Una passeggiata nella Lendinara di fine Ottocento

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 10
3.1. Servizi abilitanti DPaaS 11
3.2. Eventuali tecnologie esterne a DPaaS 14
3.2.1. Modelli in locale 14
4. Pipeline di elaborazione e trasformazione 16
4.1. Pipeline di generazione di contenuti accessibili per immagini fotografiche 16
4.1.1. Analisi dell'immagine e generazione della descrizione principale 16
4.1.2. Traduzione della descrizione in lingua inglese 16
4.1.3. Produzione delle audiodescrizioni 17
4.1.4. Generazione del testo alternativo 17
4.1.5. Revisione e validazione dei contenuti 18
4.1.6. Pubblicazione degli artefatti 18
4.2. Pipeline di estrazione, traduzione e generazione audio per documenti PDF 19
4.2.1. Acquisizione e segmentazione del documento PDF 19
4.2.2. Estrazione del testo dalle pagine 19
4.2.3. Generazione della traduzione in lingua inglese 19
4.2.4. Produzione dell'audio della trascrizione in italiano 20
4.2.5. Produzione dell'audio della trascrizione in inglese 20
4.2.6. Revisione e validazione degli artefatti 20
4.2.7. Pubblicazione degli artefatti 20
5. Esposizione e distribuzione 21
5.1. Canali di accesso 21
5.2. Utenti e sistemi destinatari 22
5.3. Autenticazione e autorizzazione 22
5.4. Formati disponibili 23
5.5. Modalità di aggiornamento 23

5.6. Limiti di utilizzo 24
5.7. Livelli di servizio 24
5.8. Accesso alla documentazione 25
5.9. Condizioni di licenza e riutilizzo 25
5.10. Fruizione tramite API 26
5.10.1. Tipologie di informazioni interrogabili 26
5.10.2. Principali operazioni disponibili 26
5.10.3. Ricerca, filtro, ordinamento e paginazione 27
5.10.4. Recupero del dettaglio di una risorsa 27
5.10.5. Servizi di aggregazione e analisi 28
5.10.6. Formati delle richieste e delle risposte 28
5.10.7. Modalità di autenticazione 28
5.10.8. Limiti di chiamata 29
5.10.9. Gestione degli errori 29
5.11. Limitazioni 30
5.12. Nota sulle condivisioni dei dati 31
6. Possibili evoluzioni future 32

1.Obiettivi del Data Product

Il Data Product ha la finalità di rendere accessibili, in modo strutturato e automatizzato, le fotografie e i documenti storici della collezione del Comune di Lendinara. Il sistema è progettato per ampliare la fruizione del patrimonio culturale verso un'utenza ampia e diversificata, con particolare attenzione alle persone con disabilità visive, cognitive o con basso livello di alfabetizzazione. Per ciascun contenuto in ingresso vengono prodotti: Manuale Operativo di Rilascio in Produzione; Data Management Lendinara; molteplici artefatti accessibili tra cui descrizioni testuali, trascrizioni, traduzioni in lingua inglese, contenuti semplificati e tracce audio generate tramite sintesi vocale.

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 dal Comune di Lendinara nell'ambito del progetto "Una passeggiata nella Lendinara di fine Ottocento".

Il dataset sorgente è identificato come DS00000119, "DS Standard 2 - Artefatti del Comune di Lendinara", e appartiene al dominio archivistico.

Gli artefatti e i relativi metadati vengono depositati dall'ente in uno spazio di archiviazione dedicato, che costituisce il punto di ingresso tecnico dei dati nel Data Product. Le tabelle di staging, il database applicativo, le pipeline di elaborazione e le aree utilizzate per memorizzare gli output sono componenti interne della soluzione, impiegate per acquisire, organizzare, elaborare e pubblicare i dati forniti dall'ente, e non costituiscono fonti dati ulteriori.

2. Soggetto responsabile o titolare della fonte

Il Comune di Lendinara è l'ente proprietario del Data Product e il soggetto responsabile del conferimento degli artefatti digitali sorgente e dei relativi metadati descrittivi.

È dunque responsabile della disponibilità, della correttezza e della completezza dei materiali trasmessi alla soluzione, nonché della corretta associazione tra ciascun artefatto e le relative informazioni descrittive.

La titolarità dei diritti sui singoli artefatti, sui fondi archivistici e sulle eventuali riproduzioni digitali deve essere definita nella documentazione amministrativa e di conferimento.

3. Tipologia e formato dei dati

I dati sorgente conferiti dal Comune di Lendinara sono costituiti principalmente da:

- immagini e fotografie storiche;
- documenti digitalizzati;
- metadati descrittivi associati agli artefatti;

- informazioni tecniche necessarie a identificare, classificare e localizzare ciascun file.

I formati sorgente gestiti comprendono JPG, JPEG, PNG, TIFF, PDF e JSON. Le immagini, comprese quelle in formato TIFF, e i documenti PDF vengono indirizzati verso specifici percorsi di elaborazione in base al formato del file.

Per ciascun artefatto sono utilizzate, ove disponibili:

- il codice identificativo dell'oggetto;
- il nome del file;
- l'estensione e il formato;
- il percorso del file nello spazio di archiviazione;
- la tipologia e sottotipologia dell'oggetto;
- il titolo e la descrizione;
- la lingua;
- ulteriori metadati descrittivi e di classificazione.

