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

Libreria integrata di rilievi 3D, fotogrammetria, analisi stratigrafica e metadati

Il Data Product: obiettivi, dati e fruizione

Il Data Product

Nell'ambito del Lotto 4 del Dialogo Competitivo (PNRR – Missione 1, Componente 3), il Data Product "Castello di San Martino - Libreria integrata di rilievi 3D, fotogrammetria, analisi stratigrafica e metadati" è stato sviluppato per il Museo archeologico del Fiume Bacchiglione (MA Bacchiglione) con l'obiettivo di supportare le attività di documentazione, analisi e restauro del bene e, allo stesso tempo, favorirne la valorizzazione e la fruizione culturale.

Il Data Product integra rilievi laser scanner, rilievi fotogrammetrici e documentazione di analisi stratigrafica in un dataset strutturato e interoperabile, arricchito attraverso processi di elaborazione e Intelligenza Artificiale.



Obiettivi e valore



Documentazione 3D

Creazione di una rappresentazione tridimensionale accurata e strutturata del bene culturale.



Integrazione dei dati

Riunione di rilievi laser scanner, dati fotogrammetrici e documentazione stratigrafica in un dataset integrato.



Supporto al restauro

Utilizzo di modelli e informazioni integrate per supportare analisi, confronti e simulazioni degli interventi.



Arricchimento semantico

Strutturazione e organizzazione di metadati per arricchire le informazioni associate al bene culturale.



Contenuti multi-target

Produzione di descrizioni differenziate per esperti e pubblico generico, con il supporto dell'Intelligenza Artificiale.



Valorizzazione e fruizione

Accesso a contenuti 3D, testuali e audio per favorire la conoscenza e la fruizione del patrimonio culturale.

I dati alla base del Data Product

Il Data Product integra diverse tipologie di dati relativi al bene culturale, combinando informazioni tridimensionali, documentazione e contenuti multimediali. I materiali sorgente vengono acquisiti nello storage S3 e resi disponibili alle successive fasi di elaborazione.

Principali tipologie di dati sorgente

- **Nuvole di punti**
  Rilievi laser scanner nei formati **LAS/LAZ**.
- **Modelli 3D**
  Rilievi e modelli fotogrammetrici nei formati **GLB/GLTF**.
- **Documentazione stratigrafica**
  Documentazione relativa all'analisi del bene nei formati **PDF/TXT**.
- **Immagini**
  Materiali visivi nei formati **TIFF/JPEG/PNG**.
- **Video**
  Contenuti audiovisivi utilizzati tra i materiali sorgente del Data Product.

* Dati e metriche da Ministero della Cultura 2024

A chi si rivolge il Data Product?

I principali profili utente

01

Data Producer

Alimentazione del Data Product attraverso il caricamento dei dati sorgente e l'avvio delle elaborazioni.

02

Ricercatori

Consultazione e utilizzo dei modelli 3D e dei dati associati per attività di studio e analisi.

03

Restauratori e operatori

Utilizzo dei dati e dei modelli a supporto delle attività di documentazione, monitoraggio e restauro.

04

Data Consumer

Consultazione, riutilizzo e integrazione di dataset 3D e metadati nei propri flussi di lavoro.

05

Visitatori

Fruizione dei contenuti derivati, tra cui scene 3D, descrizioni testuali e contenuti audio.

06

Product Manager e Admin

Supervisione delle risorse, monitoraggio dei processi e gestione degli accessi.

02/

Dal dato sorgente al dataset elaborato

La pipeline di elaborazione

01/

Acquisizione

02/

Elaborazione

03/

Arricchimento

04/

Pubblicazione

Le fasi della pipeline

La pipeline di elaborazione accompagna i dati lungo l'intero processo di trasformazione, a partire dall'acquisizione dei materiali sorgente fino alla produzione e alla pubblicazione del dataset elaborato. Il processo è organizzato in quattro fasi principali, ciascuna dedicata a uno specifico momento del trattamento e dell'arricchimento dei dati.

01

Acquisizione

I dati sorgente vengono caricati nello **storage S3**, che costituisce il punto di ingresso dei materiali nella pipeline.

