31 luglio 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

ESTRAZIONE E ARRICCHIMENTO METADATI DEI CONTENUTI DIGITALI DEI BRONZI DI RIACE

Data Product dedicato all'analisi, all'estrazione e all'arricchimento semantico dei contenuti digitali relativi ai Bronzi di Riace.

Cinquant’anni di studi, analisi e interventi conservativi sui Bronzi di Riace, raccolti e connessi in un’unica base di conoscenza.

Dagli asset digitali alle relazioni tra opere, studi e restauri

Nel 1972 i Bronzi di Riace riemergono dal mare della Calabria.

Da allora sono stati studiati, analizzati, restaurati e raccontati come pochi altri reperti al mondo.

Ogni intervento ha prodotto documentazione: relazioni di restauro, campagne fotografiche, registrazioni, analisi dei materiali e oggi modelli tridimensionali. Migliaia di file, in decine di formati diversi, che il Data Product mette in relazione tra loro.

In che modo il Data Product genera valore?

Il Data Product analizza i file, ne estrae le informazioni tecniche e identifica persone, enti, luoghi, materiali, attività, pubblicazioni e interventi conservativi.

I dati vengono organizzati in un'unica base di conoscenza, nella quale ogni contenuto è collegato alle opere e agli altri elementi pertinenti.

Diventa così possibile ricostruire la storia degli studi e della conservazione dei Bronzi, cercare informazioni trasversalmente e alimentare nuovi servizi digitali.

Ogni contenuto viene descritto nella forma e nel significato

Livello 1 Metadati tecnici

Il primo livello descrive le **caratteristiche oggettive del file**, senza interpretarne il contenuto. Il sistema riconosce automaticamente:

- nome, formato e dimensione;
- autore o creatore dichiarato;
- date di creazione e modifica;
- numero di pagine per i documenti;
- risoluzione e dimensioni per le immagini;
- durata e codec per audio e video;
- vertici, superfici e dimensioni per i modelli 3D.

Queste informazioni permettono di conoscere **consistenza, qualità e composizione tecnica dell'archivio digitale**.

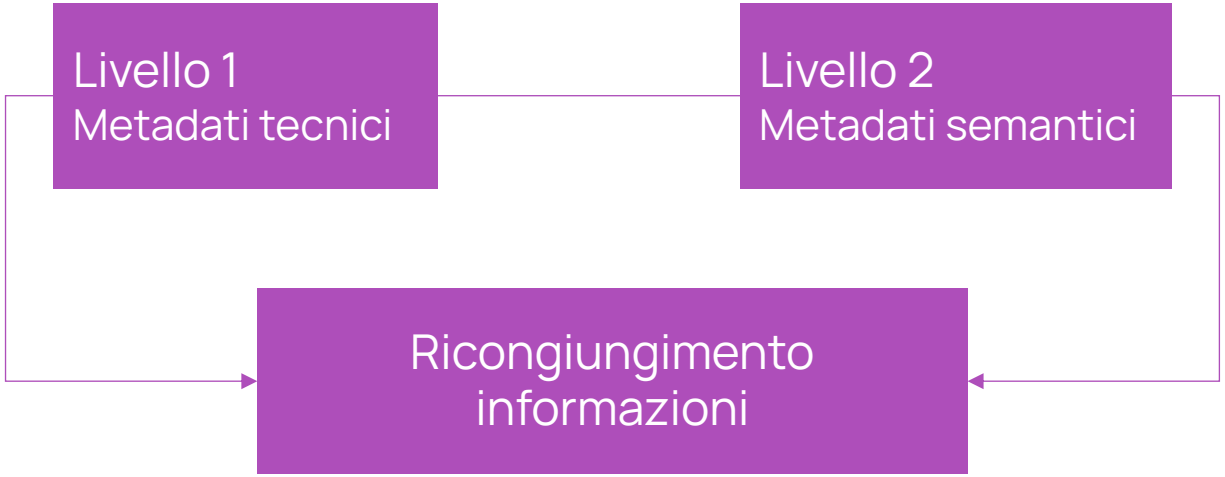
Livello 2 Metadati semantici

Il secondo livello analizza **ciò che il contenuto racconta**. Attraverso servizi di intelligenza artificiale, il sistema estrae o genera:

- titolo e sintesi del contenuto;
- persone e organizzazioni citate;
- luoghi e periodi temporali;
- materiali e tecniche;
- attività di studio, restauro e conservazione;
- soggetti tematici;
- pubblicazioni e riferimenti bibliografici.

Le informazioni vengono ricondotte a vocabolari e ontologie condivise, come Dublin Core e ArCo, per renderle **coerenti e interoperabili.**

Relazioni – La conoscenza che collega i contenuti



I due livelli vengono riuniti nel profilo completo dell'asset. Il Data Product non si limita però a descrivere i singoli file: riconosce e struttura anche le relazioni tra le informazioni. In questo modo l'archivio non è più una semplice raccolta di file, ma diventa una rete di conoscenza navigabile sui Bronzi di Riace.

Può collegare, ad esempio:

- un restauratore all'intervento eseguito;
- l'intervento al Bronzo interessato;
- un materiale alla tecnica di conservazione;
- uno studio al relativo autore e all'opera analizzata;
- un documento agli enti, ai luoghi e ai periodi citati.