Queste informazioni permettono di collegare in modo univoco il record descrittivo all'artefatto digitale corrispondente e di individuare il processo di elaborazione applicabile.

4. Modalità di acquisizione

L'acquisizione della fonte avviene mediante caricamento degli artefatti e dei relativi metadati nello spazio di archiviazione dedicato al Data Product. La modalità di origine del dataset è classificata come "Locale (upload)".

Dopo il conferimento, la pipeline di ingestion acquisisce i metadati e i riferimenti ai file, identifica i singoli oggetti digitali e registra nel database del Data Product le informazioni necessarie alla loro gestione. Il collegamento tra ciascun record descrittivo e il relativo artefatto avviene tramite il codice identificativo, il nome, l'estensione e il percorso del file.

I dati possono transitare attraverso strutture tecniche temporanee prima di essere registrati nelle tabelle applicative definitive; si tratta di strutture con funzione esclusiva

di staging, che non modificano l'identificazione della fonte originaria — il patrimonio digitale conferito dal Comune di Lendinara.

In base al formato sorgente, la soluzione distingue automaticamente il percorso di elaborazione:

- i file JPG, JPEG, PNG, TIF e TIFF vengono indirizzati alla pipeline dedicata alle immagini;
- i file PDF vengono indirizzati alla pipeline dedicata ai documenti.

5. Frequenza di aggiornamento

Il dataset sorgente non segue una frequenza di aggiornamento periodica prestabilita. L'aggiornamento avviene in relazione al conferimento, alla sostituzione o alla modifica di artefatti e metadati da parte del Comune di Lendinara.

Quando vengono resi disponibili nuovi contenuti, o quando contenuti esistenti vengono modificati, la soluzione può rieseguire i processi di acquisizione e di elaborazione interessati: l'aggiornamento è gestito su evento o su richiesta, non secondo una cadenza giornaliera, settimanale o mensile.

Modificare il contenuto di un artefatto non richiede necessariamente la rielaborazione completa di tutti i dati: la soluzione consente di riprendere l'elaborazione dal passaggio interessato dalla modifica. Ad esempio, una descrizione o una trascrizione corretta manualmente può diventare il nuovo punto di partenza per rigenerare gli output successivi, come traduzioni, testi semplificati o versioni audio.

Gli aggiornamenti riguardanti esclusivamente dati o artefatti non determinano una nuova versione del Data Product, purché restino invariati la struttura dei dati, le pipeline applicative e le modalità di esposizione. Un nuovo versionamento è invece necessario in presenza di modifiche strutturali alla soluzione.

6. Volume indicativo dei dati

La dimensione indicativa del dataset sorgente conferito dal Comune di Lendinara è pari a circa 1 GB.

Il volume si riferisce agli artefatti e ai metadati originari forniti dall'ente, e non comprende gli ulteriori contenuti generati dalla soluzione durante le elaborazioni, quali:

- copie convertite delle immagini;
- descrizioni testuali;

- testi alternativi;
- versioni semplificate o easy-to-read;
- traduzioni;
- file audio generati tramite sintesi vocale;
- miniature degli artefatti;
- file JSON contenenti gli esiti e i metadati tecnici delle elaborazioni.

Poiché da un singolo artefatto sorgente possono essere prodotti diversi contenuti derivati, il volume complessivo occupato dal Data Product al termine delle elaborazioni può superare quello della fonte grezza; 1 GB resta quindi il solo riferimento dimensionale della fonte originaria.

7. Eventuali vincoli di accesso, utilizzo o licenza

L'accesso tecnico allo spazio di archiviazione contenente i dati sorgente è consentito esclusivamente agli utenti e ai componenti applicativi autorizzati, mediante credenziali dedicate e profili di accesso specifici per il Data Product.

I file non sono esposti pubblicamente attraverso un accesso diretto e indiscriminato allo storage: l'eventuale accesso agli artefatti da parte delle applicazioni e degli utenti finali deve avvenire attraverso i meccanismi applicativi previsti dalla soluzione, mantenendo separati il livello di archiviazione e il livello di fruizione.

Il Data Product prevede una modalità di fruizione pubblica del prodotto finale; la pubblicazione e il riuso dei contenuti dovranno tuttavia essere disciplinati da una licenza esplicitamente associata al Data Product.

Le credenziali consentono l'accesso operativo ai sistemi, mentre la licenza e gli atti di conferimento stabiliscono le condizioni giuridiche secondo cui i contenuti possono essere elaborati, pubblicati e riutilizzati: autorizzazioni tecniche e licenza d'uso restano quindi su piani distinti.

8. Livello di qualità e completezza riscontrato

Il dataset contiene gli artefatti digitali e i metadati descrittivi necessari alla loro identificazione, classificazione e successiva elaborazione. La soluzione verifica la presenza delle informazioni indispensabili per individuare ciascun oggetto e collegarlo al relativo file sorgente.

I controlli riguardano in particolare:

- la presenza effettiva del file nello spazio di archiviazione;
- la corrispondenza tra il nome dichiarato 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;
- l'assenza di elaborazioni già attive o completate sul medesimo oggetto.

