Libreria di profili ceramici (Sigillata Italica) per il restauro digitale dei reperti

Documento di specifiche tecniche di elaborazione

Sommario

1. Obiettivi del Data Product 3
2. Origine dei dati ed eventuali elaborazioni 5
3. Tecnologie utilizzate all'interno del Data Product Lab 6
4. Pipeline di elaborazione e trasformazione 7
5. Esposizione e distribuzione 8
6. Limitazioni 9
   6.1. Nota sulle condivisioni dei dati 10
7. Possibili evoluzioni future 11

Indice delle figure

Figura 1 Dal disegno tecnico manuale al modello digitale 3
Figura 2 Pipeline di elaborazione dei dati 7

1.Obiettivi del Data Product

Il Data Product è stato sviluppato per DRM Toscana – Museo delle Navi di Pisa nell'ambito del progetto "Sistema integrato per il restauro digitale di ceramiche Sigillata Italica e stampa 3D" (Lotto 4 – PR 02). Il Museo conserva oltre 2.000 reperti di ceramica Sigillata Italica e ne gestisce il restauro attraverso strumenti digitali quali laser scanner, elaborazione 3D e stampa 3D; in questo contesto, il Data Product acquisisce i profili ceramici in formato vettoriale (DWG, SVG, PDF vettoriale, EPS), li converte automaticamente in formato DXF e ne verifica la congruenza geometrica rispetto ai file di partenza.

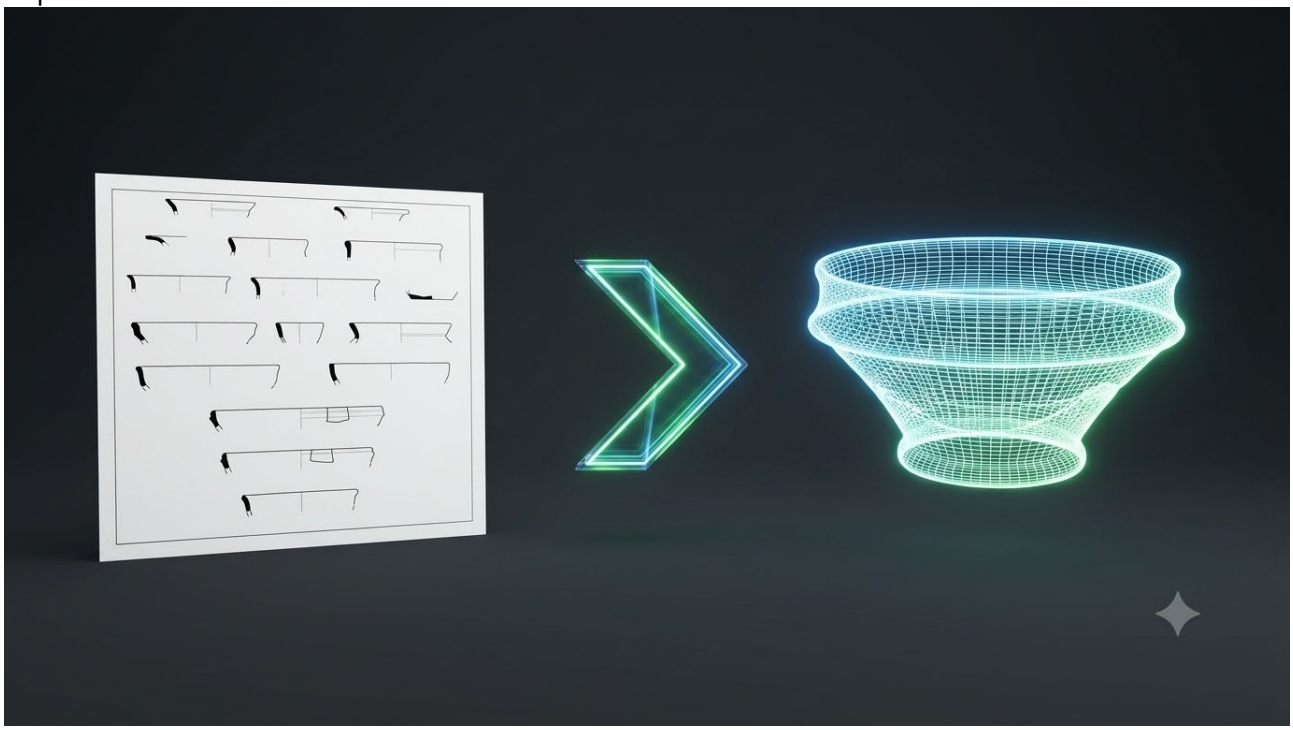


Figura 1 Dal disegno tecnico manuale al modello digitale

Gli output prodotti sono i profili ceramici convertiti in DXF standardizzato e interoperabile (il "Profilo Conspectus"), l'esito della verifica automatica di congruenza per ciascun profilo, un log/manifest dell'elaborazione con l'elenco dei file trattati, i relativi esiti ed eventuali warning, e infine un archivio ZIP con il dataset dei DXF prodotti, scaricabile dall'utente.

