

Uso e riuso di dati per la valorizzazione del patrimonio culturale italiano online: buone pratiche per l'interoperabilità

Chiara Aiola¹, Giulia Fabbris²

¹Net7, Italia – aiola@netseven.it

²Università Ca' Foscari di Venezia, Italia – giulia.fabbris@unive.it

ABSTRACT

Con questo poster si intende presentare una riflessione sulla necessità di seguire standard e buone pratiche per lo sviluppo di progetti e l'inserimento di oggetti nell'ecosistema digitale al fine di creare risorse interoperabili in un'ottica collaborativa e di riuso.

PAROLE CHIAVE

Interoperabilità – riuso di dati – patrimonio culturale digitale – piano nazionale di digitalizzazione – principi FAIR

1. INTRODUZIONE

Nonostante la fama che ormai vantano a livello internazionale, i principi [FAIR](#) non sembrano ancora alla base di molti progetti digitali. La scarsa osservanza di standard e linee guida, l'utilizzo di formati diversi e il limitato libero accesso alle risorse digitali prodotte sono grossi impedimenti per lo sviluppo di un vero e proprio ecosistema digitale del patrimonio culturale. Certamente esistono standard informatici così come linee guida per la codifica di oggetti digitali (TEI, DCMI), ma manca la volontà da parte di molti studiosi di integrare questi strumenti o “buone pratiche” nei loro progetti. È infatti necessario che le risorse prodotte siano interoperabili, poiché “[l]’ambiente digitale trova la propria essenza costitutiva nelle relazioni, ovvero nella possibilità di generare e rigenerare connessioni reciproche tra le informazioni, facilitando la produzione di nuovi significati” (PND:17). Al contrario, ogni progetto costituirebbe un’entità a sé stante senza possibilità di integrazione e di riuso.

Anche la mancanza di un’infrastruttura informatica specificamente dedicata alla conservazione dei dati nell’ambito delle Digital Humanities spesso costituisce un ostacolo al reperimento dei dati. Si assiste spesso a un proliferare di progetti e piattaforme legati a specifici progetti che non espongono API per la condivisione dei dati (o se esistono non sono documentate), oppure si sceglie di utilizzare strumenti che non nascono propriamente per la conservazione di dati ma che consentono lo storage gratuito e persistente del materiale caricato, come repository su GitHub. Questa scelta da un lato tutela i ricercatori dal rischio di perdere il proprio lavoro e dall’impossibilità di sostenere costi di hosting, consente il libero accesso alle risorse, anche attraverso il sistema di API fornite dalla piattaforma, ma dall’altro non permette di avere strumenti specifici per strutturare i metadati o i file del progetto, rendendo quindi problematico un processo di automatizzazione dell’interpretazione dei dati (cf. sotto la creazione di un’infrastruttura del patrimonio culturale promossa dal Ministero della cultura nell’ambito del [Piano Nazionale di Digitalizzazione](#)).

Un altro aspetto che è da tenere in grande considerazione è il concetto di obsolescenza. Le risorse analogiche sono deteriorabili, per questo è bene che vengano digitalizzate. Ma le risorse digitali o digitalizzate non sono meno soggette allo stesso destino, poiché la tecnologia è in continua evoluzione e linguaggi, software e hardware diventano obsoleti molto velocemente. Per andare incontro alla voce A2 dei principi FAIR¹, i dati e i metadati devono poter essere recuperabili anche quando le piattaforme o i sistemi informatici utilizzati per la loro raccolta e visualizzazione vengono dismessi o le tecnologie usate diventano obsolete, e per fare ciò è necessario adottare modelli condivisi e controllati. Quanto evidenziato finora rientra nei primi tre punti di debolezza identificati dal Piano Nazionale di Digitalizzazione (PND) del Patrimonio Culturale circa lo stato attuale dell’ecosistema digitale italiano per il patrimonio culturale (PND:5).