Per i dati sorgente è stato riscontrato un possibile disallineamento tra alcuni formati dichiarati nei metadati e quelli dei file effettivamente conferiti, in particolare immagini indicate come PNG ma presenti in formato TIFF: una situazione che richiede la validazione, ed eventualmente la correzione, delle informazioni descrittive prima dell'elaborazione.

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 – Translation;
- CPAI – Ner;
- CPAI – Tagging;
- CPAI - Image Enhance;
- CPAI – Abstract;
- CPAI - Estrazione Entità;
- CPAI - Abstract immagini;
- CPAI - Tagging Entità;
- CPAC – Yolo;
- CPAC – WhisperTTS;
- CPAC – Whisper;
- CPAC – TrOCR;
- CPAC – Tagging;
- CPAC - Stable Diffusion;
- CPAC - Real ESRGAN;
- CPAC - Qwen2;
- CPAC - P2PaLa;
- CPAC – NERs;
- CPAC - Llama 31;
- CPAC – Keras;
- CPAC – Gliner;

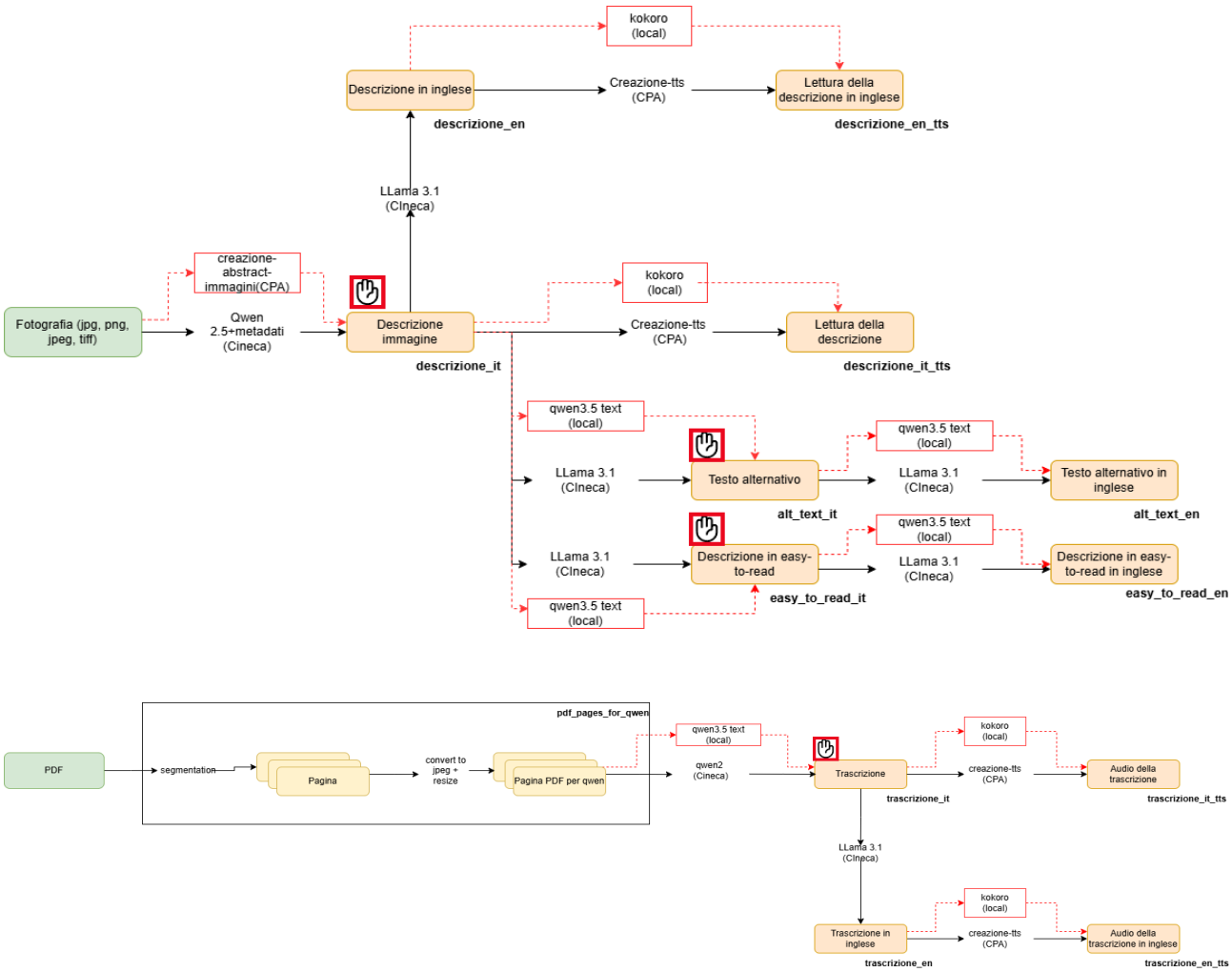
- VM;
- JupyterLab;
- LPW_DataPreparation;
- OpenSeaDragon;
- NiFi;
- Postgres;
- Elastic Search.

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
- Testo Alternativo in Italiano ed in inglese
- Descrizione in formato easy to read

INPUT: *Pdf*

OUTPUT:

- Trascrizione del pdf
- Traduzione in inglese del pdf
- Audio trascrizione in Italiano e in inglese

La descrizione approfondita delle pipeline viene riportata di seguito, nel paragrafo 4.

