Barga- Un museo accessibile a tutti

Documento di specifiche tecniche di elaborazione

Versione del documento: 1.0

Data: 28/08/2026

Sommario

1. Obiettivi del Data Product 4
2. Origine dei dati ed eventuali elaborazioni 5
3. Tecnologie utilizzate all'interno del Data Product Lab 11
3.1. Servizi abilitanti DPaaS 12
3.2. Eventuali tecnologie esterne a DPaaS 15
3.2.1. Modelli in locale 15
4. Pipeline di elaborazione e trasformazione 17
4.1. Pipeline di generazione di descrizioni e audiodescrizioni accessibili per immagini 17
4.1.1. Analisi dell'immagine e generazione della descrizione 17
4.1.2. Generazione dell'audiodescrizione in lingua italiana 17
4.1.3. Traduzione della descrizione in lingua inglese 18
4.1.4. Generazione dell'audiodescrizione in lingua inglese 18
4.1.5. Revisione e validazione degli artefatti 18
4.1.6. Pubblicazione degli artefatti 19
4.2. Pipeline di generazione di sottotitoli, trascrizioni, riassunti e audioriassunti per contenuti audiovisivi 19
4.2.1. Preparazione dell'audio e separazione dei parlanti 19
4.2.2. Trascrizione automatica e generazione dei sottotitoli 20
4.2.3. Traduzione dei sottotitoli in lingua inglese 20
4.2.4. Generazione della trascrizione completa 20
4.2.5. Generazione del riassunto in italiano 20
4.2.6. Generazione dell'audio del riassunto in italiano 21
4.2.7. Generazione del riassunto in lingua inglese 21
4.2.8. Generazione dell'audio del riassunto in inglese 21
4.2.9. Revisione e validazione degli artefatti 22
4.2.10. Pubblicazione degli artefatti 22
5. Esposizione e distribuzione 23
5.1. Canali di accesso 23

5.2. Distribuzione tramite dump 24
5.3. Utenti e sistemi destinatari 24
5.4. Autenticazione e autorizzazione 24
5.5. Formati disponibili 25
5.6. Modalità di aggiornamento 25
5.7. Limiti di utilizzo 26
5.8. Livelli di servizio 26
5.9. Accesso alla documentazione 26
5.10. Condizioni di licenza e riutilizzo 26
5.11. Fruizione tramite API 27
5.11.1. Tipologie di informazioni interrogabili 27
5.11.2. Principali operazioni disponibili 27
5.11.3. Ricerca, filtro, ordinamento e paginazione 28
5.11.4. Recupero del dettaglio di una risorsa 28
5.11.5. Servizi di aggregazione e analisi 28
5.11.6. Formati delle richieste e delle risposte 28
5.11.7. Modalità di autenticazione 29
5.11.8. Limiti di chiamata 29
5.11.9. Gestione degli errori 29
5.12. Nota sulle condivisioni dei dati 31
6. Possibili evoluzioni future 32

1.Obiettivi del Data Product

Il Data Product Barga è finalizzato alla generazione automatica di artefatti per l'accessibilità a partire dal patrimonio museale digitalizzato, comprendente fotografie di opere e oggetti, pannelli descrittivi e materiale documentale. L'obiettivo è rendere i contenuti fruibili da un'utenza ampia e diversificata, con particolare attenzione alle persone con disabilità visive, cognitive o con basso livello di alfabetizzazione.

2. Origine dei dati ed eventuali elaborazioni

1. Sistema, archivio, database, API o dataset di origine

La fonte originaria del Data Product è costituita dagli artefatti digitali e dai relativi metadati descrittivi conferiti dall'ente museale di Barga nell'ambito del progetto "Un museo accessibile a tutti".

Il dataset sorgente è identificato come DS00000105 e comprende contenuti relativi al patrimonio culturale e museale di Barga. Il Data Product è identificato come DP00000083. Gli artefatti e i relativi metadati vengono depositati in uno spazio di archiviazione dedicato, che costituisce il punto di ingresso tecnico dei dati nel Data Product.

Le strutture temporanee di appoggio, il database applicativo, i processi interni di elaborazione e le aree utilizzate per memorizzare i contenuti prodotti dalla soluzione sono componenti interne impiegate per acquisire, organizzare, elaborare e pubblicare i dati forniti dall'ente. Tali componenti non costituiscono fonti dati ulteriori: il passaggio dei dati attraverso di esse non modifica l'identificazione della fonte originaria, che rimane sempre costituita dagli artefatti digitali e dai metadati conferiti dall'ente.

2. Soggetto responsabile o titolare della fonte

L'ente museale di Barga è il soggetto responsabile del conferimento degli artefatti digitali sorgente e dei relativi metadati descrittivi.

L'ente conferente è responsabile della disponibilità, della correttezza e della completezza dei materiali trasmessi, nonché della corretta associazione tra ciascun artefatto e le relative informazioni descrittive. La titolarità del Data Product, la titolarità dei diritti sui singoli artefatti, sulle opere rappresentate e sulle eventuali riproduzioni digitali devono essere coerentemente indicate nella documentazione amministrativa e negli atti di conferimento.

Devono inoltre essere definite le condizioni applicabili ai contenuti provenienti da soggetti differenti, qualora alcuni materiali siano stati prodotti, conservati o resi disponibili da istituzioni diverse dall'ente museale.

3. Tipologia e formato dei dati

I dati sorgente conferiti per il Data Product di Barga sono costituiti principalmente da:

- fotografie e immagini relative alle opere e ai beni culturali;
- immagini di pannelli informativi e museali;
- contenuti audiovisivi;
- rappresentazioni digitali tridimensionali;
- metadati descrittivi associati agli artefatti;
- informazioni tecniche necessarie a identificare, classificare e localizzare ciascun file.

Tra i contenuti censiti sono presenti immagini in formato JPG, contenuti video in formato MP4 e rappresentazioni tridimensionali in formati quali GLB, FBX e OBJ. La disponibilità effettiva di ciascun file va sempre verificata in fase iniziale.

Per ciascun artefatto sono utilizzate, ove disponibili, le seguenti informazioni:

- codice identificativo dell'oggetto;
- nome del file;
- estensione e formato;
- percorso del file nello spazio di archiviazione;
- tipologia e sottotipologia dell'oggetto;
- titolo;
- descrizione;
- data o periodo di riferimento;
- luogo raffigurato o di provenienza;
- archivio di provenienza;
- ulteriori metadati descrittivi e di classificazione.

Queste informazioni consentono di identificare ciascun artefatto e di collegare il record descrittivo al file digitale corrispondente.

I metadati possono inoltre essere utilizzati per contestualizzare il contenuto dell'artefatto, collegandolo all'opera, al luogo, all'archivio o alla relativa sezione museale.

4. Modalità di acquisizione

L'acquisizione della fonte avviene mediante il caricamento degli artefatti digitali e dei relativi metadati nello spazio di archiviazione dedicato al Data Product.

I metadati possono essere conferiti mediante file strutturati, mentre i relativi contenuti digitali sono associati alle informazioni descrittive attraverso elementi identificativi condivisi, quali il codice dell'oggetto, il nome del file, il formato e la relativa posizione nello spazio di archiviazione.