Con questo poster si intende quindi avanzare una riflessione sulla necessità di uniformità dei dati e sull’utilizzo di buone pratiche, elaborata a partire dai risultati parziali di un progetto dottorale [PON](#) in corso di svolgimento. Il progetto, in linea con il PND, ha come scopo la creazione di un’applicazione per la valorizzazione del patrimonio culturale italiano e si basa sul riuso delle risorse già presenti nel web e disseminate in più piattaforme eterogenee (cf. Immagine 1). Il beneficio primario è quello di preservare il patrimonio culturale messo già a disposizione in rete da altre istituzioni e individui e renderlo più fruibile per una platea diversificata di utenti.

¹ “Metadata are accessible, even when the data are no longer available”

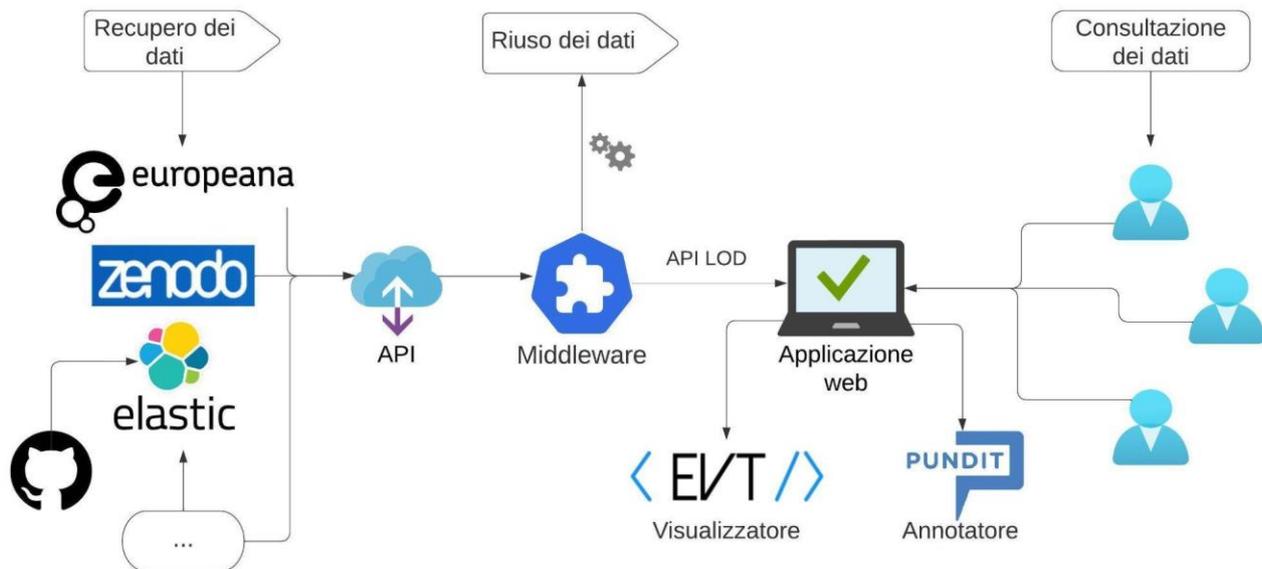


Immagine 1 – Workflow

2. STATO DELL'ARTE

Negli ultimi anni, e in particolare a seguito della pandemia, c'è stata una forte spinta verso la digitalizzazione delle risorse, sia a livello nazionale che europeo. Con l'aumento di fondi e finanziamenti indirizzati allo sviluppo digitale e una maggior attenzione alla preservazione del patrimonio culturale, diversi aggregatori sono sorti. A livello regionale possiamo citare a titolo esemplificativo e non esaustivo i portali [Puglia Digital Library](#), [Cultura Campania](#) e [Mèmorea](#) (Piemonte), mentre sul piano nazionale è opportuno fare riferimento a [CulturaItalia](#). Questo strumento è di grande rilevanza per il progetto oggetto di questo poster, poiché è il principale aggregatore italiano di dati per [Europeana](#), molti dei quali confluiranno nell'app. È inoltre un ottimo esempio di *FAIRness*, poiché recupera metadati esistenti tramite un Application Profile basato sulle linee guida della DCMI e li restituisce a Europeana in un ulteriore formato standard, EDM, basato anch'esso su capisaldi dell'interoperabilità quali OAI ORE, DC, SKOS e CIDOC-RM. (cf. <https://pro.europeana.eu/page/edm-documentation> e [3, 7])

