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

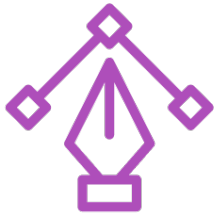
Condition Report: dataset integrato a supporto del prestito

Il Data Product: obiettivi, dati e fruizione

Il Data Product

Nell'ambito del Lotto 4 del Dialogo Competitivo (PNRR – Missione 1, Componente 3, Investimento 1.1), il Data Product "Condition Report" (DP00000117) è stato sviluppato per il Museo delle Civiltà di Roma con l'obiettivo di supportare le attività di documentazione e di certificazione dello stato conservativo dei beni culturali e, allo stesso tempo, di valorizzare il patrimonio informativo già custodito dall'istituto.

Il Data Product integra i dati catalografici con le informazioni sul restauro e sulle analisi diagnostiche in un dataset unificato e strutturato, arricchito mediante tecniche di intelligenza artificiale per l'estrazione di metadati semantici, a supporto della realizzazione del condition report per il prestito dei beni museali.



Obiettivi e valore



Efficienza dei flussi

Semplifica la produzione della reportistica relativa ai prestiti e favorisce il tracciamento dello stato conservativo del bene.



Metadati semantici

L'arricchimento anche tramite intelligenza artificiale estrae metadati che rendono il patrimonio ricercabile e riutilizzabile.



Supporto decisionale

Offre ad archivisti, conservatori e restauratori uno strumento per valutare le azioni da intraprendere per la protezione del bene.



Interoperabilità

La standardizzazione dei dati ne facilita il riutilizzo in contesti applicativi diversi, riducendo le duplicazioni.



Dati integrati

Unifica dati catalografici, documentazione archivistica e conservativa e analisi diagnostiche in un dataset coerente.



Valorizzazione

I contenuti generati rendono il patrimonio più accessibile e comprensibile anche a un pubblico non specialista.

I dati alla base del Data Product

Il Data Product integra la documentazione già custodita dal museo e i dati gestiti nella banca dati dell'istituto.

I materiali sorgente vengono caricati sullo storage S3 e resi disponibili alle successive fasi di elaborazione.

Il dataset di partenza è costituito da:

- circa 250 schede oggetto
- circa 20 indagini diagnostiche
- 10 Mostre
- 2400 elementi di archivio fotografico
- 50 GB di archivio documentale + immagini

Principali tipologie di dati sorgente

- Dati catalografici
  Schede e metadati dei beni in formato **strutturato**.
- Documentazione archivistica
  Materiali d'archivio e curatoriali relativi al bene nei formati **PDF/TXT**.
- Documentazione conservativa
  Relazioni di restauro ed esiti delle analisi diagnostiche nei formati **PDF/TXT**.
- Condition report storici
  Report già associati ai beni nella **banca dati** del museo.
- Immagini
  Materiali visivi nei formati TIFF, JPEG e PNG.

A chi si rivolge il Data Product?

I principali profili utente

01 Data Producer

Inserimento della documentazione, configurazione iniziale e monitoraggio della pipeline.

02 Ricercatori e tecnici

Utenti esperti di dominio che utilizzano la piattaforma di documentazione e gestione del bene museale.

03 Restauratori

Abilitati alla creazione dei condition report tramite l'applicazione di gestione.

04 Data Consumer

Professionisti ed enti terzi autorizzati che consultano e riutilizzano dataset e metadati dal catalogo DPaaS.

05 Visitatori

Utenti finali con accesso limitato all'interfaccia di fruizione dei contenuti.

06 Admin

Supervisione delle risorse computazionali e gestione degli accessi.

02/

Dal dato sorgente al dataset elaborato

La pipeline di elaborazione

01/

Import

02/

Ingestion

03/

Generazione

04/

Pubblicazione

Le fasi della pipeline

La pipeline accompagna i dati dall'acquisizione dei materiali sorgente fino alla pubblicazione del dataset elaborato.

Il processo è organizzato in quattro fasi, con esecuzione ad avvio manuale da parte di un operatore.

01

Import

I dati sorgente vengono caricati sullo **storage S3**, punto di ingresso dei materiali nella pipeline. Oppure importati a partire da un DP precedentemente pubblicato

02

Ingestion

Il dato sorgente viene integrato all'interno della piattaforma di gestione dei beni museali.

03

Generazione

Il dato elaborato con l'ausilio dei servizi CPA produce dati e metadati pronti per essere pubblicati all'interno del DPLab.