Dopo il conferimento, la soluzione acquisisce i riferimenti agli artefatti e le relative informazioni descrittive, identifica i singoli oggetti digitali e registra gli elementi necessari alla loro gestione. Come indicato al punto 1, i dati possono transitare attraverso le strutture temporanee di appoggio prima di essere registrati nelle strutture applicative definitive, con funzione esclusivamente interna di supporto all'acquisizione, al controllo e all'organizzazione dei dati.

Prima dell'inserimento nel Data Product, i dati possono essere sottoposti alle seguenti attività:

- acquisizione dei file e dei metadati;
- verifica dell'associazione tra record descrittivi e artefatti;

- selezione delle informazioni necessarie;
- organizzazione dei dati secondo una struttura uniforme;
- normalizzazione dei valori e delle classificazioni;
- verifica della disponibilità dei file;
- integrazione dei metadati tecnici;
- collegamento degli oggetti alle relative categorie, opere o sezioni;
- eventuale conversione dei contenuti in formati maggiormente adatti alla fruizione.

5. Frequenza di aggiornamento

Il dataset sorgente non segue necessariamente una frequenza di aggiornamento periodica prestabilita.

L'aggiornamento avviene in relazione:

- al conferimento di nuovi artefatti;
- alla consegna di contenuti inizialmente mancanti;
- alla sostituzione di immagini o altri contenuti con versioni aggiornate;
- alla modifica o integrazione dei metadati descrittivi;
- alla correzione delle associazioni tra opere e relativi contenuti;
- alla disponibilità di materiali ancora in fase di produzione.

Alcuni contenuti risultano soggetti a completamento o sostituzione, anche in relazione alla realizzazione di nuove immagini e alla consegna di media non ancora inclusi nel materiale iniziale.

L'aggiornamento è pertanto gestito in occasione della disponibilità di nuovi dati o della modifica di quelli esistenti, anziché secondo una cadenza giornaliera, settimanale o mensile.

Un nuovo versionamento dovrebbe essere associato alle modifiche che incidono sulla struttura, sulle modalità di fruizione o sulle caratteristiche sostanziali del Data Product, mentre i normali aggiornamenti dei contenuti possono essere gestiti nell'ambito della versione corrente, secondo le regole di governance definite per il prodotto.

6. Volume indicativo dei dati

Il volume del dataset sorgente deve essere determinato considerando separatamente:

- il numero complessivo degli artefatti;
- il numero dei record di metadati;
- la ripartizione dei contenuti per tipologia;
- il numero di immagini;
- il numero e la durata dei contenuti audiovisivi;

- il numero e la dimensione dei contenuti tridimensionali;
- la dimensione complessiva dei file conferiti;
- gli eventuali contenuti ancora da consegnare o sostituire.

La dimensione indicativa del dataset sorgente conferito dal Museo Civico “A. Mordini” di Barga è pari a circa 3 GB.
Il volume della fonte originaria deve inoltre essere distinto dal volume complessivo del Data Product dopo le elaborazioni. Da un singolo artefatto sorgente possono infatti essere prodotti ulteriori contenuti, quali:

- descrizioni testuali;
- testi alternativi;
- contenuti semplificati;
- traduzioni;
- file audio;
- sottotitoli;
- miniature;
- copie convertite o ottimizzate;
- informazioni tecniche;
- ulteriori contenuti derivati destinati all'accessibilità.

Il volume complessivo occupato dal Data Product può quindi essere superiore a quello dei soli artefatti e metadati originari.

7. Eventuali vincoli di accesso, utilizzo o licenza

L'accesso tecnico allo spazio di archiviazione contenente i dati sorgente è riservato agli utenti e ai componenti applicativi autorizzati, mediante credenziali e profili di accesso dedicati.

I file sorgente non devono essere esposti attraverso un accesso diretto e indiscriminato allo spazio di archiviazione. La consultazione da parte delle applicazioni e degli utenti finali deve avvenire attraverso i meccanismi di fruizione previsti dal Data Product, mantenendo separato il livello di archiviazione da quello di consultazione.

La pubblicazione e il riuso dei contenuti devono essere disciplinati da condizioni coerenti con:

- la titolarità degli artefatti;
- i diritti sulle opere rappresentate;
- i diritti sulle riproduzioni fotografiche e digitali;
- la provenienza dei materiali;
- gli eventuali obblighi di attribuzione;

- le limitazioni alla modifica, alla redistribuzione o all'uso commerciale;
- la licenza associata al Data Product;
- la licenza eventualmente associata ai singoli contenuti;
- le condizioni applicabili ai contenuti derivati.

Le credenziali stabiliscono dunque chi può accedere operativamente ai sistemi, mentre la licenza e gli atti di conferimento definiscono le condizioni giuridiche secondo cui i contenuti possono essere elaborati, pubblicati e riutilizzati: le autorizzazioni tecniche e la licenza d'uso operano su piani distinti.

Deve inoltre essere verificato se tutti gli artefatti possano essere pubblicati integralmente oppure se, per alcune tipologie, la fruizione debba riguardare esclusivamente i metadati, una copia ottimizzata, una rappresentazione ridotta o i contenuti derivati.

8. Livello di qualità e completezza riscontrato

Il dataset comprende artefatti digitali e metadati descrittivi utilizzati per identificarli, classificarli e associarli alle opere e ai relativi contesti museali.

La qualità complessiva deve essere valutata considerando sia la correttezza dei metadati sia l'effettiva disponibilità dei file indicati.

I controlli devono riguardare in particolare:

- la presenza effettiva del file nello spazio di archiviazione;
- la corrispondenza tra il contenuto censito e il file conferito;
- la corrispondenza tra il nome indicato nei metadati e il nome effettivo del file;
- la correttezza dell'estensione e del formato;
- la coerenza del percorso di archiviazione;
- la leggibilità e l'integrità tecnica del file;
- la presenza dei metadati minimi necessari;
- la corretta associazione tra l'artefatto e la relativa scheda descrittiva;
- la coerenza tra opere, media associati, categorie e sezioni museali;
- la presenza di eventuali duplicati;
- l'individuazione dei contenuti mancanti, provvisori o da sostituire.

Particolare attenzione deve essere posta alla corrispondenza tra il file descrittivo e la cartella dei contenuti, soprattutto quando sono indicati contenuti audiovisivi o tridimensionali non ancora disponibili.

La qualità della fonte è sufficiente per l'acquisizione e l'elaborazione automatizzata, a condizione che venga verificata la coerenza tra file e metadati conferiti. Per il dataset non sono stati definiti KPI quantitativi di qualità, soglie formali di accettazione o una percentuale minima di completezza.

3. Tecnologie utilizzate all'interno del Data Product Lab

Il Data Product si avvale di alcuni componenti applicativi utilizzati per l'orchestrazione dei processi di elaborazione, la gestione dei flussi di lavoro, l'esecuzione delle pipeline di Intelligenza Artificiale e il monitoraggio dello stato di avanzamento delle elaborazioni. Tali componenti operano in integrazione con le funzionalità offerte dalla piattaforma e consentono l'automazione dell'intero ciclo di generazione degli artefatti di accessibilità.

In particolare, il processo è orchestrato tramite Apache NiFi, sul quale sono configurati job schedulati che effettuano periodicamente il monitoraggio delle cartelle S3 associate ai singoli Enti conferitori. I job verificano la presenza di nuovi contenuti e controllano, attraverso appositi meccanismi di tracciamento, se i file siano già stati elaborati dal sistema.

