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

MESSINA 1908

Il patrimonio diventa accessibile

Opere, fotografie, mappe, percorsi e contenuti multimediali

Descrizioni • Audio • Sottotitoli • Easy-to-Read

OBIETTIVI

Un patrimonio inclusivo e riusabile

Pubblico eterogeneo

Particolare attenzione a esigenze visive, cognitive e a differenti livelli di alfabetizzazione.

Bilinguismo

Descrizioni, audio, sottotitoli e altri contenuti sono disponibili in italiano e inglese.

Human-in-the-loop

Il portale consente revisione, correzione e validazione degli artefatti modificabili.

Pronto per evoluzioni future

Gli artefatti generati possono alimentare nuovi servizi digitali, percorsi didattici e applicazioni di ricerca senza richiedere nuove attività di digitalizzazione.



ACCESSIBILITÀ INTEGRATA

File originali e artefatti AI sono raccolti insieme per offrire modalità di consultazione differenti.

ORIGINE DEI DATI

Un patrimonio digitale articolato



Fonte istituzionale

Artefatti e metadati conferiti dal Museo Regionale Maria Accascina.



Contenuti

Opere, fotografie, cartografia storica, dati geografici, testi, audiovisivi, tour e ricostruzioni digitali.



Formati

JSON, TIFF, GeoJSON, XLSX, JPG/PNG, DOCX/TXT, PDF, MARC21, MP3/MP4 e risorse tridimensionali.



Identificativi

Dataset DS00000108 e Data Product DP00000081, nel dominio archivistico.

VOLUMI E AGGIORNAMENTI

Conferimenti non periodici

“La dimensione del dataset sorgente è circa 20 GB.”

VOLUME INDICATIVO



Consegna di riferimento

125 fotografie, 41 opere e 70 immagini TIFF.



Evoluzione della fonte

Integrazioni e sostituzioni dipendono dalla digitalizzazione e dalle revisioni del museo.



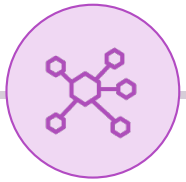
Qualità

La fonte è adeguata se sono verificate corrispondenza file-scheda, metadati, collegamenti e integrità tecnica.

ARCHITETTURA

Orchestrazione e monitoraggio

01/



S3

Area riservata per dati, immagini, GIS e materiali tematici.

02/



Apache NiFi

Job schedulati rilevano nuovi contenuti e attivano i task previsti.

03/



API Python

Traduzione dei metadati e generazione degli artefatti di accessibilità.

04/



PostgreSQL

Stati, monitoraggio, errori e retry automatici garantiscono tracciabilità e resilienza.

PIPELINE FOTOGRAFIE

Dall’immagine a quattro forme di accesso

Audiodescrizioni

Kokoro converte le descrizioni nelle corrispondenti tracce vocali.

Alt Text

Versioni sintetiche per sistemi web e tecnologie assistive.

Easy-to-Read

Testi semplificati, con frasi brevi e vocabolario immediato, in entrambe le lingue.



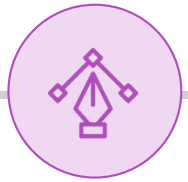
DESCRIZIONI

Analisi visuale e metadati producono descrizioni dettagliate in italiano e inglese.

PIPELINE AUDIOVISIVI

Trascrizioni e sottotitoli sincronizzati

01/



Input

File MP4 e MP3, con normalizzazione iniziale della traccia audio in WAV.

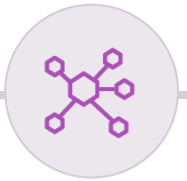
02/



Speech-to-Text

Whisper riconosce la lingua e genera la trascrizione strutturata.

03/



Sincronizzazione

Le porzioni testuali sono associate agli intervalli temporali del contenuto.

04/



Multilingua

Qwen 3.5 Text e Llama 3.1 supportano la produzione della versione inglese.

PIPELINE DOCUMENTI

Testo, traduzione e voce

Input

Documenti DOCX, PDF e TXT contenenti descrizioni e informazioni culturali.

Normalizzazione

Il testo viene estratto, ripulito e conservato come contenuto di riferimento.



Traduzione

Llama 3.1 e Qwen 3.5 Text producono la versione inglese.

Audio

Kokoro genera audiodescrizioni in italiano e inglese, pubblicate con il documento originale.

TECNOLOGIE

DPaaS, CINECA e modelli locali



Modelli esterni

Alcuni servizi sono stati eseguiti su infrastrutture esterne per stabilità e qualità dei risultati.



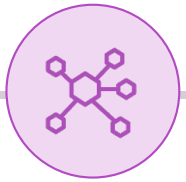
Vision AI

Qwen 3.5 Vision è stato il riferimento durante sviluppo e collaudo per i contenuti multimodali.

ESPOSIZIONE

Portale, catalogo e API read-only

01/



Portale web

Canale principale per visitatori e operatori autorizzati.

02/



Catalogo DPaaS

Scoperta del Data Product, informazioni descrittive e richiesta di accesso.

03/



API REST

Ricerca, dettaglio, relazioni, classificazioni, filtri e paginazione con risposte JSON.

04/



Sicurezza

Le credenziali interne e lo storage restano confinati nei servizi backend.

POSSIBILI EVOLUZIONI

Un’architettura modulare

“Aggiornamento o sostituzione dei modelli AI senza modifiche significative ai processi.”

MODELLI EVOLUTIVI



Pipeline come CPA

Servizi condivisi e riutilizzabili per altri enti e Data Product.



Aggiornamento dinamico

Nuovi file potrebbero attivare automaticamente elaborazione e pubblicazione.



Vincolo attuale

La gestione completamente automatica dell’evoluzione dei contenuti non è implementata nella versione corrente.