04

Pubblicazione

Gli output confluiscono nel **dataset elaborato** su S3, esposto tramite l'**API REST di GolangAPI**; i metadati sono persistiti su **PostgreSQL.**

Tecnologie e componenti

L'architettura del Data Product integra componenti dedicati allo sviluppo applicativo, alla gestione e persistenza dei dati e alla loro esposizione.

L'elaborazione AI non è eseguita dal Data Product, ma delegata ai servizi CPA a catalogo, esterni al suo perimetro.

Componenti principali

- **GolangAPI**
  Web service che espone l'API REST del DP e gestisce le entità del dataset.
- **PostgreSQL**
  Database relazionale per i metadati e i log di elaborazione.
- **Amazon S3**
  Object storage per i dati sorgente e per il dataset elaborato.
- **Servizi CPA**
  Servizi a catalogo DPaaS che eseguono le elaborazioni AI, esterni al perimetro del DP.
- **DP Lab**
  Contenitore VM per lo sviluppo full-code del Data Product.

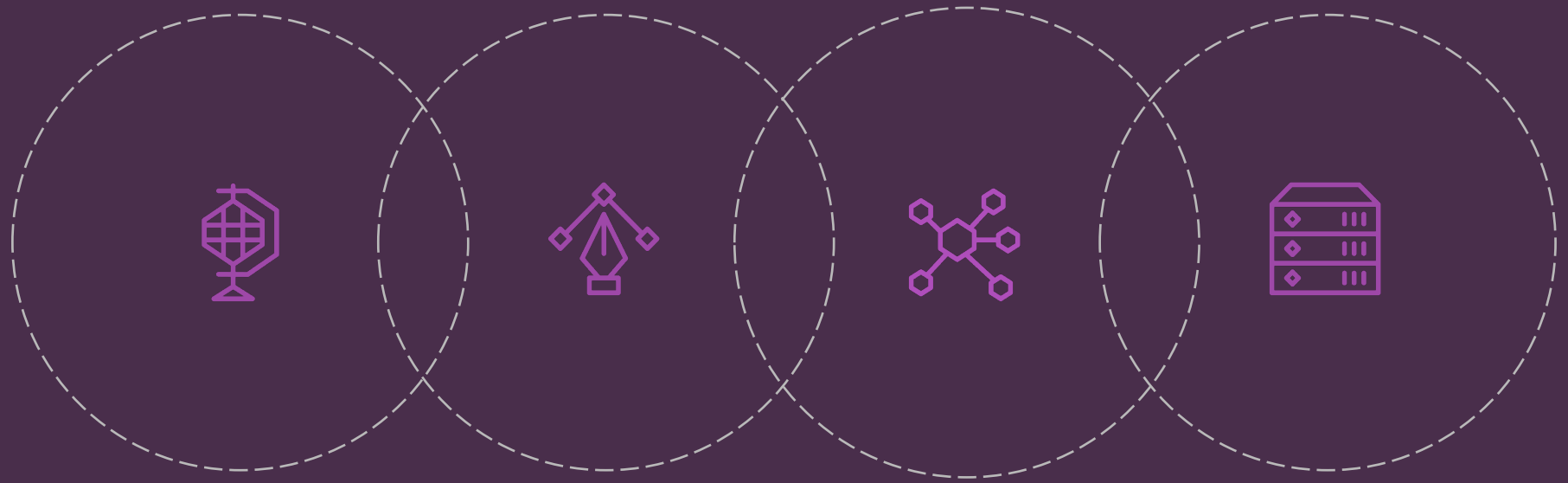
03/

I contenuti del Data Product

Quali contenuti produce il Data Product?

Gli output del Data Product

Al termine delle elaborazioni il Data Product rende disponibile la documentazione elaborata sotto forma di testi strutturati e categorizzati, insieme ai metadati semantici estratti dai materiali sorgente. Gli output sono disponibili in formato ZIP oppure in JSON o PDF e sono esposti tramite l'API REST del Data Product.



Testi strutturati

Metadati semantici

Dataset elaborato

Condition report

L’organizzazione del dataset

Le risorse del Data Product sono organizzate attraverso specifiche entità gestite su database relazionale PostgreSQL.

Le relazioni tra le entità collegano a ciascun reperto i condition report prodotti nel tempo, le immagini e la documentazione che ne testimonia lo stato.