Il Data Product risponde a due casi d'uso principali.

- La standardizzazione e l'interoperabilità: i profili vettoriali eterogenei in ingresso vengono convertiti in DXF e validati per congruenza, così da ottenere profili standardizzati e riusabili all'interno della piattaforma.

- La campionatura di reperti privi di forme catalogate: in assenza di forme di riferimento standardizzate,come accade per ampie porzioni del Conspectus della Sigillata Italica, lo strumento consente di campionare i reperti a supporto della definizione di nuove classificazioni.

I profili utente target sono due: il Data Product è anzitutto rivolto al gruppo tecnico che gestisce la conversione e la verifica dei profili; si rivolge inoltre a restauratori e archeologi, che utilizzano i DXF tramite Ecomic per il restauro digitale e la stampa dei reperti, oltre ad accedervi dal catalogo per finalità di ricerca, consultazione e download.

2. Origine dei dati ed eventuali elaborazioni

Gli input del Data Product sono profili ceramici in formato vettoriale (DWG, SVG, PDF vettoriale, EPS), relativi ai reperti di ceramica Sigillata Italica del Museo delle Navi di Pisa. Tali profili vengono caricati sullo storage del DP Lab da personale tecnico o da un Data Engineer, che costituiscono quindi il punto di ingresso dei dati nella piattaforma.

Prima che il profilo entri nel flusso di elaborazione, viene effettuato un controllo sulla qualità del dato in ingresso: si verifica che il file sorgente sia in uno dei formati supportati (DWG, SVG, PDF vettoriale, EPS) e che risulti correttamente leggibile. Questo controllo rappresenta l'unica elaborazione svolta sui dati prima del loro ingresso nel DP Lab; le elaborazioni vere e proprie, ossia la conversione automatica in formato DXF e la verifica di congruenza geometrica rispetto all'input, avvengono invece all'interno del DP Lab.

3. Tecnologie utilizzate all'interno del Data Product Lab

La tecnologia adottata per la realizzazione del Data Product è Python, eseguito in un contenitore VM full-code del DP Lab. Python copre l'intero processo di elaborazione: viene utilizzato per l'acquisizione dei profili ceramici in formato vettoriale (DWG, SVG, PDF vettoriale, EPS), per la conversione automatica dei profili in formato DXF e per la verifica di congruenza geometrica tra l'output prodotto e il file di origine. Non sono quindi impiegati componenti applicativi o servizi distinti per le diverse fasi della pipeline: l'intera catena di elaborazione, dal caricamento del profilo alla produzione del Dataset File in output, è gestita da un'unica base tecnologica.

4. Pipeline di elaborazione e trasformazione

La pipeline di elaborazione di un lotto di profili si articola in tre fasi sequenziali, eseguite in un contenitore VM full-code del DP Lab.

Nella prima fase vengono caricati i profili ceramici in formato vettoriale (DWG, SVG, PDF vettoriale, EPS). Nella seconda fase, che costituisce il cuore della pipeline, i profili caricati vengono convertiti automaticamente in formato DXF standardizzato: la conversione è realizzata in Python. Nella terza fase viene verificata la congruenza geometrica del profilo DXF prodotto rispetto all'input di partenza.

Al termine delle tre fasi, la pipeline restituisce in output i profili ceramici in formato DXF (Profilo Conspectus), l'esito della verifica di congruenza per ciascun profilo e il relativo log; il tutto viene reso disponibile come Dataset File.

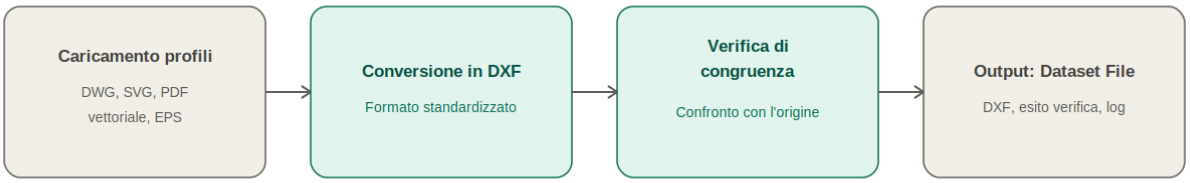


Figura 2 Pipeline di elaborazione dei dati

5. Esposizione e distribuzione

Il dato prodotto viene reso disponibile attraverso due canali di accesso, entrambi con accessibilità Public. Il primo è il download dal catalogo DPaaS: gli utenti autorizzati scaricano i profili in formato DXF, come singoli file o come archivio, direttamente dalla libreria pubblicata nel catalogo. Il secondo canale è una REST API: tramite un endpoint dedicato è possibile accedere alla libreria ceramica standardizzata, per il consumo da parte dei livelli 2 e 3 della piattaforma.

