

CLEF 2.0. Solutions for Native Linked Data Cataloguing of Italian Digital Cultural Heritage*

Sebastiano Giacomini^(a), Marilena Daquino^(b), Francesca Tomasi^(c),
Laurent Antoine Fintoni^(d)

a) Alma Mater Studiorum University of Bologna, <https://orcid.org/0009-0007-7813-0939>

b) Alma Mater Studiorum University of Bologna, <https://orcid.org/0000-0002-1113-7550>

c) Alma Mater Studiorum University of Bologna, <https://orcid.org/0000-0002-6631-8607>

d) Alma Mater Studiorum University of Bologna, <https://orcid.org/0000-0002-8656-4602>

Contact: Sebastiano Giacomini, sebastiano.giacomin2@unibo.it; Marilena Daquino, marilena.daquino2@unibo.it;
Francesca Tomasi, francesca.tomasi@unibo.it; Laurent Antoine Fintoni, laurent.fintoni2@unibo.it

Received: 29 February 2024; **Accepted:** 12 July 2024; **First Published:** 15 January 2025

ABSTRACT

The Semantic Web had a significant impact on the GLAM domain, where the need to connect knowledge has grown so important that it has sparked numerous crowdsourcing initiatives and collaborative native Linked Open Data cataloguing projects. One of the key challenges in facing these collaborative activities is the heterogeneity of content and levels of expertise among users. While existing solutions can meet several important requirements, developing new features is often affected by the practical use of such platforms in real-world work settings. This article explores novel demands and presents the solution proposed by CLEF 2.0, the native Linked Open Data cataloguing software adopted in various case studies concerning the description of Italian Digital Cultural Heritage.

KEYWORDS

Crowdsourcing; Cataloguing; Linked Open Data; Cultural Heritage.

CLEF 2.0. Soluzioni per la catalogazione nativa Linked Data del patrimonio digitale culturale italiano*

ABSTRACT

L'affermazione del Web Semantico ha avuto un impatto significativo nel settore delle istituzioni GLAM, per le quali la connessione dei saperi ha assunto una rilevanza tale da produrre numerose iniziative di *crowdsourcing* e progetti collaborativi di catalogazione nativa Linked Open Data. Una sfida attuale che interessa tali attività collaborative riguarda l'eterogeneità dei contenuti e dei gradi di competenza posseduti dagli utenti. Se da un lato soluzioni esistenti riescono a soddisfare i requisiti minimi in questo ambito di lavoro, spesso a dettare le linee guida dello sviluppo di nuove funzionalità è il concreto impiego di queste stesse piattaforme in contesti di lavoro pratici. Il presente articolo intende analizzare queste esigenze e presentare la soluzione proposta da CLEF 2.0, il software per la catalogazione nativa Linked Open Data adottato in alcuni casi di studio inerenti alla descrizione del patrimonio culturale digitale italiano.

PAROLE CHIAVE

Crowdsourcing; Catalogazione; Linked Open Data; Patrimonio Culturale.

* Attribuzione di responsabilità degli autori: Francesca Tomasi è responsabile per la sezione 1; Marilena Daquino è responsabile per la sezione 2 e 4.1; Sebastiano Giacomini è responsabile per le sezioni 3.2, 4.2 e 6; Laurent Fintoni è responsabile per le sezioni 3.1. Tutti gli autori sono responsabili per la sezione 5.

1. Introduzione

L'affermazione delle tecnologie del Web Semantico ha spinto istituzioni e professionisti a riconsiderare l'organizzazione dei propri saperi e le metodologie per la condivisione della conoscenza (Poblet, Casanovas, e Rodríguez-Doncel 2019). Questa visione del Web, con i suoi processi trasformativi, ha avuto un impatto significativo sulle istituzioni GLAM (*Galleries, Libraries, Archives, and Museums*), offrendo una prospettiva di rinnovamento (Davis e Heravi 2021) e di superamento del tradizionale isolamento tra le diverse collezioni di dati (Marden et al. 2013). Non a caso, l'importanza della condivisione dei saperi tra istituzioni e cittadini ha assunto in tempi recenti una rilevanza tale da produrre una proliferazione di iniziative di *crowdsourcing*, che non sono però sempre in grado di sfruttare appieno le possibilità offerte dal *Semantic Web* (Daquino 2021).