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 contenuti accessibili per immagini fotografiche

La pipeline è progettata per trasformare immagini digitali del patrimonio culturale in una serie di contenuti accessibili destinati a supportare differenti modalità di fruizione. A partire da fotografie nei formati **JPG, PNG, JPEG e TIFF**, il sistema genera automaticamente descrizioni testuali, audiodescrizioni, testi alternativi per tecnologie assistive e versioni semplificate dei contenuti. Tutti gli artefatti sono prodotti sia in lingua italiana sia in lingua inglese e vengono memorizzati nel Data Product per essere successivamente resi disponibili attraverso il portale di consultazione. Gli elementi contrassegnati dall'icona della mano possono essere modificati e validati dagli utenti autorizzati direttamente dal portale.

4.1.1. Analisi dell'immagine e generazione della descrizione principale

La pipeline prende in ingresso una fotografia digitale e la sottopone a una fase di analisi automatica mediante il modello multimodale **Qwen 2.5**, eseguito presso l'infrastruttura **CINECA**. Durante questa fase il modello elabora sia il contenuto visivo dell'immagine sia i metadati associati, identificando soggetti, oggetti, contesto, elementi artistici e relazioni presenti nella scena.

A supporto di questa elaborazione viene utilizzato il servizio **Creazione Abstract Immagini (CPA)**, che contribuisce alla strutturazione delle informazioni rilevate e alla generazione di una descrizione organica e coerente.

L'output prodotto è una **descrizione dell'immagine in italiano** (*descrizione_it*), che rappresenta il principale artefatto della pipeline e costituisce il punto di partenza per tutte le elaborazioni successive.

Questo contenuto può essere revisionato e modificato dagli operatori attraverso il portale per correggere eventuali imprecisioni, integrare dettagli storico-artistici o migliorare la qualità della descrizione.

4.1.2. Traduzione della descrizione in lingua inglese

La descrizione italiana viene successivamente elaborata tramite il modello **Llama 3.1**, eseguito presso l'infrastruttura CINECA, per produrre la corrispondente **descrizione in lingua inglese** (*descrizione_en*).

La traduzione mantiene il contenuto informativo della descrizione originale e consente la consultazione del patrimonio culturale anche da parte di utenti internazionali.

4.1.3. Produzione delle audiodescrizioni

Le descrizioni testuali vengono convertite in contenuti audio mediante il servizio **Creazione TTS (CPA)** che utilizza il motore di sintesi vocale **Kokoro**, eseguito localmente.

A partire dalla descrizione italiana viene generato l'artefatto:

- **descrizione_it_tts**: lettura audio della descrizione in italiano.

Analogamente, dalla descrizione inglese viene prodotto:

- **descrizione_en_tts**: lettura audio della descrizione in inglese.

Le audiodescrizioni consentono una fruizione alternativa dei contenuti da parte di utenti che preferiscono o necessitano di una modalità di accesso basata sull'ascolto.

4.1.4. Generazione del testo alternativo

A partire dalla descrizione principale, la pipeline genera un **testo alternativo** (*alt_text_it*) utilizzando una combinazione dei modelli **Llama 3.1** e **Qwen 3.5 Text**.

L'obiettivo del testo alternativo è fornire una descrizione breve, sintetica e compatibile con gli standard di accessibilità digitale, in particolare per l'utilizzo da parte di screen reader e tecnologie assistive.

Poiché la qualità di questi contenuti è essenziale per garantire un'efficace accessibilità, il sistema consente agli operatori autorizzati di modificare il testo alternativo direttamente dal portale.

Successivamente il modello **Llama 3.1**, supportato da **Qwen 3.5 Text**, genera la corrispondente versione inglese:

- **alt_text_en**: testo alternativo in lingua inglese.
- **Generazione delle descrizioni Easy-to-Read**

Per favorire la comprensione dei contenuti da parte di un pubblico più ampio, la pipeline produce una versione semplificata della descrizione denominata **Easy-to-Read**.

Attraverso l'utilizzo combinato di **Llama 3.1** e **Qwen 3.5 Text**, viene generato l'artefatto:

- **easy_to_read_it**: descrizione semplificata in italiano.

Questa versione adotta un linguaggio più semplice, una sintassi meno complessa e una struttura più immediata, facilitando la comprensione anche da parte di utenti che possono beneficiare di contenuti a ridotta complessità linguistica.

L'artefatto è modificabile dal portale e può essere revisionato dagli operatori per migliorarne chiarezza e leggibilità.

A partire dalla versione italiana viene quindi generata la corrispondente traduzione:

- **easy_to_read_en**: descrizione semplificata in inglese.

4.1.5. Revisione e validazione dei contenuti

La pipeline prevede un processo di supervisione umana sugli artefatti più rilevanti. In particolare, possono essere modificati e validati:

- la descrizione principale dell'immagine (*descrizione_it*);
- il testo alternativo (*alt_text_it*);
- la descrizione Easy-to-Read (*easy_to_read_it*).

Questo meccanismo permette di combinare le capacità di generazione automatica dell'Intelligenza Artificiale con la verifica da parte di operatori esperti, garantendo elevati livelli di qualità, accuratezza e aderenza alle esigenze di accessibilità.