Entità principali

- **Reperto**
  Oggetti con identificativo, codici inventariali, tipologia, cronologia, materiali, provenienza e stato di conservazione.
- **Exhibitions/Movimentazioni**
  Metadati delle mostre, prestiti e movimentazioni collegate ai reperti e che hanno determinato la generazione di condition report
- **Image**
  Asset immagine con i relativi metadati e le relazioni verso i beni.
- **Condition report**
  Versioni ed epoche del report associate al bene, confrontabili nel tempo.

04/

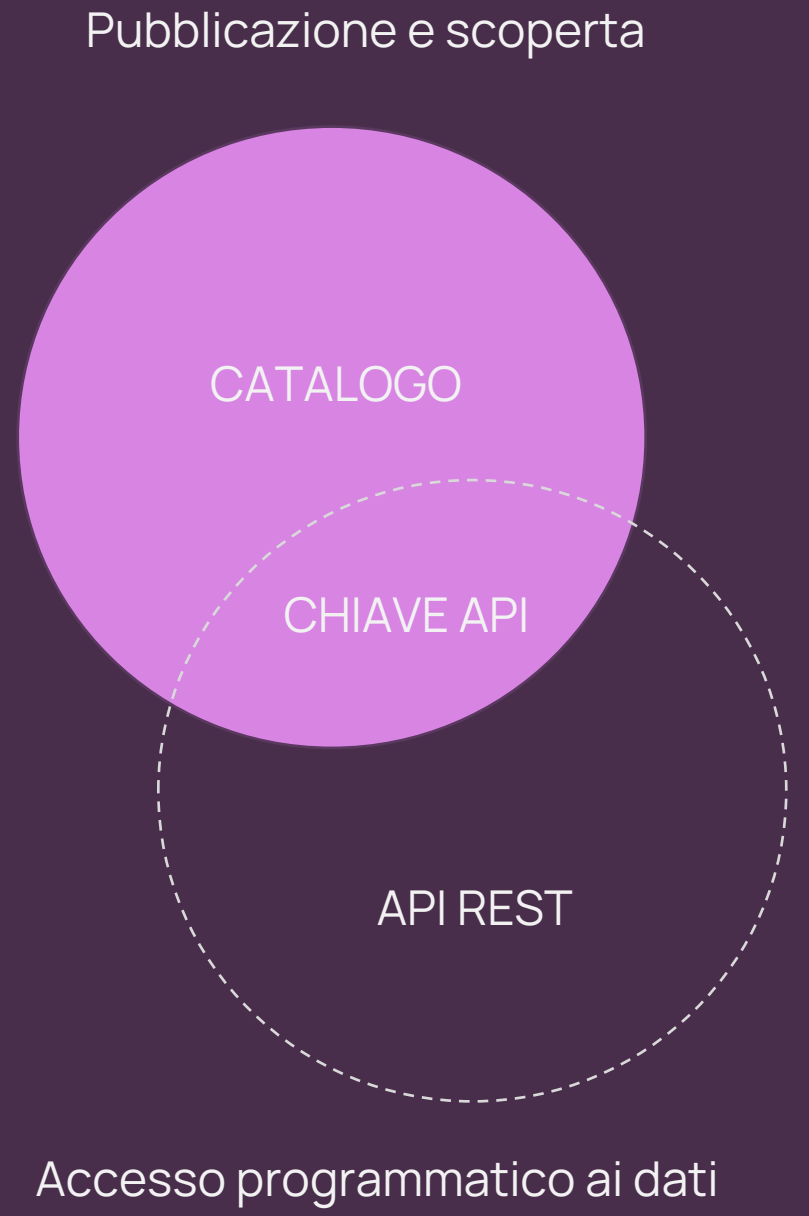
Fruizione del Data Product

Come si accede al Data Product?

Le modalità di fruizione

Il Data Product è classificato come Public ed è fruibile come API REST. L'accesso ai dati passa per due momenti:

- il catalogo DPaaS, dove il Data Product è censito e reso individuabile dagli enti autorizzati;
- l'accesso tramite API REST, che permette di interrogare le entità del dataset attraverso il web service GolangAPI e di integrarle in applicazioni e sistemi esterni.



Accesso tramite API REST

L'accesso al Data Product avviene tramite chiamate REST API esposte dal web service GolangAPI, che consentono di consultare in modo strutturato le entità del dataset e di integrarle in servizi e flussi applicativi esterni.

Gli endpoint esposti sono:

/cr/objects

/cr/objects/{IDobject}

/cr/objects/{IDobject}/reports

/cr/reports/{IDreport}



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