02

Elaborazione

I contenuti tridimensionali vengono sottoposti a operazioni di conversione e ottimizzazione, mentre documenti e immagini vengono elaborati attraverso **OCR, Kosmos2 e Image-to-Image**.

03

Arricchimento

Le informazioni ottenute vengono elaborate attraverso **LLAMA31** per la produzione di descrizioni multi-target e tramite **Text-to-Speech** per la generazione delle relative audio guide.

04

Pubblicazione

Gli output confluiscono nel **dataset elaborato**, memorizzato nello storage S3 e reso accessibile tramite **download** e **API REST del WSDP**.

Tecnologie e componenti

L'architettura del Data Product integra componenti dedicati allo sviluppo applicativo, alla gestione e persistenza dei dati e alla loro elaborazione. Le diverse tecnologie supportano l'intero processo, dall'acquisizione dei materiali sorgente fino alla gestione e all'esposizione del dataset elaborato.

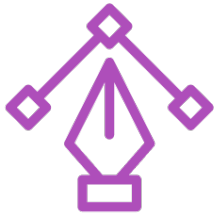
* Dati e metriche da Ministero della Cultura 2024

Componenti principali

- **Python**
  Linguaggio per lo sviluppo delle componenti applicative e della pipeline di elaborazione.
- **FastAPI / WSDP**
  Framework per l'esposizione delle API REST e l'accesso programmatico alle risorse del Data Product.
- **S3**
  Storage a oggetti per la gestione dei dati sorgente e degli output del dataset elaborato.
- **PostgreSQL**
  Database relazionale per la gestione di metadati, log e informazioni strutturate.
- **Ottimizzazione multi-risoluzione**
  Processo applicato a modelli 3D e nuvole di punti per predisporli alla successiva fruizione.
- **Servizi CPA**
  Servizi DPaaS dedicati alle elaborazioni AI e all'arricchimento dei diversi contenuti.

L'Intelligenza Artificiale nella pipeline

L’Intelligenza Artificiale interviene in diverse fasi della pipeline di elaborazione, supportando l’estrazione, l’analisi e l’arricchimento delle informazioni associate al bene culturale. Le principali elaborazioni vengono realizzate attraverso i servizi CPA (Content Processing Advanced) disponibili sulla piattaforma DPaaS, ciascuno specializzato nel trattamento di specifiche tipologie di contenuto. I servizi consentono di operare su immagini e documenti, estrarre e analizzare informazioni, generare descrizioni rivolte a differenti tipologie di pubblico e produrre contenuti audio. A questi servizi si affianca Trellis, utilizzato come tecnologia esterna a DPaaS per la generazione di modelli tridimensionali a partire dall’immagine dell’oggetto rappresentato.



Servizi e modelli AI

Kosmos2
Analisi semantica delle immagini ed estrazione di informazioni testuali.

IMMAGINI
SERVIZIO CPA

OCR
Estrazione dei contenuti testuali dalla documentazione stratigrafica.

TESTO
SERVIZIO CPA

LLAMA31
Elaborazione dei testi e generazione di descrizioni divulgative multi-target.

TESTO
SERVIZIO CPA

Text-to-Speech
Generazione di audio guide a partire dalle descrizioni testuali.

AUDIO
SERVIZIO CPA

Image-to-Image
Elaborazione delle immagini di rilievo e della documentazione.

IMMAGINI
SERVIZIO CPA

Trellis
Generazione di modelli 3D a partire dall'immagine dell'oggetto rappresentato.