Nel caso in cui un contenuto risulti non ancora processato, NiFi attiva una sequenza di chiamate verso due API REST esposte sul Livello 1 dell'architettura applicativa. Entrambe le API sono state sviluppate in linguaggio Python e costituiscono componenti esterne rispetto alla piattaforma DPaaS.

La prima API è dedicata alla traduzione automatica dei metadati in lingua inglese, necessaria per garantire la disponibilità delle informazioni descrittive nelle lingue supportate dal Data Product. La seconda API implementa invece la pipeline di generazione degli artefatti, responsabile della creazione dei contenuti derivati quali descrizioni, audiodescrizioni, trascrizioni, sottotitoli, riassunti e ulteriori risorse di accessibilità.

Per ciascun Ente sono stati configurati specifici task di elaborazione, definiti sulla base dei requisiti funzionali e delle esigenze espresse durante la fase di raccolta dei fabbisogni. Ogni task identifica una particolare tipologia di elaborazione da eseguire sui contenuti conferiti dall'Ente e viene gestito in maniera indipendente all'interno della pipeline applicativa.

Lo stato di avanzamento delle elaborazioni viene registrato all'interno di un database PostgreSQL, anch'esso utilizzato come componente di supporto esterno alla piattaforma DPaaS. Per ogni elaborazione sono memorizzati identificativo del task, stato di esecuzione, informazioni di monitoraggio ed eventuali errori rilevati durante il processamento, consentendo il tracciamento completo del ciclo di vita delle lavorazioni.

A livello operativo, Apache NiFi implementa inoltre specifici meccanismi di retry automatico, che consentono di rieseguire le chiamate verso le API in caso di errori

temporanei o malfunzionamenti transitori, aumentando la resilienza complessiva del sistema e riducendo il rischio di perdita delle elaborazioni.

Al completamento di tutti i task previsti per uno specifico Ente, gli artefatti generati e i relativi metadati vengono considerati disponibili per la pubblicazione e possono essere successivamente fruiti dai portali applicativi che accedono al Data Product.

3.1. Servizi abilitanti DPaaS

Di seguito vengono riportati tutti i servizi abilitanti DPaas che sono stati associati dall'ente al Data Product e sono stati utilizzati al fine di generare gli artefatti:

- CPAI – OCR;
- CPAI - Speech2Text;
- CPAI - Text2Speech;
- CPAI – Ner;
- CPAI – Tagging;
- CPAI - Image Enhance;
- CPAI – Abstract;
- CPAI - Estrazione Entità;
- CPAI - Epigrafi Latine;
- CPAI - Abstract immagini;
- CPAI - Tagging Entità;
- CPAC – Yolo;
- CPAC – WhisperTTS;
- CPAC – Whisper;
- CPAC – Tagging;
- CPAC - Stable Diffusion;
- CPAC - Real ESRGAN;
- CPAC - Qwen2;
- CPAC - P2PaLa;
- CPAC – NERs;
- CPAC - Llama 31;
- CPAC – LayoutParser;
- CPAC – Keras;
- CPAC - HCR_Inference;
- CPAC- Gliner;

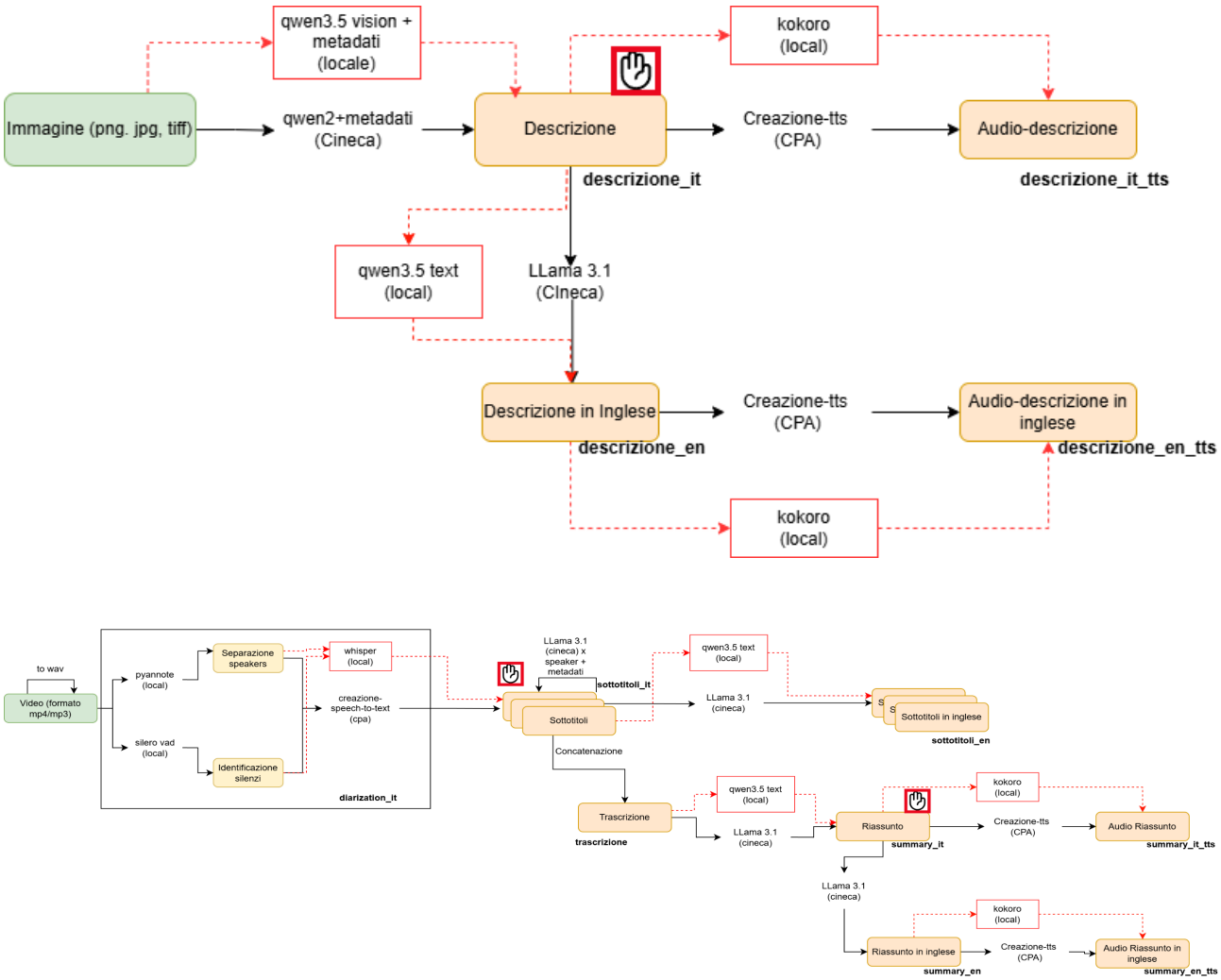
- HCR_DataPreparation;
- VM;
- JupyterLab;
- OpenSeaDragon;
- NiFi;
- Postgres.

