31 luglio 2026

Finanziato dall'Unione europea NextGenerationEU

MiC 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

VALORIZZAZIONE DIGITALE DELLA VIA APPIA ATTRAVERSO L'ESTRAZIONE E L'ARRICCHIMENTO SEMANTICO DEI METADATI

Data Product realizzato nell'ambito del progetto Appia Time Machine per l'estrazione automatica di metadati, entità semantiche e relazioni tra beni culturali a partire da contenuti digitali eterogenei

Duemila anni di storia, paesaggi e testimonianze della Via Appia, connessi in un'unica infrastruttura di conoscenza digitale.

Fonti diverse costruiscono una conoscenza unitaria della Via Appia

La Via Appia è documentata da un patrimonio digitale ampio ed eterogeneo: fotografie storiche, acquerelli, incisioni, documenti e pubblicazioni, registrazioni audio e video, rilievi e modelli 3D.

Questi contenuti descrivono luoghi, monumenti, paesaggi, materiali, trasformazioni e attività di conservazione. La varietà delle fonti e dei formati rende però difficile consultarli insieme e ricostruire le relazioni tra i diversi tratti della Via, i beni culturali e gli studi che li riguardano.

In che modo il Data Product genera valore?

Il Data Product acquisisce e analizza gli asset digitali, estraendo metadati tecnici e informazioni culturali.

Persone, luoghi, materiali, eventi, interventi conservativi e riferimenti bibliografici vengono normalizzati e organizzati secondo standard condivisi.

I contenuti possono così essere collegati tra loro anche quando provengono da formati o repository differenti.

Il risultato è una base di conoscenza interoperabile e interrogabile, pronta per alimentare catalogazione, ricerca, mappe, percorsi narrativi, visualizzazioni ed esperienze immersive dedicate alla Via Appia.

Per ogni asset, un profilo che ne descrive caratteristiche tecniche e significato culturale

Livello 1
Identità tecnica

Il sistema estrae automaticamente le proprietà oggettive del file:

- formato, dimensione, autore e date;
- pagine e testo per i documenti;
- risoluzione e caratteristiche delle 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 dell'archivio digitale.

Livello 2
Profilo culturale

L'intelligenza artificiale analizza ciò che l'asset rappresenta o racconta, individuando:

- persone, enti e organizzazioni;
- luoghi, monumenti e tratti della Via;
- materiali e tecniche;
- attività, eventi e periodi storici;
- interventi di conservazione;
- studi e riferimenti bibliografici;
- titolo, abstract e soggetti tematici.

Relazioni – La conoscenza che collega i contenuti



I metadati vengono organizzati secondo Dublin Core, ArCo e CIDOC-CRM e serializzati in JSON-LD.
Il sistema può così collegare un luogo ai beni presenti, un intervento all'opera interessata, uno studioso alla relativa pubblicazione o un modello 3D al monumento rappresentato.

La pipeline traduce l’eterogeneità in conoscenza condivisa

1

INPUT

01 · Acquisizione

Il sistema importa documenti, immagini, audio, video e modelli 3D, verifica eventuali duplicati ed estrae i metadati tecnici.

2

ANALISI

02 · Interpretazione

OCR e servizi di intelligenza artificiale riconoscono testi, persone, luoghi, materiali, eventi, interventi conservativi e studi.

3

OUTPUT

03 · Disponibilità

Le informazioni vengono normalizzate secondo Dublin Core, ArCo e CIDOC-CRM, archiviate in PostgreSQL e rese disponibili tramite API ed export CSV/JSON.

Un’architettura dati per analisi, mappe ed esperienze immersive

Il Data Product è un microservizio cloud-native che integra componenti specializzati per trattare documenti, immagini, contenuti audiovisivi, dati strutturati e modelli 3D all’interno di un unico processo.



Acquisizione e accesso

Gli asset vengono acquisiti dallo storage cloud e gestiti attraverso API REST. Il sistema calcola l’impronta SHA-256 per evitare duplicazioni e assegna un identificativo univoco a ogni contenuto.