È inoltre recentissima la pubblicazione della versione 1.1 del PND, tramite il quale il Ministero della cultura intende promuovere e, fondamentale, organizzare la trasformazione digitale dei beni culturali in Italia, andando a supportare gli enti che operano in questo settore. Un importante obiettivo esposto nel Piano è la creazione di un'infrastruttura software del patrimonio culturale per organizzare le risorse del patrimonio culturale e dei metadati in maniera corretta, affidabile, sicura ed efficiente (PND:35). Questo spazio è pensato per essere un supporto integrativo nel quale far confluire le diverse risorse provenienti dai vari domini dell'ecosistema per agevolarne il confronto.

Un altro aspetto molto importante è la documentazione pubblicata a disposizione dei fruitori dei vari siti e progetti (cf. [11]). Nella loro indagine, Franzini et al. 2019 cercano di capire quali sono le aspettative degli utenti e gli usi più comuni circa le edizioni digitali di testi. Sebbene il progetto in sviluppo dalla proponente si occupi di diversi tipi di risorse del patrimonio, alcuni risultati ottenuti dallo studio appena citato sono perfettamente estendibili a progetti digitali di respiro più ampio. In particolare, alla domanda 8) *How important is detailed editorial and technical documentation?* risponde il 69% dei partecipanti con *important* o *very important*. Gli autori commentano dicendo che “[d]ocumenting the process of creation behind any type of project serves to communicate development and quality, and to give appropriate context. With documentation, creators [...] facilitate the reuse of a resource. One of the basic principles and assumptions of research is *reproducibility* or, in other words, the ability of one researcher to take the work of another researcher follow the original researcher's pathway and arrive at the same results. Reproducibility is key to research acceptance and validation” ([9]:10-11). La riproducibilità e quindi la presenza di documentazione è stato un aspetto chiave per la scelta dei progetti da integrare nell'applicazione che si sta sviluppando. Infatti, diversi lavori che erano stati inizialmente presi in considerazione per l'integrazione dei dati sono stati poi esclusi poiché non esponevano le modalità di recupero di questi.

Infine, è necessario citare la *reusability* dei metadati, poiché è un aspetto cruciale per far sì che la risorsa digitale sia preservabile nel tempo. Barbuti 2021 propone un'espansione della *R* dei principi FAIR in *reusable*, *relevant*, *reliable* e *resilient*, ritenendo che questo ampliamento, con particolare riferimento ai metadati descrittivi, permetta di rendere le risorse digitali sostenibili e permanenti se presentano una corretta proporzione tra qualità e quantità di contenuto. Su

questo concetto in particolare si fonda il presente progetto dottorale, poiché è previsto esclusivamente il riuso di dati e strumenti già presenti nell'ecosistema digitale, modellati in base alle esigenze specifiche della ricerca in corso al fine di offrire una fruizione alternativa delle risorse.

3. OBIETTIVI

Lo scopo di questo poster sarà quindi di mettere in luce alcune delle soluzioni FAIR previste per lo sviluppo del progetto dottorale della proponente.