4.1.6. Pubblicazione degli artefatti

Al termine dell'elaborazione vengono resi disponibili attraverso il Data Product i seguenti contenuti:

- Descrizione dell'immagine in italiano (*descrizione_it*);
- Descrizione dell'immagine in inglese (*descrizione_en*);
- Audiodescrizione in italiano (*descrizione_it_tts*);
- Audiodescrizione in inglese (*descrizione_en_tts*);
- Testo alternativo in italiano (*alt_text_it*);
- Testo alternativo in inglese (*alt_text_en*);
- Descrizione Easy-to-Read in italiano (*easy_to_read_it*);
- Descrizione Easy-to-Read in inglese (*easy_to_read_en*).

L'insieme di questi artefatti consente di offrire diverse modalità di accesso allo stesso contenuto culturale, favorendo una fruizione più inclusiva e adattabile alle differenti esigenze informative e di accessibilità degli utenti.

4.2. Pipeline di estrazione, traduzione e generazione audio per documenti PDF

La pipeline è progettata per trasformare documenti PDF in contenuti testuali e audio accessibili, consentendo la consultazione delle informazioni sia in forma scritta che mediante ascolto. Il processo combina tecniche di segmentazione documentale, riconoscimento ottico del testo, traduzione automatica e sintesi vocale per produrre trascrizioni multilingua e relative versioni audio. Tutti gli artefatti generati vengono archiviati all'interno del Data Product e resi disponibili tramite il portale di consultazione. Gli elementi contrassegnati con l'icona della mano possono essere revisionati e modificati dagli operatori autorizzati prima della pubblicazione.

4.2.1. Acquisizione e segmentazione del documento PDF

La pipeline prende in ingresso un documento in formato **PDF**. Nella fase iniziale il documento viene suddiviso nelle singole pagine che lo compongono attraverso un processo di segmentazione.

Successivamente ogni pagina viene convertita in formato immagine e opportunamente ridimensionata per ottimizzare le prestazioni dei modelli di Intelligenza Artificiale utilizzati nelle fasi successive. Il risultato è un insieme di rappresentazioni grafiche delle pagine del documento, predisposte per l'estrazione automatica dei contenuti testuali.

4.2.2. Estrazione del testo dalle pagine

Le immagini generate vengono elaborate mediante il modello **Qwen 2**, eseguito presso l'infrastruttura **CINECA**, supportato dal modello **Qwen 3.5 Text** eseguito localmente.

L'obiettivo di questa fase è riconoscere il contenuto testuale presente all'interno delle pagine e produrre una trascrizione digitale completa del documento.

Il testo individuato nelle diverse pagine viene progressivamente ricostruito e organizzato fino a ottenere una **trascrizione integrale del documento in lingua italiana** (*trascrizione_it*).

Poiché la qualità della trascrizione è fondamentale per le successive elaborazioni e per la corretta consultazione dei contenuti, il sistema consente agli operatori autorizzati di modificare e validare il testo direttamente dal portale. Questo permette di correggere eventuali errori di riconoscimento, migliorare la formattazione e garantire una maggiore accuratezza del risultato finale.

4.2.3. Generazione della traduzione in lingua inglese

La trascrizione italiana validata viene successivamente elaborata tramite il modello **Llama 3.1**, eseguito presso l'infrastruttura CINECA, per produrre una **trascrizione in lingua inglese** (*trascrizione_en*).

Questa fase consente di rendere accessibile il contenuto documentale anche a utenti internazionali, mantenendo la coerenza semantica con il documento originale e preservando il significato delle informazioni contenute nel testo.

4.2.4. Produzione dell'audio della trascrizione in italiano

La trascrizione italiana viene convertita in un contenuto audio attraverso il servizio **Creazione TTS (CPA)** che utilizza il motore di sintesi vocale **Kokoro**, eseguito localmente.

Il risultato di questo processo è l'artefatto:

- **trascrizione_it_tts**: lettura audio della trascrizione italiana.

Questa funzionalità consente agli utenti di accedere alle informazioni del documento anche attraverso modalità di ascolto, migliorandone la fruibilità e l'accessibilità.

4.2.5. Produzione dell'audio della trascrizione in inglese

Analogamente, la trascrizione inglese viene sottoposta allo stesso processo di sintesi vocale mediante il servizio CPAI Text-to-Speech basato sul modello **Kokoro**.

Il risultato è il seguente artefatto:

- **trascrizione_en_tts**: lettura audio della trascrizione inglese.

La disponibilità della versione audio in lingua inglese amplia ulteriormente le possibilità di accesso al contenuto documentale da parte di utenti internazionali.

4.2.6. Revisione e validazione degli artefatti

La pipeline integra un meccanismo di supervisione umana sulla trascrizione estratta dal documento. L'artefatto **trascrizione_it** può essere modificato e validato tramite il portale da utenti dotati di specifici privilegi amministrativi o redazionali.

Questa attività di controllo qualitativo consente di combinare l'efficienza dell'estrazione automatica con la verifica umana, garantendo risultati più accurati e affidabili, soprattutto nel caso di documenti complessi, scansioni di bassa qualità o contenuti specialistici.

4.2.7. Pubblicazione degli artefatti

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

- Trascrizione del documento in italiano (*trascrizione_it*);

- Traduzione del documento in inglese (*trascrizione_en*);
- Audio della trascrizione italiana (*trascrizione_it_tts*);
- Audio della trascrizione inglese (*trascrizione_en_tts*).