Si evidenzia che **non tutti i servizi, modelli e componenti riportati nell'elenco delle tecnologie disponibili sono stati effettivamente utilizzati nell'ambito della presente fase progettuale**. La loro inclusione è stata richiesta all'Ente al fine di consentire un'attività di confronto tra diverse soluzioni tecnologiche, sia in termini di prestazioni sia di qualità dei risultati ottenuti, nonché per valutare la fattibilità di specifici requisiti funzionali e non funzionali.

L'associazione delle tecnologie al Data Product ha pertanto avuto anche una finalità esplorativa e di assessment, volta a identificare gli strumenti maggiormente idonei per l'evoluzione della soluzione. Alcuni dei servizi elencati potrebbero quindi non essere stati utilizzati nelle pipeline attualmente implementate, ma rappresentano opzioni tecnologiche valutate durante la fase di analisi o candidate per future evoluzioni del sistema.

La presenza di tali componenti nella documentazione non implica pertanto il loro impiego operativo nella versione corrente del Data Product, ma ne documenta la disponibilità e la potenziale adozione in successive fasi di sviluppo, ottimizzazione o ampliamento delle funzionalità previste.

Di seguito viene riportato l'albero elaborativo che descrive, per ogni tipologia di input, il flusso che è stato costruito e i servizi utilizzati per le varie trasformazioni.



Come è possibile vedere dagli alberi elaborativi, al fine di garantire l'accessibilità delle risorse vengono generati:

INPUT: *Immagine*
OUTPUT :

- Descrizione in Italiano ed in inglese
- Audio descrizione in Italiano ed in inglese

INPUT: *Video*
OUTPUT :

- Trascrizione in Italiano
- Riassunto in Italiano e in inglese

- Audio riassunto in Italiano e in inglese
- Sottotitoli simultanei in Italiano ed in inglese

3.2. Eventuali tecnologie esterne a DPaaS

3.2.1. Modelli in locale

Durante la fase di realizzazione del Data Product sono stati valutati e testati i servizi di Intelligenza Artificiale messi a disposizione dall'ecosistema DPaaS, verificandone l'applicabilità rispetto agli specifici casi d'uso, agli obiettivi progettuali e ai requisiti di accessibilità definiti dagli Enti coinvolti. Per alcune funzionalità, al fine di supportare le attività di sviluppo e collaudo e di assicurare la necessaria continuità nella sperimentazione, si è ritenuto opportuno affiancare ai servizi disponibili modelli eseguiti su infrastrutture esterne gestite dal fornitore. Tale configurazione ha mantenuto la piena integrazione con i workflow di orchestrazione e con i meccanismi di pubblicazione della piattaforma DPaaS.
Per la maggior parte delle funzionalità sono stati adottati modelli funzionalmente equivalenti ai servizi presenti sulla piattaforma DPaaS. L'impiego di istanze locali durante la fase di sviluppo ha consentito di configurare, verificare e ottimizzare i modelli in relazione agli specifici contenuti e requisiti del Data Product. Per la componente di elaborazione multimodale è stato utilizzato **Qwen 3.5 Vision**, mentre la piattaforma DPaaS rende attualmente disponibile la versione **Qwen2**.
La scelta di utilizzare Qwen 3.5 Vision è stata determinata dalla necessità di elaborare contenuti multimodali, in particolare immagini di opere, fotografie, pagine manoscritte e documenti digitalizzati. Rispetto a Qwen2, la versione 3.5 introduce capacità più evolute di interpretazione visiva, una migliore integrazione tra informazioni testuali e contenuti grafici, una maggiore accuratezza nell'estrazione del contenuto documentale e una qualità superiore nella generazione di descrizioni contestualizzate. Tali caratteristiche si sono dimostrate particolarmente rilevanti per la realizzazione degli artefatti di accessibilità previsti dal progetto, quali descrizioni di immagini, audiodescrizioni, trascrizioni di manoscritti e contenuti derivati destinati alla consultazione museale.
Considerata la rilevanza di tale componente per il corretto funzionamento delle pipeline implementate, è stato richiesto ai responsabili della piattaforma DPaaS di rendere disponibile e ospitare una versione del modello **Qwen 3.5 Vision** all'interno dell'infrastruttura ufficiale. L'adozione del medesimo modello consentirebbe infatti di mantenere allineati i risultati prodotti durante la fase di sviluppo con quelli generati successivamente in esercizio.

Si evidenzia pertanto che, a valle della messa in produzione del Data Product, tutti i nuovi contenuti caricati o modificati verranno elaborati mediante i modelli disponibili sulla piattaforma DPaaS. Qualora il modello **Qwen 3.5 Vision** non venga reso disponibile all'interno dell'infrastruttura, le elaborazioni che richiedono capacità di analisi visuale verranno eseguite utilizzando **Qwen2**. In tale scenario, pur mantenendo invariata la funzionalità complessiva del sistema, gli artefatti generati potrebbero presentare un livello di dettaglio, accuratezza e qualità descrittiva differente rispetto a quello ottenuto durante la fase di realizzazione e collaudo del Data Product, nella quale è stato utilizzato Qwen 3.5 Vision come modello di riferimento.

4. Pipeline di elaborazione e trasformazione

4.1. Pipeline di generazione di descrizioni e audiodescrizioni accessibili per immagini

La pipeline è progettata per trasformare immagini digitali del patrimonio culturale in contenuti descrittivi e audio destinati a migliorare l'accessibilità delle risorse digitali. A partire da immagini nei formati **PNG, JPG e TIFF**, il processo genera automaticamente descrizioni testuali e audiodescrizioni in lingua italiana e inglese, consentendo la fruizione dei contenuti anche attraverso modalità di ascolto. Tutti gli artefatti prodotti vengono archiviati nel Data Product e successivamente resi disponibili tramite il portale di consultazione. Gli elementi contrassegnati dall'icona della mano possono essere revisionati e modificati dagli utenti autorizzati prima della pubblicazione.

4.1.1. Analisi dell'immagine e generazione della descrizione

La pipeline prende in ingresso un'immagine digitale e la sottopone a un processo di analisi automatica mediante il modello multimodale **Qwen2**, eseguito presso l'infrastruttura di calcolo del **CINECA**.

Per migliorare la qualità dell'interpretazione del contenuto visivo, il modello utilizza anche i metadati associati all'immagine. Parallelamente interviene il modello **Qwen 3.5 Vision**, eseguito localmente, che contribuisce all'arricchimento delle informazioni estratte e alla costruzione di una rappresentazione più accurata del contenuto.

L'analisi consente di identificare gli elementi presenti nell'immagine, il contesto rappresentato, eventuali particolarità artistiche, architettoniche o documentarie e tutte le informazioni utili alla comprensione del contenuto.

Il risultato di questa fase è la generazione della **descrizione in lingua italiana** (*descrizione_it*), che rappresenta l'artefatto principale della pipeline.

Poiché tale contenuto costituisce la base per tutte le successive elaborazioni, il sistema consente agli utenti autorizzati di modificarlo e validarlo direttamente dal portale, correggendo eventuali inesattezze o arricchendo la descrizione con informazioni aggiuntive.

4.1.2. Generazione dell'audiodescrizione in lingua italiana

La descrizione italiana viene successivamente elaborata dal servizio **Creazione TTS (CPA)** per produrre una versione audio del contenuto.