MODELLI 3D
SERVIZIO ESTERNO

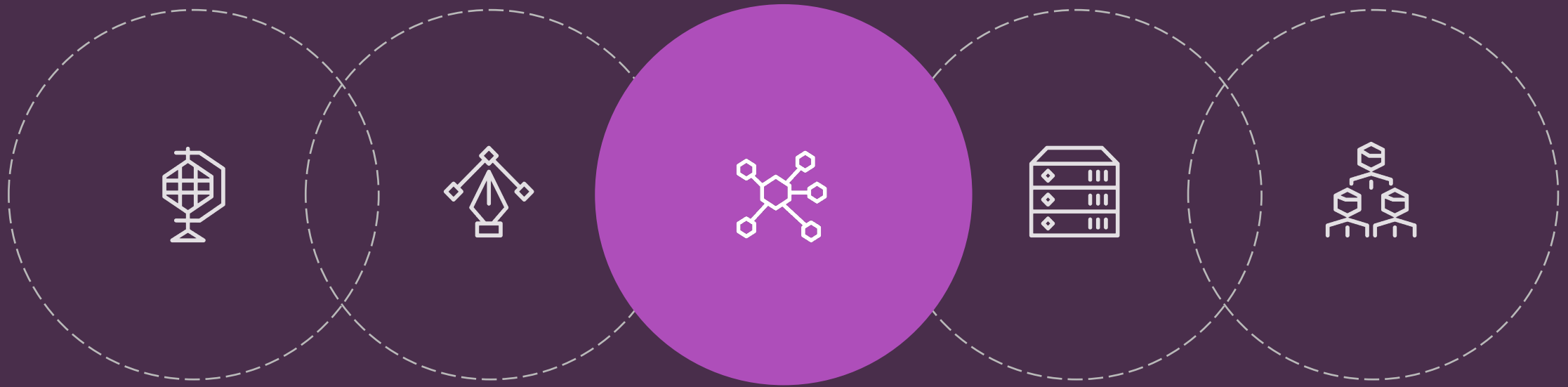
03/

I contenuti del Data Product

Quali contenuti produce il Data Product?

Gli output del Data Product

Al termine delle diverse fasi di elaborazione, il Data Product rende disponibili contenuti tridimensionali, informazioni semantiche e risorse multimediali, organizzati all'interno del dataset elaborato e destinati alle diverse modalità di fruizione. Gli output sono disponibili in formati differenti in funzione della tipologia di contenuto: OBJ/GLTF per i modelli 3D, LAS/LAZ per le nuvole di punti, JSON e TXT/JSON per metadati e contenuti testuali e MP3/WAV per i contenuti audio.



Modelli e dati 3D

Metadati sematici

Dataset Elaborato

Contenuti testuali

Contenuti audio

L’organizzazione del dataset

Le risorse del Data Product sono organizzate attraverso specifiche entità, che ne definiscono la struttura informativa.

Queste consentono di organizzare e mettere in relazione i diversi contenuti che compongono il dataset.

Entità principali

- **Target**
  Pubblico a cui sono destinati i contenuti, distinto tra generico e specialista.
- **Contenuti**
  Descrizioni testuali associate a punti di interesse, luoghi e reperti.
- **Audio e Immagini**
  Risorse multimediali corredate dai relativi metadati descrittivi.
- **Modelli 3D**
  Risorse tridimensionali corredate da formato, riferimento spaziale, LOD e bounding box
- **Luoghi**
  Riferimenti spaziali organizzati gerarchicamente e associati alle diverse risorse.
- **Reperti storici**
  Elementi associabili a luoghi, immagini, modelli 3D e contenuti.