La libreria elaborata viene salvata nella sotto-cartella ./export/.
È importante delimitare il perimetro di questo Data Product rispetto ai livelli successivi: gli artefatti morfologici (bundle, conspectus, firme) e le interfacce di matching non appartengono a DP00000118, ma alla piattaforma CMS con Extension Blender di Livello 2, che li costruisce a partire dalla libreria DXF qui distribuita.

Fruizione tramite download
La fruizione tramite download si articola in tre passaggi: l'utente individua il Data Product DP00000118 nel catalogo DPaaS, scarica il Dataset File con i profili ceramici in formato DXF e importa i DXF nella piattaforma/extension di restauro digitale di Livello 2 per le lavorazioni a valle.

Fruizione tramite API
La fruizione tramite API si basa su tre entità principali. Il Lotto (o Job) rappresenta l'insieme dei profili elaborati in una stessa esecuzione. Il Profilo (o File) rappresenta la coppia formata dal sorgente vettoriale in input e dal relativo DXF in output, con l'esito della verifica di congruenza, lo stato e i percorsi associati. Il Task rappresenta la singola esecuzione del tool, con le relative statistiche (input, convertiti, saltati), i tempi e lo stato. Attraverso l'API è ad esempio possibile recuperare la lista paginata delle librerie elaborate, oppure i dati di una singola libreria.

6. Limitazioni

Il Data Product presenta alcuni limiti tecnici e di ambito che è opportuno tenere presenti nell'utilizzo.

Sul piano dell'input, lo strumento richiede che il profilo sia già disponibile in formato vettoriale (DWG, SVG, PDF vettoriale, EPS); non è prevista la vettorializzazione a partire da immagini raster.

Sul piano dell'accuratezza, la fedeltà della conversione in DXF è oggetto di una verifica automatica di congruenza rispetto al file di origine: i profili che risultano non conformi vengono segnalati, ma restano comunque un limite intrinseco al processo di conversione automatica.

Infine, la metadatazione dei dataset sorgente può non essere disponibile per tutti i profili trattati.

6.1. Nota sulle condivisioni dei dati

I dati prodotti dal Data Product — i profili ceramici in formato DXF standardizzato (Profilo Conspectus), i relativi esiti di verifica di congruenza e i log/manifest di elaborazione — sono resi disponibili con accessibilità Public, sia tramite download dal catalogo DPaaS sia tramite REST API.

Non sono trattati dati personali, riservati o classificati: i profili ceramici rappresentano esclusivamente la geometria vettoriale dei reperti archeologici, privi di riferimenti a persone fisiche. Non si rendono pertanto necessarie attività di anonimizzazione o pseudonimizzazione.

L'accesso ai dati è aperto a due categorie di utenti: il gruppo tecnico che gestisce la conversione e la verifica dei profili, e i profili a valle (restauratori, archeologi, altri utenti del catalogo) che consultano o scaricano i DXF per finalità di ricerca, restauro digitale e stampa 3D tramite Ecomic. L'accesso in scrittura, ossia il caricamento dei profili sorgente, resta riservato al personale tecnico e ai Data Engineer.

I dati sono di titolarità di DRM Toscana – Museo delle Navi di Pisa, in quanto ente conservatore dei reperti da cui i profili sono derivati. La libreria DXF può essere riutilizzata dai livelli successivi della piattaforma, in particolare dalla piattaforma CMS e con Extension Blender di Livello 2, per la costruzione degli artefatti morfologici, restando inteso che tali elaborazioni non rientrano nel perimetro del DP00000118.

7. Possibili evoluzioni future

Sulla base delle esigenze emerse durante la realizzazione e l'utilizzo del Data Product, sono state individuate alcune possibili direttrici di sviluppo:

- estensione dei formati di input supportati, ad esempio l'introduzione della vettorializzazione automatica a partire da immagini raster (scansioni, fotografie dei disegni tecnici), oggi esclusa;
- estensione ad altre classi tipologiche o collezioni, oltre alla Sigillata Italica, per supportare la campionatura e la standardizzazione di ulteriori reperti ceramici del Museo o di altre istituzioni della rete DRM Toscana;
- affinamento della verifica di congruenza geometrica, con l'introduzione di metriche più granulari (ad es. deviazione locale del profilo) e soglie di tolleranza configurabili per tipologia di reperto;
- automazione end-to-end, per ridurre l'intervento manuale nel caricamento e nella validazione preliminare dei profili sorgente;

Per ciascuna direttrice, in una fase successiva di pianificazione, verrà specificato il beneficio atteso, i principali destinatari, le dipendenze tecniche od organizzative, gli eventuali prerequisiti, il livello di priorità e una possibile collocazione temporale.