La sintesi vocale viene realizzata utilizzando il modello **Kokoro**, eseguito localmente, che converte il testo in una narrazione audio naturale e facilmente comprensibile.

Da questa fase viene generato l'artefatto:

- **descrizione_it_tts**: audiodescrizione in lingua italiana.

L'audiodescrizione rappresenta una modalità alternativa di accesso all'informazione e consente agli utenti di fruire del contenuto mediante ascolto.

4.1.3. Traduzione della descrizione in lingua inglese

La descrizione italiana viene successivamente elaborata dal modello **Llama 3.1**, eseguito presso l'infrastruttura CINECA, per produrre la corrispondente versione inglese.

Durante questa fase interviene anche il modello **Qwen 3.5 Text**, eseguito localmente, che contribuisce alla qualità linguistica del contenuto tradotto e alla corretta preservazione del significato originale.

Il risultato è il seguente artefatto:

- **descrizione_en**: descrizione dell'immagine in lingua inglese.

La disponibilità della descrizione in lingua inglese consente la consultazione del patrimonio digitale anche da parte di utenti internazionali.

4.1.4. Generazione dell'audiodescrizione in lingua inglese

La descrizione inglese viene convertita in audio attraverso il servizio **Creazione TTS (CPA)** basato sul motore di sintesi vocale **Kokoro**.

Il risultato è il seguente artefatto:

- **descrizione_en_tts**: audiodescrizione in lingua inglese.

Questo contenuto permette la fruizione audio delle informazioni anche da parte di utenti non italofoni.

4.1.5. Revisione e validazione degli artefatti

La pipeline prevede una fase di supervisione umana sulla descrizione principale dell'immagine. In particolare può essere modificato e validato attraverso il portale il seguente artefatto:

- **descrizione_it**: descrizione dell'immagine in italiano.

L'integrazione tra Intelligenza Artificiale e revisione umana consente di garantire elevati livelli di accuratezza, qualità descrittiva e coerenza con il contesto culturale dei contenuti pubblicati.

4.1.6. Pubblicazione degli artefatti

Al termine dell'elaborazione il Data Product rende disponibili i seguenti contenuti:

- Descrizione dell'immagine in italiano (*descrizione_it*);
- Audiodescrizione in italiano (*descrizione_it_tts*);
- Descrizione dell'immagine in inglese (*descrizione_en*);
- Audiodescrizione in inglese (*descrizione_en_tts*).

Attraverso questa pipeline l'immagine originale viene trasformata in un insieme di contenuti descrittivi e multilingua, accessibili sia in formato testuale sia audio, contribuendo a migliorare l'inclusività e la fruibilità del patrimonio culturale digitale da parte di differenti categorie di utenti.

4.2. Pipeline di generazione di sottotitoli, trascrizioni, riassunti e audioriassunti per contenuti audiovisivi

La pipeline è progettata per trasformare contenuti audiovisivi in una serie di artefatti accessibili, multilingua e facilmente consultabili. A partire da file video e audio nei formati **MP4** e **MP3**, il processo combina tecnologie di riconoscimento del parlato, identificazione dei parlanti, traduzione automatica, sintesi e generazione di riassunti per produrre contenuti destinati sia alla consultazione tradizionale sia a modalità di fruizione assistita. Tutti gli artefatti generati vengono archiviati nel Data Product e successivamente pubblicati sul portale. Gli elementi contrassegnati con l'icona della mano possono essere revisionati e modificati dagli utenti autorizzati prima della pubblicazione definitiva.

4.2.1. Preparazione dell'audio e separazione dei parlanti

La pipeline prende in ingresso contenuti audiovisivi in formato **MP4** o **MP3**. Nella fase iniziale la traccia audio viene convertita in formato WAV per agevolare le successive elaborazioni.

Successivamente vengono avviati due processi paralleli:

- **Speaker Diarization**, realizzato mediante il framework **PyAnnote**, eseguito localmente, per identificare i diversi parlanti presenti nella registrazione;

- **Voice Activity Detection**, realizzato mediante **Silero VAD**, che individua automaticamente le pause e i segmenti di silenzio.

L'obiettivo di questa fase è segmentare accuratamente il contenuto audio, identificare chi sta parlando nei diversi momenti della registrazione e migliorare la qualità complessiva della trascrizione.

4.2.2. Trascrizione automatica e generazione dei sottotitoli

Le informazioni prodotte dalla diarizzazione vengono utilizzate dal servizio **Speech-to-Text (CPA)** basato sul modello **Whisper**, eseguito localmente.

Whisper genera la trascrizione del parlato mentre il modello **Llama 3.1**, eseguito presso il CINECA, integra il risultato con le informazioni provenienti dalla diarizzazione, dai metadati disponibili e dall'identificazione degli speaker.

Il risultato è la produzione dei **sottotitoli in lingua italiana** (*sottotitoli_it*), comprensivi della sincronizzazione temporale e delle informazioni utili all'identificazione dei parlanti.

Poiché i sottotitoli rappresentano uno degli artefatti principali della pipeline, essi possono essere modificati e validati dagli operatori autorizzati direttamente dal portale.

4.2.3. Traduzione dei sottotitoli in lingua inglese

I sottotitoli italiani vengono successivamente elaborati tramite **Llama 3.1** e **Qwen 3.5 Text** per generare la corrispondente versione inglese.

Il risultato di questa fase è:

- **sottotitoli_en** – sottotitoli in lingua inglese.

Questa elaborazione consente di rendere il contenuto audiovisivo accessibile anche a utenti internazionali.

4.2.4. Generazione della trascrizione completa

Parallelamente alla produzione dei sottotitoli, il sistema concatena tutti i segmenti temporizzati generati durante la trascrizione per ottenere una rappresentazione continua del contenuto parlato.

L'artefatto prodotto è:

- **trascrizione** – testo completo della registrazione.

La trascrizione rappresenta la base documentale per le elaborazioni di sintesi e permette di accedere integralmente ai contenuti senza la necessità di consultare il materiale audiovisivo originale.

4.2.5. Generazione del riassunto in italiano

La trascrizione viene elaborata mediante **Llama 3.1**, supportato dal modello **Qwen 3.5 Text**, per produrre una sintesi strutturata del contenuto della registrazione.

Il risultato è l'artefatto:

- **summary_it** – riassunto del contenuto in italiano.

Il riassunto evidenzia gli argomenti principali, i concetti più rilevanti e le informazioni maggiormente significative emerse durante il contenuto audiovisivo.

Questo artefatto può essere modificato e validato tramite il portale dagli utenti autorizzati, consentendo un controllo qualitativo e una verifica della correttezza delle informazioni prodotte automaticamente.

4.2.6. Generazione dell'audio del riassunto in italiano

Il riassunto italiano viene successivamente convertito in contenuto audio attraverso il servizio **Creazione TTS (CPA)**.

La sintesi vocale viene realizzata utilizzando il modello **Kokoro**, eseguito localmente.

Il risultato è:

- **summary_it_tts** – audio del riassunto in italiano.

L'utilizzo della sintesi vocale consente di fornire una modalità alternativa di accesso al contenuto, particolarmente utile per utenti che preferiscono l'ascolto alla lettura.