Sebbene lo sviluppo del Web Semantico miri ad una progressiva automatizzazione delle proprie attività, l'utilizzo delle tecnologie coinvolte resta fortemente ancorato all'intervento umano. A determinare questo vincolo, è anzitutto la natura stessa delle attività in esame, le quali richiedono una conoscenza approfondita della materia e dipendono fortemente dal contesto di impiego (Sarasua et al. 2015), causando spesso uno scollamento tra le esigenze reali di progetti collaborativi e lo stato dell'arte di soluzioni tecnologiche per il *crowdsourcing* nativo Linked Open Data (LOD). Nonostante i numeri registrati da alcune iniziative (es. Wikidata¹) abbiano rivelato nel tempo l'enorme potenziale di un approccio collaborativo nell'ambito della creazione di Linked Data, il *crowdsourcing* continua a porre sfide ancora attuali.

Una questione fondamentale riguarda l'eterogeneità dei contenuti e dei gradi di competenza posseduti dagli utenti. Negli ultimi anni sono emerse una serie di applicazioni e *content management systems* (CMS) dedicati alla creazione condivisa e collaborativa di collezioni di Linked Data. In tali soluzioni si va alla ricerca di un equilibrio tra le opposte esigenze di garantire un sistema di catalogazione omogeneo e coerente e di evitare semplificazioni della realtà, garantendo ad utenti con diversi livelli di esperienza la possibilità di condividere facilmente i loro saperi senza porre limitazioni alle possibilità descrittive dei più esperti.

Se da un lato alcune caratteristiche risultano indispensabili per soddisfare i requisiti minimi di qualità e di corretto funzionamento di qualsiasi soluzione in questo ambito di lavoro (es. la possibilità di utilizzare vocabolari controllati, di poter collegare contenuti multimediali e metadati, di avere ruoli e privilegi diversi all'interno del processo di revisione editoriale dei dati), spesso a dettare le linee guide dello sviluppo di nuove funzionalità per le applicazioni di *crowdsourcing* è il concreto impiego di queste stesse piattaforme in contesti di lavoro pratici. Tuttavia, se l'obiettivo di un sistema è quello di offrire supporto ad un'ampia gamma di progetti, l'integrazione di nuove funzionalità andrà effettuata nel rispetto di tale esigenza, senza compromettere l'usabilità per l'utente finale, specie se con limitata esperienza nel settore delle tecnologie semantiche. Proprio in merito a questo auspicato compromesso si rivelano carenti la maggior parte delle soluzioni sin qui sviluppate, le quali tendono a sacrificare, alternativamente, l'usabilità, la qualità, gli strumenti di automatizzazione e le funzionalità fondamentali per la creazione di una collezione di Linked Data. Negli ultimi anni alcune applicazioni hanno dato risposta alle richieste di iniziative di catalogazione LOD, ma l'emergere di nuovi progetti sempre più complessi ha segnato in tempi recenti la

¹ <https://www.wikidata.org/wiki/Wikidata:Statistics/it>.

necessità di apportare ulteriori modifiche sostanziali agli strumenti in questione. Alla necessità di soddisfare importanti livelli di qualità, usabilità e affidabilità si vanno infatti affiancando nuovi requisiti di completezza, ricchezza e precisione descrittiva.

Il presente articolo intende analizzare queste esigenze e presentare la soluzione proposta da CLEF 2.0, la nuova versione di CLEF (Crowdsourcing Linked Entities via web Form)², un'applicazione web sviluppata dal centro di ricerche in Digital Humanities (/DH.arc) dell'Università di Bologna (Daquino et al. 2023). L'esame di questo aggiornamento prende le mosse dall'analisi di tre casi di studio: Global Education and Learning (GEL)³, un progetto co-finanziato dall'UNESCO per la creazione di un catalogo bibliografico sulla letteratura internazionale sull'educazione alla cittadinanza globale; ATLAS⁴, un progetto finanziato dall'Unione europea – Next Generation EU (PRIN 2022) per la pubblicazione di un catalogo LOD di progetti di ricerca in Digital Humanities; KNOT⁵, un progetto pilota dell'Università di Bologna e l'Istituto Centrale per la Digitalizzazione del Patrimonio Culturale del Ministero della Cultura per la valorizzazione del patrimonio culturale digitale degli Atenei italiani.