Come funziona, passo per passo

01

Ingestione

Il file viene prelevato dal repository e identificato con un'impronta univoca del contenuto.

02

Estrazione

Se ne misurano le proprietà fisiche, diverse per ogni famiglia di formato.

03

Comprensione

Testo, immagini e audio vengono letti: l'intelligenza artificiale ne ricava entità e argomenti.

04

Sintesi

Un modello linguistico genera un riassunto e struttura gli eventi descritti nel documento.

05

Allineamento

Ogni concetto viene ricondotto ai vocabolari ufficiali dei beni culturali.

06

Pubblicazione

Il profilo completo entra nel catalogo, pronto per ricerche, analisi ed export.

Dalla fonte alla risposta contestualizzata

1

INPUT

01 · Estrazione

Il sistema acquisisce il file e ne rileva metadati tecnici e impronta univoca

2

ANALISI

02 · Arricchimento

AI e servizi specialistici riconoscono testo, immagini, audio, entità ed eventi

3

OUTPUT

03 · Connessione

Ontologie e database organizzano relazioni, statistiche, API ed export.

Componenti specializzati, un unico profilo del patrimonio

Il Data Product integra diversi componenti, ciascuno dedicato a una fase dell'elaborazione.

Documenti, immagini, audio, video e modelli 3D vengono trattati con strumenti specifici, ma confluiscono in un unico modello informativo dedicato ai Bronzi di Riace.



API e storage – Acquisire e rendere accessibili i contenuti

Gli asset digitali vengono acquisiti dallo storage S3 attraverso un'API REST. Il sistema assegna a ogni elaborazione un identificativo, controlla lo stato delle attività e rende disponibili le funzioni di consultazione, analisi ed esportazione.

Durante l'acquisizione viene calcolata l'impronta SHA-256 del file, che permette di riconoscere eventuali duplicati.



Ontologie e database – Organizzare e collegare dati

Le informazioni estratte vengono ricondotte a concetti condivisi.

Metadati tecnici, informazioni semantiche e relazioni vengono salvati in PostgreSQL.

Il risultato è un **unico profilo del patrimonio**, accessibile tramite API ed esportabile in formato CSV o JSON.



Estrazione e AI – Leggere formati e contenuti

Apache Tika estrae testo e metadati tecnici da documenti e contenuti multimediali. Componenti dedicati analizzano immagini, audio, video e modelli 3D, rilevando informazioni come risoluzione, durata, codec, vertici, superfici e dimensioni geometriche.

I servizi di intelligenza artificiale completano il profilo dell'asset:

- OCR per riconoscere il testo nelle immagini;
- Whisper per trascrivere audio e video;
- Kosmos-2 per descrivere i contenuti visivi;
- NERS per identificare persone, enti, luoghi e date;
- Tagging per classificare i temi secondo il Nuovo Soggettario;
- Llama 3.1 per produrre abstract e strutturare interventi conservativi e studi.

Quattro scenari applicativi

01/ Ricerca, conservazione, valorizzazione e divulgazione

Interrogazione trasversale della documentazione, consultazione, e correlazione tra dati

02/ Analizzare studi, materiali e tecniche

Esplorare pubblicazioni e relazioni scientifiche per riconoscere materiali, tecniche diagnostiche e restauri

03/ Consultare il profilo completo di un asset

Accedere, in un unico punto, a caratteristiche tecniche e contenuto semantico di documenti, immagini, audio, video e 3D

04/ Ricerca filtrata

Limitare i risultati per fonte o tipologia documentale.

DPaaS Piattaforma data product as a service

Riusabilità

Un modello per collezioni complesse

Il Data Product può essere applicato a collezioni per le quali opere, studi, documentazione tecnica e attività conservative devono essere ricondotti a un'unica base di conoscenza.

Condizioni per il riuso: individuare il corpus da elaborare, verificare formati e qualità delle fonti e definire le entità e le relazioni più significative per il proprio patrimonio. I risultati prodotti dall'intelligenza artificiale possono essere sottoposti a validazione curatoriale prima della pubblicazione o del riuso.

Connettere archivi e collezioni

Musei, archivi e istituti culturali possono utilizzare la soluzione per riunire documenti, immagini, contenuti audiovisivi, dati strutturati e modelli 3D.

Ogni asset viene descritto attraverso metadati tecnici e semantici, permettendo di superare la frammentazione tra formati, repository e sistemi documentali differenti.

Supportare ricerca e conservazione

Il modello può essere configurato per riconoscere le entità rilevanti per il patrimonio dell'ente: opere, persone, organizzazioni, luoghi, materiali, tecniche, attività, periodi, interventi e pubblicazioni.

Le relazioni estratte consentono di ricostruire la storia degli studi e della conservazione, individuare i documenti associati a un'opera e confrontare informazioni provenienti da fonti diverse.

Alimentare nuovi servizi digitali

Profili, metadati e relazioni sono disponibili tramite API REST ed esportazioni CSV e JSON. Possono quindi alimentare:

- motori di ricerca semantica;
- strumenti di analisi e dashboard;
- archivi digitali evoluti;
- chatbot basati sulle fonti;
- applicazioni museali e percorsi di visita;
- visualizzazioni delle relazioni;
- esperienze immersive e contenuti interattivi.

Quando un archivio diventa una rete di metadati, fonti e relazioni, il patrimonio si trasforma in conoscenza interrogabile