Attraverso questa pipeline il documento PDF viene trasformato in un insieme di contenuti multilingua e multimodali, consentendo differenti modalità di accesso all'informazione e contribuendo a migliorare sensibilmente l'accessibilità e l'inclusione nella consultazione del patrimonio documentale digitale.

5. Esposizione e distribuzione

Il Data Product "Una passeggiata nella Lendinara di fine Ottocento" viene pubblicato nel catalogo della piattaforma DPaaS e reso fruibile attraverso un portale web dedicato e mediante API REST in sola lettura.

Il Data Product consente la consultazione strutturata delle fotografie, delle cartoline e dei documenti storici conferiti dal Comune di Lendinara, 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 portale web, destinata agli utenti finali e agli operatori autorizzati;
- fruizione machine-to-machine tramite API, destinata al portale, al Backend for Frontend e agli altri sistemi consumer autorizzati.

Il portale rappresenta il principale canale di consultazione per gli utenti finali. Consente di esplorare il patrimonio per oggetto, tema, macroarea, luogo e tipologia di contenuto e di accedere alle immagini storiche, alle informazioni descrittive, alle trascrizioni e alle versioni accessibili e multilingua prodotte dal Data Product.

La read-only API costituisce il canale di interoperabilità per il portale e per gli altri sistemi applicativi. Espone gli oggetti e i relativi metadati, le classificazioni, le geolocalizzazioni e le relazioni con gli artefatti accessibili generati. Le operazioni disponibili sono esclusivamente di consultazione e non consentono ai consumer di modificare direttamente i dati esposti.

5.1. Canali di accesso

I canali di accesso previsti sono:

- portale web pubblico dedicato a Lendinara;
- API REST 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 accesso;
- eventuale snapshot del database in formato JSON, prodotto per esigenze di trasferimento o fruizione offline.

La parte pubblica del portale permette la consultazione dei contenuti senza autenticazione. Le funzionalità riservate, in particolare quelle destinate alla gestione, revisione e validazione dei contenuti, richiedono autenticazione e un profilo autorizzativo adeguato.

5.2. Utenti e sistemi destinatari

Il Data Product è destinato principalmente a:

- cittadini e visitatori interessati al patrimonio storico e culturale di Lendinara;
- persone con disabilità visive o cognitive;
- utenti che beneficiano di contenuti semplificati o di modalità alternative di consultazione;
- utenti internazionali, grazie alla disponibilità di contenuti in lingua inglese;
- operatori del Comune, della biblioteca e degli archivi;
- operatori incaricati della revisione e validazione degli artefatti;
- portale Lendinara e relativo Backend for Frontend;
- 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.

5.3. Autenticazione e autorizzazione

La consultazione dei contenuti pubblici attraverso il portale non richiede autenticazione.

Le funzionalità riservate agli operatori, quali revisione, modifica e validazione degli artefatti generati, sono accessibili tramite autenticazione SPID/CIE e sulla base dei ruoli applicativi assegnati.

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

- la descrizione principale dell'immagine in italiano;
- il testo alternativo in italiano;
- la descrizione Easy-to-Read in italiano;
- la trascrizione italiana dei documenti PDF.

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

Per l'accesso machine-to-machine ai servizi del Data Product, le API sono protette mediante una chiave identificativa dell'utente o dell'applicazione e il codice identificativo del Data Product.

Le credenziali tecniche necessarie per l'accesso al database PostgreSQL e allo storage S3 sono utilizzate esclusivamente dai componenti backend e non sono esposte ai consumer.

Le modalità definitive di autenticazione applicativa, la gestione delle chiavi, la loro durata, la rotazione e le eventuali restrizioni di rete sono definite nell'ambiente di esercizio e configurate sul layer infrastrutturale di esposizione.

5.4.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 JPG, JPEG, PNG, TIF e TIFF;
- documenti in formato PDF;
- descrizioni testuali;
- trascrizioni dei documenti;
- contenuti testuali in italiano e inglese;
- testi alternativi per tecnologie assistive;
- descrizioni in formato Easy-to-Read;
- audiodescrizioni e letture audio generate tramite sintesi vocale;
- metadati tecnici e descrittivi strutturati;
- informazioni relative a temi, macroaree, cataloghi, esposizioni, tag, luoghi e geolocalizzazioni.

Le risorse binarie, quali immagini e tracce audio, sono conservate in uno storage S3 privato. I client non accedono direttamente al bucket. L'applicazione recupera le risorse dallo storage e le restituisce al consumer in streaming attraverso endpoint dedicati.

5.5.Modalità di aggiornamento

L'aggiornamento del Data Product avviene attraverso pipeline orchestrate tramite Apache NiFi. Le pipeline monitorano periodicamente le cartelle dello storage associate all'Ente, identificano i nuovi contenuti o i contenuti non ancora elaborati e attivano automaticamente i task necessari.

Lo stato di ogni elaborazione viene registrato nel database PostgreSQL. Per ciascuna esecuzione vengono tracciati:

- identificativo del job;
- task eseguiti;
- stato di avanzamento;
- esito delle singole elaborazioni;
- eventuali errori;
- informazioni necessarie al monitoraggio e alla diagnosi.

Gli artefatti prodotti vengono associati all'oggetto sorgente e resi disponibili al completamento delle elaborazioni previste.