Il progetto nello specifico si propone di raccogliere in un'unica piattaforma risorse relative al patrimonio culturale italiano e integrare tra loro diversi strumenti di visualizzazione e ricerca. Più in generale, questo poster vuole essere esemplificativo di un approccio in linea con i principi FAIR per la creazione di un nuovo portale, i quali vengono osservati sin dal concepimento del lavoro. Infatti, non è sufficiente che essi vengano integrati in fasi successive a fini, per esempio, di pubblicazione, poiché tale approccio tende a indebolire il potenziale di riuso e integrazione dei dati. È necessario quindi che la ricerca aspiri continuamente alla *FAIRness* e ai fondamenti della gestione e condivisione dei dati durante tutto il processo di concepimento e realizzazione di un progetto. [6] A questo proposito, i cardini su cui si fonda questa ricerca sono:

- il riutilizzo delle risorse già disponibili e riutilizzabili, selezionate in base ai temi della ricerca;
- la scelta di strumenti *open source* per la visualizzazione e rielaborazione dei dati già esistenti;
- l'integrazione tra risorse e applicazioni;
- l'arricchimento dei contenuti.

Le domande implicite a cui questo progetto cerca di rispondere sono: è davvero possibile con gli strumenti attualmente esistenti realizzare un progetto basandosi esclusivamente su risorse e dati esistenti, sviluppando al massimo un "connettore" per mettere in comunicazione il tutto e aggregarlo in un unico ambiente che risulti comunque di semplice utilizzo per un utente specialista e non? Che sforzo richiede questa operazione e che tipo di conoscenze e competenze è stato necessario possedere?

Come accennato sopra, le risorse che saranno visualizzabili tramite questa applicazione verranno recuperate da aggregatori online tramite API. Per ora sono state sviluppate per Europeana, Zenodo, GitHub e si sta lavorando sulla connessione a un endpoint SPARQL sul patrimonio culturale di origine germanica in Veneto, anch'esso in sviluppo con un progetto PON. Viene estratto un insieme di metadati che vengono ricondotti al formato di metadati DCMI per uniformarne la consultazione così da creare dati ricercabili e interoperabili.

Nella piattaforma saranno integrati un software di visualizzazione e uno strumento per le annotazioni. Per implementare la prima funzionalità è stato scelto [Edition Visualization Technology](#), il quale dovrà essere adattato ai bisogni specifici del progetto per poter permettere all'applicazione di recuperare risorse esterne alla build. Questo strumento è ampiamente usato nei progetti di filologia digitale poiché è stato concepito per essere *open source* (il codice è scaricabile liberamente da [GitHub](#)), basato sugli standard del web e molto flessibile per essere adattato, appunto, alle esigenze di progetti differenti. In particolare, riceve come input file codificati in XML/TEI, è implementato nel framework [Angular](#) e supporta la visualizzazione di immagini [IIIF](#) (cf. [4], [8]).

Le annotazioni saranno invece possibili tramite l'installazione di un'estensione, [Pundit](#)², sviluppata dall'azienda partner del progetto [Net7](#), la quale è disponibile anche sull'[EOSC Marketplace](#). Anch'essa è rilasciata con licenza pubblica, può essere usata in ottica collaborativa e offre funzionalità quali evidenziazione, commenti, tag e annotazioni semantiche. Quest'ultime possono essere annotazioni in senso stretto (metadati) oppure grafi risultanti dai metadati e le relazioni nelle varie risorse web che l'utente ha creato con Pundit. È presente anche una dashboard che permette di gestire tutte le annotazioni che possono essere anche filtrate ed esportate in vari formati standard e/o aperti ([5]). Allo stesso modo, l'interfaccia utente sarà simile a quella visualizzabile nella [demo](#) di Muruca, sempre sviluppata dalla suddetta azienda. In particolare, questo strumento permetterà di organizzare le risorse in collezioni e creare percorsi ([1], [10]).

Un primo tentativo di risposta alle domande sopra avanzate può essere quindi che un progetto di questo tipo è implementabile perché gli strumenti e le risorse FAIR esistono e possono essere integrati per dar loro valore aggiunto. Tuttavia, sarà necessario adattarli ai bisogni specifici della ricerca e bisognerà comunque fare i conti con i "limiti" e la qualità più o meno adeguata di ciò che è presente in rete. La risposta implicita alla seconda domanda è che oltre alla collaborazione tra esperti di discipline umanistiche e sviluppatori informatici occorre sviluppare una sensibilità comune alle varie questioni di riusabilità dei dati e del codice. La multidisciplinarietà del settore delle Digital Humanities consiste

² È disponibile anche un video introduttivo: <https://youtu.be/sfGS6aYsV3s>

anche in questo: un affiancamento di esperti di vari ambiti ma che condividono una conoscenza comune di buone pratiche e metodologie.