04/

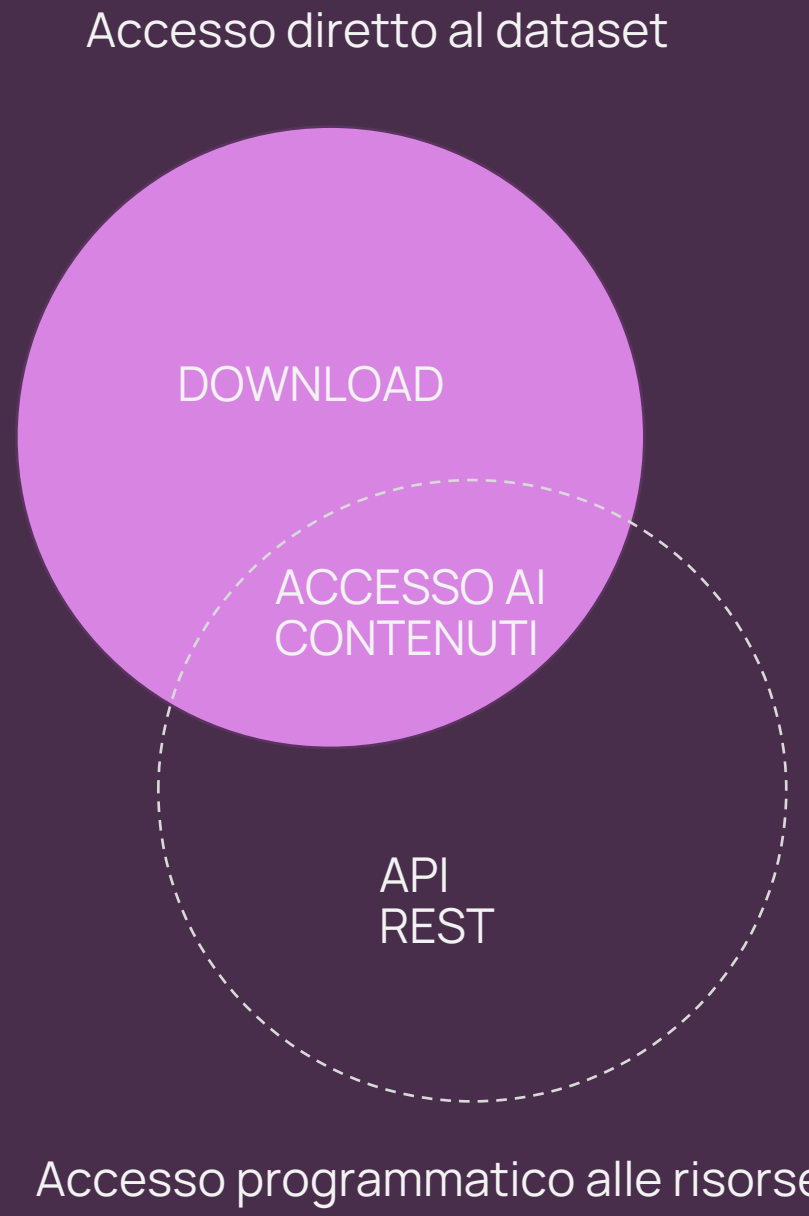
Fruizione del Data Product

Come si accede al Data Product?

Le modalità di fruizione

Il Data Product rende disponibili dati e contenuti attraverso due modalità di accesso:

- il download del dataset, che consente di consultare e riutilizzare direttamente i dati e i contenuti elaborati;
- l'accesso tramite API REST, che permette di interrogare le risorse del Data Product attraverso il Web Service del Data Product (WSDP) e di integrarle in applicazioni e sistemi esterni.



Accesso tramite API REST

L'accesso programmatico al Data Product avviene attraverso le API REST esposte dal Web Service del Data Product (WSDP), che consentono di interrogare in modo selettivo le risorse del dataset e recuperare le informazioni di interesse senza dover acquisire l'intero contenuto.

I dati vengono scambiati in formato JSON e possono essere utilizzati per integrare le risorse del Data Product all'interno di applicazioni e sistemi esterni. L'accesso ai servizi è gestito attraverso un meccanismo di autenticazione basato su API key.



Limitazioni e condizioni di utilizzo

La fruizione dei contenuti tridimensionali è condizionata principalmente dalla dimensione e dalla complessità dei modelli 3D e delle nuvole di punti, che possono incidere sui tempi di elaborazione, caricamento e visualizzazione.

Per migliorare le prestazioni, i contenuti possono essere sottoposti a ottimizzazione multi-risoluzione e gestiti attraverso livelli di dettaglio (LOD) e streaming progressivo tramite CESIUM.



Dimensione dei contenuti

Modelli complessi possono richiedere maggiori risorse di elaborazione.



Prestazioni

La complessità dei modelli può incidere sulla fluidità della fruizione.



Multi-risoluzione

L'ottimizzazione consente una gestione più efficiente dei contenuti complessi.



Tempi di caricamento

Il peso dei contenuti può rallentarne il trasferimento e la visualizzazione.



Conversione 3D

Dimensioni elevate possono rendere più onerose le operazioni di conversione.



LOD e streaming

LOD e streaming progressivo tramite CESIUM favoriscono una visualizzazione più efficiente.

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