L'articolo è organizzato come segue. Nella sezione 2 presentiamo le principali soluzioni per la catalogazione in LOD e i requisiti generali a cui rispondono. In sezione 3 presentiamo i requisiti raccolti a partire dai nuovi casi di studio e confrontiamo le soluzioni esistenti, giustificando la scelta di estendere CLEF, le cui integrazioni vengono presentate in sezione 4. In sezione 5 discutiamo i risultati raggiunti, presentando gli aspetti di maggiore impatto e le limitazioni attuali. Concludiamo in sezione 6 con le prospettive future.

2. Stato dell'arte

L'attuale panorama dei sistemi di *crowdsourcing* LOD è caratterizzato da una varietà di proposte. Tra queste, spicca anzitutto Semantic MediaWiki (SMW)⁶, un'estensione aperta e gratuita di MediaWiki, l'applicazione alla base di servizi come Wikipedia e Wikidata. Questo sistema nasce con lo scopo di arricchire i tradizionali wiki, costituiti da solo testo, con delle annotazioni semantiche, producendo dati disponibili come LOD e conservati presso il triplestore del progetto.

Se SMW offre una soluzione per la generazione e visualizzazione di dati all'interno dei wiki stessi, un tool con caratteristiche simili, Wikibase⁷, è invece progettato per consentire la creazione e gestione collaborativa di Linked Data riutilizzabili in applicazioni esterne (Diefenbach et al. 2021). Tutti i dati prodotti in Wikibase sono nativamente conservati in un database relazionale, incluse le informazioni fondamentali circa la gestione delle attività dei collaboratori e dei loro permessi. Gli stessi dati sono inoltre disponibili, tramite esportazione, in un triplestore Blazegraph.

Un'altra importante realtà è rappresentata da Omeka S⁸. Il progetto, nato come evoluzione del

² <https://polifonia-project.github.io/clef/>, rilasciato con Licenza ISC.

³ <https://projects.dharc.unibo.it/digestgel/>.

⁴ <https://dh-atlas.github.io/>

⁵ https://icdp-digital-library.github.io/KNOT/website/IT/index_it.html.

⁶ https://www.semantic-mediawiki.org/wiki/Semantic_MediaWiki.

⁷ <https://wikiba.se/>.

⁸ <https://omeka.org/s/>.

software Omeka Classic per la creazione di esposizioni online, si propone di mettere a disposizione uno strumento flessibile ed estendibile. I dati prodotti sono conservati in un database relazionale e sono accessibili come documenti JSON-LD tramite API, mentre non sono disponibili serializzazioni in altri formati RDF e i dati non possono essere interrogati in linguaggio SPARQL.

Sempre in ambito GLAM figura ResearchSpace⁹, un progetto open source ideato dal British Museum di Londra e in collaborazione con Metaphacts (Oldman e Tanese 2018). Il risultato tangibile di questo ambizioso progetto coincide con la pubblicazione di collezioni di LOD, accessibili mediante interfacce costruite dagli utenti tramite la compilazione di Template che richiedono una conoscenza approfondita dei linguaggi HTML, SPARQL e JavaScript. I dati prodotti sono immediatamente disponibili nel triplestore e possono essere interrogati tramite query SPARQL.

Chiude questa breve rassegna dei principali CMS, Sinopia¹⁰, un ambiente di editing di Linked Data sviluppato nell'ambito dell'iniziativa Linked Data for Production. Sebbene altre ontologie siano riutilizzabili al suo interno, Sinopia nasce ottimizzata per implementare il data model di BIBFRAME¹¹ e dunque per catalogare risorse bibliografiche.

