15 ottobre 2026

Finanziato dall'Unione europea NextGenerationEU

MINISTERO DELLA CULTURA

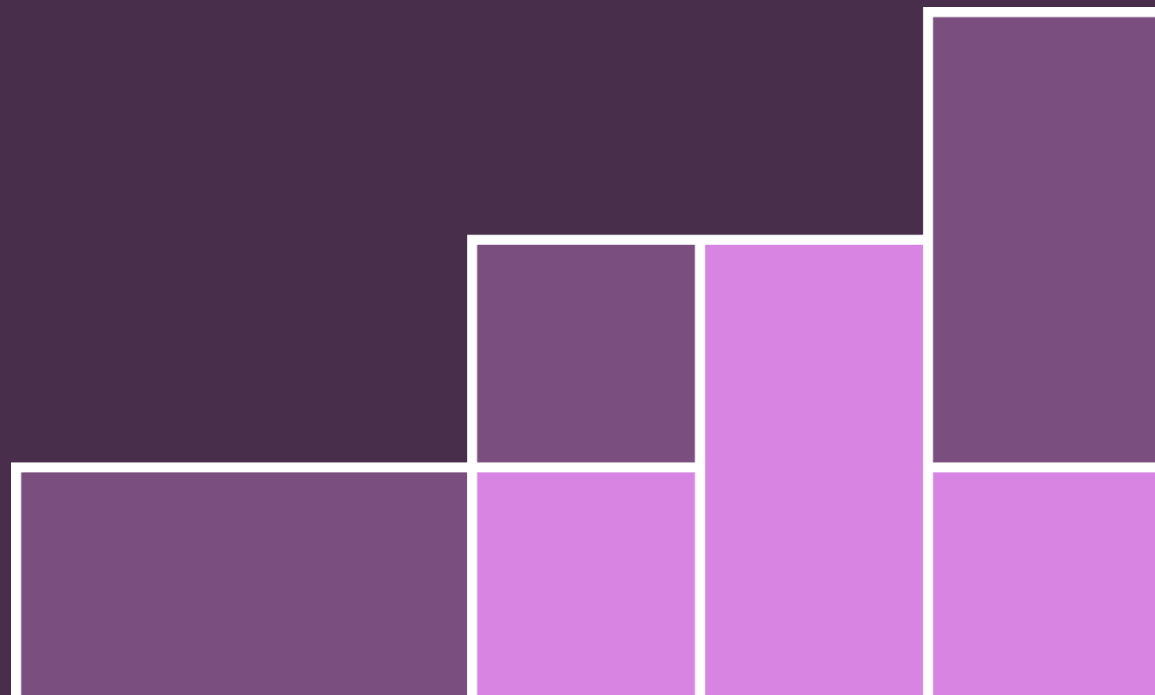
DIGITAL LIBRARY
Istituto centrale per la digitalizzazione del patrimonio culturale

Ecomic ecosistema digitale per la cultura

DPaaS Piattaforma data product as a service

In viaggio su ali di Carta

Archivio di Stato di Genova



Scopo del Data product

Rendere accessibile e fruibile un corpus di documenti archivistici medievali dell'Archivio di Stato di Genova attraverso schede digitali strutturate e arricchite, utilizzate anche per alimentare l'applicazione immersiva 3D per tablet e VR "In viaggio su ali di carta".

Le fonti

Materiale fornito dai responsabili dell'Archivio di Stato di Genova

- Corpus di 21 schede: documenti archivistici medievali con regesto, edizione bibliografica e testo divulgativo
- 19 immagini a corredo (riproduzioni dei documenti originali)
- Cronologia: dal diploma del 958 sulla nascita del Comune ad un documento del 1462
- Categorie principali: vita quotidiana e istruzione, diritto e istituzioni, commercio e mercatura, marineria e spedizioni, diplomazia con l'Oriente, eventi politici e militari genovesi
- Apparato critico: regesto archivistico ed edizione bibliografica per ciascun documento
- 4 schede su 21 prive di immagini a corredo

Ingestione e Normalizzazione

- Schede catalografiche: titolo, descrizione, metadati, immagini
- Indicizzazione da manifest json delle schede
- Descrizioni segmentate in blocchi testuali (allineati con le traduzioni e l'audio)
- Blocco extra strutturato per ciascun documento: regesto, edizione, testo divulgativo
- Immagini normalizzate per l'inferenza (max 1000×1000 px)
- Il testo italiano originale resta la fonte originale

Generazione dei Contenuti arricchiti:

Testo (Qwen2, pipeline cloud):

- Traduzione di titolo e descrizione → EN, FR, DE, ES
- Traduzione dei metadati strutturati (regesto, edizione, testo divulgativo)

Immagini (Qwen2 multimodale):

- Descrizione visiva delle immagini dei documenti

Audio:

- Kokoro TTS: sintesi vocale in italiano, per blocco testuale

Ciclo di Vita e Approvazione

Ogni contenuto generato automaticamente (una traduzione, una descrizione, un audio) passa attraverso un controllo con Governance Human-in-the-Loop prima di essere pubblicato: prima viene creato dal sistema, poi controllato, e solo dopo l'ok di una persona diventa visibile al pubblico.

- Niente viene pubblicato in automatico dopo l'elaborazione: **serve sempre l'approvazione di un operatore**
- Il controllo è puntuale: ad esempio si può pubblicare il titolo e le foto di una scheda anche se l'audio non è ancora stato revisionato

Il Dataset in Numeri

- **21 schede** (documenti dal 958 al 1462)
- **19 immagini**
- **6 lingue supportate** (IT + traduzioni EN/FR/DE/ES)
- **20 testi divulgativi**

Finanziato dall'Unione europea
NextGenerationEU

MINISTERO DELLA CULTURA

DIGITAL LIBRARY
Istituto centrale per la digitalizzazione del patrimonio culturale

Ecomic
ecosistema digitale per la cultura

DPaaS
Piattaforma data product as a service

Grazie

