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

I reperti del Museo "Domenico Ridola" di Matera

Museo Archeologico Nazionale "Domenico Ridola" – Matera

Data Product DP00000045

Scopo del Data product

Rendere accessibile e fruibile un corpus di schede reperto del Museo Ridola attraverso schede digitali strutturate e arricchite, utilizzate anche per alimentare l'app di gamification urbana "Ridolhunting".

Le fonti

Materiale fornito dai responsabili del Museo Archeologico Nazionale "Domenico Ridola" – Matera

- Corpus di 17 schede reperto del Museo Archeologico Nazionale "Domenico Ridola" di Matera
- 41 immagini dei reperti (fotografie ICCD)
- Cronologia: dal Neolitico Medio al III sec. a.C. (prevalenza ceramica greca/apula, VI–IV sec. a.C.)
- Categorie principali: ceramica a figure rosse/nere, ceramica geometrica e corinzia, oggetti per uso cerimoniale-magico-votivo, scultura/statuaria
- Contesti di rinvenimento: prevalentemente da Timmari (MT), Lamia S. Francesco, San Salvatore e da Montescaglioso (MT), Serra d'Alto (MT)
- Apparato di glossario: 45 note terminologiche a corredo delle schede (es. scialbatura, chitone, tenia, bacellata)

Ingestione e Normalizzazione

- Schede catalografiche: titolo, descrizione, metadati, immagini
- Indicizzazione da manifest json delle schede
- Descrizioni reperti segmentate per immagini e descrizioni immagini testuali (allineati con le traduzioni e audio dei blocchi di testo)
- Scheda tecnica con metadati
- Immagini normalizzate per l'inferenza (max 1000×1000 px)
- Il testo italiano originale resta la fonte originale

Generazione dei Contenuti arrichiti:

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: **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

- **17 schede reperto**
- **41 immagini**
- **5 lingue supportate** (IT + traduzioni EN/FR/DE/ES)
- **45 note di glossario**

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