4.2.7. Generazione del riassunto in lingua inglese

A partire dal riassunto italiano, il modello **Llama 3.1** genera una versione tradotta in lingua inglese.

L'artefatto risultante è:

- **summary_en** – riassunto in inglese.

Questa elaborazione permette di rendere disponibili anche a utenti internazionali le informazioni più rilevanti contenute nella registrazione.

4.2.8. Generazione dell'audio del riassunto in inglese

La versione inglese del riassunto viene infine elaborata dal servizio di sintesi vocale CPA basato sul modello **Kokoro**.

Il risultato è:

- **summary_en_tts** – audio del riassunto in inglese.

In questo modo il contenuto risulta fruibile sia in formato testuale sia in formato audio nelle due lingue supportate.

4.2.9. Revisione e validazione degli artefatti

La pipeline prevede la supervisione umana sui contenuti maggiormente rilevanti per l'esperienza utente. In particolare, risultano modificabili dal portale:

- **sottotitoli_it**;
- **summary_it**.

Questa funzionalità consente agli operatori di correggere eventuali errori, integrare informazioni contestuali e migliorare la qualità dei contenuti prima della pubblicazione.

4.2.10. Pubblicazione degli artefatti

Al termine dell'elaborazione il Data Product rende disponibili i seguenti artefatti:

- Sottotitoli in italiano (*sottotitoli_it*);
- Sottotitoli in inglese (*sottotitoli_en*);
- Trascrizione completa (*trascrizione*);
- Riassunto in italiano (*summary_it*);
- Audio del riassunto in italiano (*summary_it_tts*);
- Riassunto in inglese (*summary_en*);
- Audio del riassunto in inglese (*summary_en_tts*).

Attraverso questa pipeline il contenuto audiovisivo viene trasformato in un insieme strutturato di artefatti testuali e audio che ne facilitano la consultazione, la comprensione e l'accessibilità, favorendo una fruizione inclusiva del patrimonio digitale da parte di utenti con esigenze e modalità di accesso differenti.

5. Esposizione e distribuzione

Il Data Product “Barga - Un museo accessibile a tutti” viene pubblicato nel catalogo della piattaforma DPaaS e reso fruibile agli utenti attraverso un’app mobile dedicata. Il backoffice dell’ap scarica un dump del Data Product contenente gli oggetti digitali, i metadati e gli artefatti necessari alla consultazione e utilizza localmente i contenuti acquisiti, evitando la necessità di interrogare le API in tempo reale durante la normale fruizione. Le API read-only rimangono disponibili come canale di interoperabilità e accesso programmatico al Data Product.

Il Data Product consente la consultazione strutturata di opere, oggetti museali digitalizzati e contenuti audiovisivi associati al patrimonio museale, insieme ai relativi metadati descrittivi e tecnici e agli artefatti di accessibilità generati dalle pipeline di elaborazione.

L’esposizione è articolata su due livelli funzionali:

- fruizione tramite app mobile dedicata, destinata agli utenti finali;
- distribuzione tramite dump del Data Product, scaricato e utilizzato localmente per la predisposizione dei dati per l’app mobile.

L’app mobile dedicata rappresenta il principale canale di consultazione per gli utenti finali. Consente di esplorare gli oggetti digitali e di accedere alle immagini, ai contenuti audiovisivi, alle informazioni descrittive e agli artefatti accessibili e multilingua prodotti dal Data Product. Per ridurre la dipendenza dalla connettività di rete, l’app utilizza una copia locale dei contenuti acquisita mediante il download del dump.

L’API read-only costituisce un canale separato di interoperabilità per i sistemi applicativi autorizzati. Espone gli oggetti e i relativi metadati, le classificazioni, le geolocalizzazioni, i cataloghi, le esposizioni e le relazioni con gli artefatti generati. Le operazioni disponibili sono esclusivamente di consultazione e non consentono ai consumer di modificare direttamente i dati esposti. La normale consultazione mediante app mobile non richiede chiamate API in tempo reale.

5.1. Canali di accesso

I canali di accesso previsti sono:

- app mobile dedicata a Barga;
- API read-only accessibili tramite protocollo HTTPS;
- catalogo pubblico DPaaS, utilizzato per la scoperta del Data Product, la consultazione delle informazioni descrittive e la richiesta di access.

Per gli utenti finali, la modalità ordinaria di fruizione è rappresentata dall'app mobile. L'app acquisisce una copia dei contenuti mediante il download delle informazioni del Data Product e la utilizza localmente, riducendo la dipendenza dalla connettività di rete durante la consultazione. Le funzionalità di revisione e validazione sono riservate agli operatori autorizzati e richiedono autenticazione e un profilo autorizzativo adeguato.

5.2. Distribuzione tramite dump

L'app mobile mediante la sua componente di backend accede ai contenuti attraverso un dump del Data Product. Il dump contiene le informazioni e le risorse necessarie alla consultazione degli oggetti digitali e dei relativi artefatti accessibili. Questa modalità separa il momento di acquisizione dei contenuti dal momento della loro fruizione.

L'utilizzo di una copia locale è finalizzato a mantenere fruibili i contenuti in condizioni di connettività assente o instabile. Le modalità di produzione del dump, la frequenza con cui viene reso disponibile e le regole con cui il backend dell'app verifica o acquisisce gli aggiornamenti dipendono dalla configurazione applicativa.

5.3. Utenti e sistemi destinatari

Il Data Product è destinato principalmente a:

- cittadini e visitatori interessati al patrimonio museale di Barga;
- persone che beneficiano di modalità alternative di consultazione dei contenuti;
- utenti internazionali, grazie alla disponibilità di contenuti in lingua inglese;
- operatori museali e personale incaricato della gestione e valorizzazione delle collezioni digitali;
- operatori incaricati della revisione e validazione degli artefatti;
- app mobile dedicata a Barga e relativi componenti applicativi;
- sistemi applicativi autorizzati che necessitano di interrogare i contenuti tramite API;
- enti culturali, ricercatori e sviluppatori interessati al riutilizzo dei dati e dei metadati, nel rispetto delle condizioni di licenza applicabili.

5.4. Autenticazione e autorizzazione

La consultazione dei contenuti avviene attraverso l'app mobile dedicata, secondo le modalità di accesso previste dall'applicazione. Le funzionalità riservate agli operatori, quali revisione, modifica e validazione degli artefatti generati, sono accessibili mediante funzionalità applicative riservate, previa autenticazione e sulla base dei ruoli assegnati.

In particolare, possono essere sottoposti a revisione da parte degli operatori autorizzati:

- la descrizione principale dell'immagine in italiano;

- i sottotitoli in italiano;
- il riassunto in italiano.

La revisione di uno di questi contenuti può determinare la rigenerazione degli artefatti derivati, quali traduzioni e contenuti audio.

Per l'accesso machine-to-machine, le API sono protette mediante credenziali applicative previste dall'ambiente di esercizio. Le credenziali dei componenti backend e dei sistemi di memorizzazione non sono esposte ai consumer.

Le modalità definitive di gestione delle credenziali, della loro durata e delle eventuali restrizioni di rete sono definite nell'ambiente di esercizio.