Le API permettono di avviare e monitorare l’elaborazione, consultare i profili degli asset, interrogare le viste analitiche ed esportare i dati.



Normalizzazione semantica

Le entità estratte vengono associate a concetti ontologici condivisi. Il sistema uniforma termini provenienti da fonti differenti e costruisce relazioni tra beni, luoghi, persone, materiali, interventi e pubblicazioni.

Le informazioni sono organizzate in un grafo semantico serializzato in JSON-LD, interoperabile con altri sistemi culturali.



Estrazione tecnica e analisi AI

Apache Tika estrae testo e metadati tecnici da documenti, immagini, audio e video. Componenti dedicati analizzano i modelli 3D, rilevando vertici, superfici e dimensioni geometriche.

I servizi AI completano il profilo attraverso:

- OCR delle immagini;
- trascrizione di audio e video;
- descrizione dei contenuti visivi;
- riconoscimento di persone, luoghi e date;
- estrazione di materiali, attività e soggetti;
- generazione di abstract;
- identificazione studi bibliografici.



Persistenza, analisi e riuso

Metadati tecnici, informazioni semantiche e relazioni vengono archiviati in PostgreSQL. Le viste rendono disponibili il profilo completo degli asset, le entità frequenti, le relazioni, gli autori e le statistiche per formato e periodo.

API REST ed export CSV/JSON permettono di utilizzare i dati per analisi, mappe digitali, percorsi narrativi, visualizzazioni avanzate ed esperienze immersive dedicate alla Via Appia.

Quattro scenari applicativi

01/ Catalogazione e analisi del patrimonio

Organizzare automaticamente documenti, immagini, contenuti audiovisivi e modelli 3D, consultando statistiche, autori, formati ed entità ricorrenti.

02/ Ricerca e conservazione

Collegare beni, luoghi, materiali, interventi conservativi, studiosi e pubblicazioni, ricostruendo il profilo completo di monumenti e tratti della Via.

03/ Mappe e percorsi narrativi

Utilizzare luoghi, eventi, periodi e relazioni semantiche per creare cartografie digitali, itinerari tematici e visualizzazioni del patrimonio diffuso.

04/ Esperienze immersive

Alimentare applicazioni VR e servizi interattivi con contenuti e dati strutturati, mantenendo il collegamento con gli asset e le fonti originali.

Riusabilità

Il Data Product può essere adottato da parchi archeologici, musei, archivi, istituti culturali e reti territoriali che gestiscono fonti eterogenee riferite a beni, luoghi e percorsi.

Condizioni per il riuso: l'ente deve selezionare il corpus, verificare qualità e formati degli asset e definire le entità e le relazioni più significative per il proprio patrimonio.

Unificare fonti e formati differenti

Documenti, immagini, audio, video, dati strutturati e modelli 3D provenienti da repository diversi confluiscono in un'unica base di conoscenza.
Ogni asset viene descritto attraverso metadati tecnici e culturali, superando la frammentazione tra formati e sistemi di catalogazione.

Collegare beni, luoghi e conoscenze

Il modello può riconoscere e mettere in relazione beni, monumenti, luoghi, persone, materiali, eventi, interventi conservativi e pubblicazioni.
Ogni contenuto viene così ricondotto al proprio contesto territoriale, storico e scientifico.

Supportare ricerca e interoperabilità

Curatori e ricercatori possono consultare profili, relazioni e statistiche sul corpus. Dublin Core, ArCo, CIDOC-CRM e JSON-LD facilitano lo scambio con altri sistemi.
API REST ed export CSV/JSON rendono i dati disponibili per analisi e applicazioni esterne.

Creare nuove esperienze

La base di conoscenza può alimentare mappe digitali, itinerari culturali, portali, visualizzazioni e applicazioni immersive.

La conoscenza della Via Appia prende forma quando dati, luoghi e testimonianze diventano parte di un sistema connesso.