Le pipeline dispongono di meccanismi di retry automatico per la gestione degli errori temporanei. È inoltre prevista una modalità di ripresa manuale che permette di correggere uno specifico artefatto sorgente e rigenerare a cascata gli output dipendenti, evitando la riesecuzione completa dell'intera pipeline.

Quando l'aggiornamento riguarda esclusivamente i dati o gli artefatti, senza modificare il modello dati, la struttura delle API o la pipeline applicativa, non è necessario introdurre una nuova versione del Data Product.

Un nuovo versionamento è invece richiesto in caso di modifiche significative riguardanti:

- modello o struttura dei dati;
- contratti delle API;
- pipeline di elaborazione;
- logiche funzionali dell'applicazione;
- modalità di esposizione del Data Product.

5.6. Limiti di utilizzo

Le funzioni di ricerca adottano un sistema di paginazione basato sul numero di pagina e sul numero di risultati restituiti per pagina. L'utilizzo della paginazione consente di limitare la dimensione delle risposte e di gestire in modo efficiente la consultazione di insiemi numerosi di oggetti.

Gli eventuali limiti relativi a numero di richieste per unità di tempo, richieste concorrenti, dimensione massima delle risposte e della pagina, timeout e soglie di throttling sono definiti a livello di ambiente di esercizio e applicati tramite API Gateway, ingress controller o reverse proxy.

5.7. Livelli di servizio

Il servizio dispone di endpoint tecnici dedicati al monitoraggio dello stato applicativo e alla verifica della disponibilità del container.

Sono previsti:

- un endpoint di readiness, utilizzato per verificare che il servizio sia pronto a ricevere richieste;
- un endpoint di liveness, utilizzato per verificare che il processo applicativo sia attivo.

In ambiente containerizzato, tali endpoint possono essere utilizzati dalle probe Kubernetes per gestire automaticamente il bilanciamento del traffico e il riavvio del servizio in caso di indisponibilità.

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

5.8. 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 OpenAPI;
- interfaccia Swagger per la consultazione e il test delle API.

La documentazione pubblica descrive a livello funzionale le modalità di fruizione del Data Product. La documentazione tecnica di dettaglio, comprendente endpoint, parametri, modelli dati, formati delle risposte ed esempi di chiamata, è resa disponibile secondo le modalità previste dal catalogo DPaaS e dal processo di richiesta di consumo.

5.9. 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 singole immagini, ai documenti storici, ai fondi fotografici e archivistici, agli asset messi a disposizione dal Comune di Lendinara e agli artefatti derivati prodotti dalle elaborazioni automatiche.

Le condizioni di riutilizzo devono specificare almeno:

- possibilità di copia e redistribuzione;
- possibilità di modifica o elaborazione;
- eventuali obblighi di attribuzione;
- eventuali restrizioni all'utilizzo commerciale;
- modalità di citazione della fonte;
- eventuali vincoli derivanti da diritti di terzi.

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

5.10. Fruizione tramite API

L'API di Lendinara fornisce accesso programmatico, in sola lettura, agli oggetti culturali del Data Product e ai relativi contenuti accessibili. Consente alle applicazioni consumer di ricercare, recuperare e navigare fotografie, cartoline e documenti storici, insieme ai metadati descrittivi e tecnici, alle informazioni tematiche e geografiche e agli artefatti generati dalle pipeline di accessibilità.

5.10.1. Tipologie di informazioni interrogabili

Le principali tipologie di informazioni interrogabili comprendono:

- oggetti digitali e relativi identificativi;
- fotografie, cartoline e documenti storici;
- riferimenti ai file e agli asset associati;
- titolo, descrizione, autore, creator o fondo di provenienza;
- datazione e copertura temporale;
- tipologia, sottotipo e stato del ciclo di vita;
- lingua, metadati descrittivi e metadati tecnici;
- temi, macroaree, tag e parole chiave;
- cataloghi, esposizioni, luoghi, punti di interesse e geolocalizzazioni;
- relazioni tra oggetti e tra oggetti originali e artefatti derivati;
- descrizioni, alt-text, contenuti Easy-to-Read e trascrizioni in italiano e inglese;
- audiodescrizioni e tracce audio.

5.10.2. Principali operazioni disponibili

Le principali operazioni messe a disposizione dall'API sono:

- ricerca avanzata degli oggetti digitali;
- recupero del dettaglio di un singolo oggetto;
- recupero dei metadati completi di 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 immagini e delle risorse audio.

Le operazioni di consultazione non modificano il contenuto del Data Product. Le funzioni di revisione e validazione sono gestite dal portale e dai servizi applicativi riservati agli operatori autorizzati e non fanno parte dell'interfaccia pubblica read-only.

5.10.3. Ricerca, filtro, ordinamento e paginazione

L'API mette a disposizione una funzione di ricerca avanzata che consente di combinare più criteri. La ricerca può essere filtrata in base a:

- codice dell'oggetto;
- titolo;
- genere;
- autore o creator;
- data di caricamento;
- lingua;
- stato;
- tipologia e sottotipo;
- catalogo ed esposizione;
- tema o macroarea;
- luogo o geolocalizzazione;
- presenza di un oggetto parent;
- uno o più tag.