5.5. Formati disponibili

I dati e i metadati restituiti dalle API sono disponibili in formato JSON. Le risorse gestite dal Data Product comprendono:

- immagini nei formati PNG, JPG e TIFF;
- contenuti audiovisivi nei formati MP4 e MP3;
- descrizioni e audiodescrizioni in italiano e inglese;
- sottotitoli in italiano e inglese;
- trascrizioni;
- riassunti in italiano e inglese;
- audio-riassunti in italiano e inglese;
- metadati tecnici e descrittivi strutturati;
- informazioni relative a tag, cataloghi, esposizioni, relazioni e geolocalizzazioni.

Le risorse binarie, quali immagini, video e contenuti audio, vengono rese disponibili ai consumer attraverso i canali applicativi dedicati, senza esporre direttamente i sistemi interni di memorizzazione.

5.6. Modalità di aggiornamento

Il Data Product ha natura prevalentemente statica e non real-time. Gli aggiornamenti dei contenuti possono essere effettuati manualmente, a seguito del caricamento o della revisione di file e metadati, oppure mediante esecuzioni schedulate delle pipeline secondo la configurazione dell'ambiente.

Al completamento delle elaborazioni, i contenuti e gli artefatti destinati alla fruizione mobile vengono inclusi nel dump del Data Product. L'app utilizza la copia scaricata dal backend rielaborata e non necessita di interrogare le API in tempo reale durante la normale consultazione.

La disponibilità di un contenuto aggiornato nell'app dipende dalla sua pubblicazione nel Data Product, dalla messa a disposizione del dump aggiornato e dal successivo download da parte del backend.

La correzione di un artefatto revisionabile può richiedere la rigenerazione degli output dipendenti. Quando l'aggiornamento riguarda esclusivamente dati o artefatti, senza modificare il modello informativo, i contratti delle API o le logiche funzionali, non è necessario introdurre una nuova versione del Data Product.

5.7. Limiti di utilizzo

Le funzioni di ricerca adottano meccanismi di paginazione per limitare la dimensione delle risposte e rendere gestibile la consultazione di insiemi numerosi di oggetti.

Gli eventuali limiti relativi al numero di richieste, alle richieste concorrenti, alla dimensione massima delle risposte, ai timeout e alle soglie di utilizzo sono definiti nella configurazione dell'ambiente di esercizio.

5.8. Livelli di servizio

Il servizio prevede controlli tecnici dedicati alla verifica della disponibilità applicativa e dello stato di funzionamento dei componenti esposti.

I livelli quantitativi di servizio, inclusi disponibilità, tempi di risposta, tempi di ripristino, finestre di manutenzione e modalità di supporto, sono definiti nell'accordo di esercizio della piattaforma.

5.9. Accesso alla documentazione

Il Data Product è corredato dalla documentazione prevista per la pubblicazione nel catalogo DPaaS, comprendente:

- descrizione pubblica del Data Product;
- manuale utente;
- manuale operativo;
- specifiche tecniche di elaborazione;
- documentazione tecnica delle API;
- specifica formale delle API e relativa interfaccia di consultazione.

La documentazione pubblica descrive a livello funzionale le modalità di fruizione. Il dettaglio tecnico delle operazioni, dei parametri, dei modelli dati e delle risposte è reso disponibile secondo le modalità previste dal catalogo DPaaS e dal processo di richiesta di consumo.

5.10. Condizioni di licenza e riutilizzo

I dati, i metadati e gli artefatti sono riutilizzabili secondo la licenza indicata nella scheda del Data Product pubblicata nel catalogo DPaaS. Restano applicabili gli eventuali diritti e vincoli associati alle opere, agli oggetti museali, ai contenuti audiovisivi, agli altri asset messi a disposizione dall'Ente e agli artefatti derivati dalle elaborazioni.

La licenza definitiva applicabile al Data Product e ai suoi contenuti è quella selezionata e pubblicata dal titolare nella relativa scheda del catalogo DPaaS.

5.11. Fruizione tramite API

L'API di Barga fornisce accesso programmatico, in sola lettura, agli oggetti digitali del Data Product e ai relativi contenuti accessibili. Consente alle applicazioni consumer di ricercare, recuperare e navigare le risorse del patrimonio museale, insieme ai metadati descrittivi e tecnici, alle informazioni di classificazione e geolocalizzazione e agli artefatti generati dalle pipeline di accessibilità.

La normale consultazione dei contenuti attraverso l'app mobile non richiede l'interrogazione in tempo reale delle API. L'app scarica un dump del Data Product e utilizza nel backend le informazioni e le risorse in esso contenute. Le API rimangono disponibili per le esigenze di interoperabilità e per gli altri accessi programmatici autorizzati.

5.11.1. Tipologie di informazioni interrogabili

Le principali tipologie di informazioni interrogabili comprendono:

- oggetti digitali e relativi identificativi;
- opere, oggetti museali digitalizzati, immagini e contenuti audiovisivi;
- riferimenti ai file e agli asset associati;
- titolo, descrizione, autore o creator, quando presenti nei metadati;
- tipologia, sottotipo, lingua e stato del ciclo di vita;
- metadati descrittivi e tecnici;
- tag e raggruppamenti tematici;
- cataloghi, esposizioni, luoghi e geolocalizzazioni;
- relazioni tra oggetti e tra oggetti originali e artefatti derivati;
- descrizioni, audiodescrizioni, sottotitoli, trascrizioni, riassunti e audio-riassunti disponibili.

5.11.2. Principali operazioni disponibili

Le principali operazioni di consultazione comprendono:

- ricerca degli oggetti digitali;
- recupero del dettaglio di un singolo oggetto;

- recupero dei metadati associati a una risorsa;
- recupero degli oggetti collegati e consultazione delle relazioni;
- consultazione di stati, lingue, tipi, sottotipi, tag e tipi di relazione;
- consultazione di cataloghi, esposizioni e geolocalizzazioni;
- recupero delle risorse binarie associate agli oggetti.

Le operazioni di consultazione non modificano il contenuto del Data Product. Le funzioni di revisione e validazione sono riservate alle funzionalità applicative destinate agli operatori autorizzati.

5.11.3. Ricerca, filtro, ordinamento e paginazione

La ricerca consente di combinare criteri riferiti agli identificativi e ai metadati degli oggetti. I criteri documentati comprendono codice, titolo, autore o creator, lingua, stato, tipologia e sottotipo, catalogo, esposizione, luogo o geolocalizzazione e tag.

I risultati possono essere ordinati rispetto ai campi supportati e vengono restituiti con paginazione. Per il patrimonio di Barga assumono particolare rilevanza la tipologia dell'oggetto museale, il catalogo o l'esposizione di appartenenza, il luogo, la lingua e la disponibilità degli artefatti accessibili.

5.11.4. Recupero del dettaglio di una risorsa

Il dettaglio di una risorsa viene recuperato tramite il relativo identificativo. La risposta può comprendere le informazioni principali dell'oggetto, la classificazione, la tipologia e il sottotipo, lo stato, i riferimenti agli asset, il catalogo, l'esposizione, le informazioni geografiche e i metadati disponibili.

Le relazioni consentono di individuare oggetti collegati e artefatti generati a partire dalla risorsa originale.

5.11.5. Servizi di aggregazione e analisi

Le API rendono disponibili aggregazioni funzionali alla navigazione per cataloghi, esposizioni e geolocalizzazioni, includendo il numero di oggetti associati quando previsto.

