metadato strutturato aggiuntivo)

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