I tag possono essere combinati secondo logica OR, quando è sufficiente la corrispondenza con almeno uno dei tag indicati, oppure secondo logica AND, quando è richiesta la corrispondenza con tutti i tag indicati.

Per la specificità del patrimonio di Lendinara assumono particolare rilevanza i criteri relativi al soggetto rappresentato, all'autore o fondo fotografico, alla macroarea tematica, al periodo storico, al luogo o punto di interesse, alla tipologia del contenuto, alla lingua e alla disponibilità degli artefatti accessibili.

I risultati possono essere ordinati in modalità ascendente o discendente rispetto ai campi supportati. La paginazione viene gestita attraverso il numero della pagina richiesta e il numero di risultati da restituire per pagina.

5.10.4. Recupero del dettaglio di una risorsa

Il dettaglio di una risorsa viene recuperato attraverso il relativo identificativo univoco. La risposta contiene le informazioni principali dell'oggetto, tra cui identificativo, codice,

classificazione, tipologia e sottotipo, stato, riferimento al file, catalogo, esposizione e informazioni geografiche.

I metadati completi sono disponibili tramite una specifica operazione di consultazione e possono comprendere titolo, genere, autore o creator, datazione, descrizione, note, lingua, metadati tecnici e descrittivi, parole chiave, informazioni storiche, caratteristiche dell'oggetto, temi, luoghi, geolocalizzazione e artefatti di accessibilità disponibili.

Le relazioni della risorsa consentono di individuare oggetti figli o componenti, oggetti appartenenti alla stessa aggregazione, oggetti culturalmente o strutturalmente collegati e artefatti generati a partire dalla risorsa originale.

5.10.5. Servizi di aggregazione e analisi

L'API mette a disposizione servizi di aggregazione per cataloghi, esposizioni e geolocalizzazioni. Gli aggregati possono includere il conteggio degli oggetti associati a ciascun elemento, consentendo al portale di costruire modalità di navigazione per collezione, esposizione o luogo.

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

5.10.6. Formati delle richieste e delle risposte

Le richieste utilizzano il protocollo HTTPS. Le operazioni di consultazione sono effettuate principalmente tramite metodo HTTP GET e utilizzano path parameter per identificare una specifica risorsa e query parameter per definire filtri, lingua, ordinamento e paginazione.

Le risposte sono restituite in formato JSON con media type application/json. I contenuti binari, quali immagini e tracce audio, vengono restituiti in streaming con il media type corrispondente al formato della risorsa. La struttura delle richieste, delle risposte e dei modelli informativi è descritta nella specifica OpenAPI.

5.10.7. Modalità di autenticazione

L'accesso machine-to-machine alle API avviene mediante una chiave identificativa dell'utente o dell'applicazione e il codice del Data Product. La chiave permette di identificare il consumer e di applicare le regole di accesso previste dall'ambiente di esercizio.

Le credenziali per l'accesso al database e allo storage non sono trasferite ai consumer e rimangono confinate nei componenti backend. Le funzionalità riservate agli operatori richiedono autenticazione SPID/CIE e un ruolo applicativo autorizzato.

5.10.8. Limiti di chiamata

L'API utilizza la paginazione per controllare la quantità di dati restituita nelle operazioni di ricerca. Gli eventuali limiti relativi a frequenza delle chiamate, quota giornaliera o oraria, chiamate concorrenti, dimensione massima della pagina, timeout e dimensione massima delle risposte sono definiti nella configurazione dell'ambiente di esercizio e applicati dal layer infrastrutturale di esposizione.

5.10.9. Gestione degli errori

L'API utilizza codici di stato HTTP standard per comunicare l'esito delle richieste. Le principali categorie di errore riguardano:

- parametri mancanti o non validi;
- formato non corretto di un parametro;
- identificativo della risorsa non valido o risorsa non esistente;
- autenticazione assente o non valida;
- accesso non autorizzato;
- superamento degli eventuali limiti di utilizzo;
- indisponibilità temporanea del database o dello storage;
- errore interno durante l'elaborazione della richiesta.

Le risposte di errore sono restituite in formato JSON e contengono le informazioni necessarie per identificare la tipologia e la causa del problema. Gli errori di validazione dei parametri vengono distinti dagli errori applicativi o infrastrutturali.

Gli endpoint tecnici di health check e readiness consentono ai sistemi di monitoraggio di verificare separatamente che il processo applicativo sia attivo e che il servizio sia effettivamente pronto a gestire le richieste.

Le pipeline di alimentazione gestiscono separatamente gli errori di elaborazione. Lo stato del job, lo step interessato, il codice di risposta e il messaggio di errore vengono registrati per consentire il monitoraggio, il retry automatico o la ripresa manuale dell'elaborazione.

5.11. Limitazioni

Il Data Product Lendinara è costituito da un insieme circoscritto di oggetti digitali e metadati conferiti dal Comune di Lendinara. Non rappresenta quindi l'intero patrimonio storico e culturale del territorio 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 della revisione di file e metadati;
- **mediante esecuzioni schedulate delle pipeline**, secondo la configurazione dell'ambiente;

tramite eventuali elaborazioni di arricchimento successive al caricamento.

Di conseguenza, tra la modifica della fonte e la pubblicazione del dato aggiornato può intercorrere un intervallo legato alla frequenza di esecuzione e al completamento delle pipeline.

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.