Le API non costituiscono un sistema di Business Intelligence e non forniscono KPI o reporting multidimensionale. Le aggregazioni sono finalizzate principalmente alla navigazione del patrimonio e alla costruzione dell'esperienza di consultazione.

5.11.6. Formati delle richieste e delle risposte

Le richieste utilizzano il protocollo HTTPS. Le risposte informative sono restituite in formato JSON; i contenuti binari vengono restituiti nel formato coerente con la risorsa.

La struttura delle richieste, delle risposte e dei modelli informativi è descritta nella documentazione tecnica delle API.

5.11.7. Modalità di autenticazione

L'accesso machine-to-machine avviene mediante credenziali applicative previste dall'ambiente di esercizio. Le credenziali dei sistemi interni di memorizzazione non vengono trasferite ai consumer. L'accesso degli utenti finali avviene tramite l'app mobile dedicata, secondo le modalità previste dall'applicazione, mentre le funzionalità riservate agli operatori richiedono autenticazione e un ruolo applicativo autorizzato.

5.11.8. Limiti di chiamata

La paginazione controlla la quantità di dati restituita dalle operazioni di ricerca. Gli eventuali limiti relativi a frequenza delle chiamate, chiamate concorrenti, dimensione massima della pagina, timeout e dimensione delle risposte sono definiti nella configurazione dell'ambiente di esercizio.

5.11.9. Gestione degli errori

Le API utilizzano codici di stato standard per comunicare l'esito delle richieste. Le principali categorie di errore riguardano parametri mancanti o non validi, risorse non esistenti, autenticazione assente o non valida, accesso non autorizzato, superamento degli eventuali limiti di utilizzo, indisponibilità temporanea dei servizi interni ed errori applicativi.

Le risposte di errore sono restituite in formato strutturato e contengono le informazioni necessarie per identificare la tipologia del problema. Gli errori della pipeline di alimentazione vengono gestiti separatamente mediante la registrazione dello stato del task, dello step interessato e del relativo esito, così da consentire monitoraggio e ripresa dell'elaborazione.

6. Limitazioni

Il Data Product Barga è costituito dall'insieme di oggetti digitali, metadati e contenuti audiovisivi conferiti nell'ambito del progetto. Non rappresenta necessariamente l'intero patrimonio museale dell'Ente e la sua completezza dipende dalla qualità e dalla disponibilità dei materiali ricevuti.

Il Data Product ha natura prevalentemente statica e non real-time. Gli aggiornamenti non avvengono in modo continuo, ma sono effettuati manualmente, a seguito della consegna o revisione di file e metadati, oppure mediante esecuzioni schedulate delle pipeline. Tra la modifica della fonte e la pubblicazione del dato aggiornato può quindi intercorrere un intervallo legato all'avvio e al completamento delle elaborazioni.

Gli artefatti generati automaticamente possono richiedere verifica da parte degli operatori autorizzati. La revisione documentata riguarda la descrizione dell'immagine in italiano, i sottotitoli in italiano e il riassunto in italiano. La qualità degli output dipende anche dalla qualità dei contenuti e dei metadati in ingresso.

Le API sono di sola lettura. La modifica e la validazione dei contenuti non sono disponibili ai consumer esterni tramite le API, ma sono gestite mediante funzionalità applicative riservate agli operatori autorizzati.

Poiché l'app mobile utilizza una copia locale del Data Product, i contenuti consultabili sul dispositivo corrispondono a quelli presenti nel dump scaricato. Eventuali aggiornamenti apportati successivamente al Data Product diventano disponibili al backend dell'app solo dopo la messa a disposizione e il download di un dump aggiornato.

5.12. Nota sulle condivisioni dei dati

La condivisione, la pubblicazione e il riutilizzo dei dati e degli artefatti del Data Product sono regolati dalle condizioni indicate nella scheda pubblicata nel catalogo DPaaS e dagli eventuali diritti associati ai manoscritti, alle immagini digitali e ai fondi archivistici di provenienza.

La fruizione pubblica riguarda i contenuti dichiarati pubblicabili dal titolare. Le funzionalità di revisione, modifica e validazione sono riservate agli utenti autenticati e autorizzati. L'accesso machine-to-machine è subordinato alle credenziali e alle regole configurate nell'ambiente di esercizio.

La condivisione può riguardare le immagini originali, i metadati, una selezione di attributi o i soli artefatti derivati, secondo quanto definito dal titolare e riportato nella licenza applicabile. Gli obblighi di attribuzione, gli eventuali divieti di modifica o riutilizzo commerciale e i vincoli derivanti da diritti di terzi devono essere rispettati dai consumer.

Le credenziali tecniche e le informazioni necessarie all'accesso diretto alle componenti backend non rientrano nei dati condivisi. I consumer accedono esclusivamente attraverso i canali di esposizione previsti dal Data Product.

6. Possibili evoluzioni future

L'architettura realizzata è stata concepita in modo modulare, al fine di consentire future evoluzioni sia dal punto di vista tecnologico sia dal punto di vista funzionale. Una prima area di miglioramento riguarda la possibilità di sostituire o aggiornare i modelli di Intelligenza Artificiale attualmente impiegati con versioni più recenti o con modelli differenti, qualora disponibili sulla piattaforma DPaaS o ritenuti maggiormente performanti rispetto agli specifici casi d'uso. Tale approccio consentirebbe di beneficiare dei continui progressi tecnologici nel campo dell'AI senza richiedere modifiche significative ai processi di orchestrazione e pubblicazione del Data Product.

Un'ulteriore evoluzione riguarda la trasformazione delle pipeline attualmente sviluppate nell'ambito del progetto in veri e propri **servizi CPA (Common Processing Activities)** integrati all'interno dell'ecosistema DPaaS. In questo scenario, le logiche di generazione degli artefatti verrebbero rese disponibili come servizi condivisi e riutilizzabili, consentendo agli Enti di invocare un unico servizio applicativo per ottenere direttamente gli artefatti di accessibilità richiesti. Questa evoluzione permetterebbe di semplificare i processi di integrazione, ridurre la complessità operativa e favorire il riuso delle componenti sviluppate anche in altri Data Product o progetti futuri.

Dal punto di vista funzionale, potrebbe inoltre essere introdotto un meccanismo di **aggiornamento dinamico del Data Product**, che consenta l'acquisizione automatica di nuovi contenuti successivamente alla pubblicazione iniziale. In tale scenario, l'aggiunta di nuovi file di input da parte degli Enti attiverebbe automaticamente le pipeline di elaborazione e la successiva produzione degli artefatti derivati, mantenendo il Data Product costantemente aggiornato senza richiedere interventi manuali di riconfigurazione.

Tale funzionalità non è stata implementata nella presente versione della soluzione a causa di alcune limitazioni tecnologiche e operative riscontrate nell'attuale piattaforma DPaaS, che non consentono una gestione completamente automatizzata dell'evoluzione del contenuto del Data Product una volta completato il ciclo di pubblicazione iniziale. Qualora tali limitazioni venissero superate in future versioni della piattaforma, sarebbe possibile estendere l'architettura per supportare scenari di aggiornamento continuo e incrementale dei contenuti, migliorando ulteriormente la capacità del sistema di adattarsi all'evoluzione del patrimonio digitale degli Enti aderenti.