BIBLIOGRAFIA

- [1] Aiola, Chiara, Giulio Andreini, Francesca Di Donato, e Tiziana Lombardo. "Sharing knowledge digitally, the Muruca case study." In *AIUCD 2021 – DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale. Raccolta degli abstract estesi della 10° conferenza nazionale 2021*, (2021): 428-431. <https://aiucd2021.labcd.unipi.it/book-of-abstracts/>
- [2] Barbuti, Nicola. "Thinking digital libraries for preservation as digital cultural heritage: by R to R⁴ facet of FAIR principles." *International Journal on Digital Libraries* 22, (2021): 309–318. <https://doi.org/10.1007/s00799-020-00291-7>
- [3] Buonazia, Irene, Maria Emilia Masci. "Il PICO Application Profile. Un Dublin Core Application Profile per il Portale della Cultura Italiana." In *Interoperabilità di contenuti e servizi digitali: metadati, standard e linee guida*. (2007): 45-51. <https://hdl.handle.net/11384/13790>
- [4] Cacioli, Giulia, Giacomo Cerretini, Chiara Di Pietro, Sara Maenza, Roberto Rosselli Del Turco e Simone Zenzaro. "There and back again: what to expect in the next EVT version." In *AIUCD 2022 - Digital cultures. Intersections: philosophy, arts, media. Proceedings of the 11th national conference, Lecce, 1-3 June 2022*, Fabio Ciraci, Giulia Miglietta e Carola Gatto (ed.). (2022): 212-217. <http://amsacta.unibo.it/6848/>
- [5] De Santis, Luca, Duccio Breschi, Niccolò Cardelli, Edgar Gomez, Massimiliano Pardini, Marco Zizi, Maxime Bouillard, Simone Kopeinik e Laurent Capelli. "TRIPLE Deliverable: D5.5 Report on the Open Annotation Tool (Draft)." *Zenodo*. (2021). <https://doi.org/10.5281/zenodo.5653219>
- [6] Dempsey, William, Ian Foster, Scott Fraser e Carl Kesselman. "Sharing Begins at Home: How Continuous and Ubiquitous FAIRness Can Enhance Research Productivity and Data Reuse." In *Harvard data science review* 4, no. 3 (2022). <https://doi.org/10.1162/99608f92.44d21b86>
- [7] Di Giorgio, Sara. "Culturaitalia, the Italian National Content Aggregator in Europeana." In *Procedia Computer Science* 38, (2014): 40-43. <https://doi.org/10.1016/j.procs.2014.10.008>
- [8] Di Pietro, Chiara, Chiara Martignano e Roberto Rosselli Del Turco. "Progettazione e implementazione di nuove funzionalità per EVT 2: lo stato attuale dello sviluppo." *Umanistica Digitale*, no. 7 (2019). <http://doi.org/10.6092/issn.2532-8816/9322>
- [9] Franzini, Greta, Melissa Terras e Simon Mahony. "Digital Editions of Text: Surveying User Requirements in the Digital Humanities." *ACM Journal on Computing and Cultural Heritage* 12, no. 1 (2019). <https://doi.org/10.1145/3230671>
- [10] Lombardo, Tiziana e Chiara Aiola. "Storytelling through Digital Scholarly Editions." *Zenodo*. (2022). <https://doi.org/10.5281/zenodo.6594346>
- [11] Warwick, Claire, Isabel Galina, Jon Rimmer, Melissa Terras, Ann Blandford, Jeremy Gow e George Buchanan. "Documentation and the users of digital resources in the humanities." *Journal of Documentation* 65, no. 1 (2009): 33-57. Emerald Group Publishing Ltd. <https://www.doi.org/10.1108/00220410910926112>