I tool illustrati sono accomunati da un funzionamento simile, basato sulla suddivisione dei compiti tra diversi livelli di utenti e sulla creazione di modelli (*template*) per la generazione di schede catalografiche (*record* o *item*). Generalmente, un template corrisponde ad un web form, costituito da una serie di campi ciascuno con le proprie peculiarità ed associato ad una classe ontologica. La definizione di un template non può prescindere da un certo grado di competenza in materia di produzione di Linked Data. Proprio in ragione di questo assetto organizzativo, apportare modifiche a un'applicazione di *crowdsourcing* e introdurre al suo interno nuove funzionalità significa, in prima battuta, stabilire chi dovrà avere accesso ai nuovi strumenti.

Catalogare nativamente in LOD non significa soltanto raccogliere dati e pubblicarli secondo i formati del Web Semantico. Dietro questi processi creativi si nascondono infatti importanti sfide tecniche e teoriche (Hawkins 2021). Un primo aspetto fondamentale riguarda l'affidabilità dei dati pubblicati dagli utenti. Non a caso, un'ampia letteratura pone al cuore del Web Semantico il concetto di *trust* (fiducia), termine col quale riferirsi a un metodo attraverso cui un individuo accetta la caratterizzazione proposta da altri circa il loro lavoro (O'Hara et al. 2004).

Tra i meccanismi messi a punto per garantire in concreto l'affidabilità dei dati condivisi assume grande importanza la documentazione della provenienza (*provenance*) degli asserti (Tomasì 2023). Un punto, questo, spesso ignorato dai tool per la generazione di LOD, che tuttavia appare fondamentale per prevenire incongruenze, porre l'accento sulle responsabilità dei contenuti e, conseguentemente, rafforzare la fiducia nei dati raccolti (Daquino et al. 2023).

Al concetto di *provenance* è riservato un ruolo cruciale anche nell'ambito dei requisiti FAIR (Findability, Accessibility, Interoperability, Reusability), un insieme di principi volti a guidare la produzione e la pubblicazione di dati (Wilkinson et al. 2016). Questa serie di linee guida ha avuto un forte impatto sul dominio dei beni culturali, dove solo un'attenta implementazione dei principi FAIR ha reso possibile la creazione di Linked Data adatti al riuso e all'analisi secondo i metodi di

⁹ <https://researchspace.org/>.

¹⁰ <https://sinopia.io/>.

¹¹ <https://www.loc.gov/bibframe/>.

ricerca delle Digital Humanities (Hermon e Niccolucci 2021). Tuttavia, i criteri FAIR hanno talvolta assunto ruolo marginale in progetti di *crowdsourcing*, dove i dati raccolti sono spesso trattati come dati di serie B, subiscono processi diversi da quelli dei cataloghi ufficiali e portano quindi a scarsa trasparenza e attendibilità.

A questo problema ha tentato di porre rimedio CLEF, una recente proposta software per il *crowdsourcing* di LOD (Daquino et al. 2023). CLEF nasce con lo scopo di supportare progetti di piccole e medie dimensioni nella creazione collaborativa di collezioni LOD, facendo ricorso ad uno strumento familiare e di facile utilizzo come quello del Web Form. Attraverso un'interfaccia improntata all'usabilità, l'obiettivo di CLEF è quello di estendere e facilitare l'accesso alla creazione di Linked Data a un bacino di collaboratori sempre più vasto, indipendentemente dalla conoscenza delle tecnologie del Web Semantico. Un ruolo centrale nel percorso di creazione di un catalogo CLEF è riservato al rispetto dei principi FAIR e alla gestione delle informazioni circa la provenienza dei dati. Documentare la provenienza significa registrare i processi creativi dei dati nel loro stato attuale (Mitchell et al. 2022). Per fare ciò, CLEF recupera il data model definito dalla Provenance Ontology¹² per registrare le informazioni sul processo editoriale all'interno di *named graphs*. Queste informazioni, insieme ai dati delle risorse catalogati, sono resi immediatamente disponibili tramite triplestore (Blazegraph) e interrogabili tramite un apposito endpoint SPARQL, diversamente da altri CMS basati su database relazionali. CLEF pone una marcata attenzione per l'usabilità dell'applicazione finale. La sua struttura modulare, tanto a livello di interfaccia quanto a livello di codice, consegna allo sviluppatore uno strumento facilmente maneggevole per l'introduzione di nuove funzionalità.

A chiusura di questa breve introduzione allo stato dell'arte, in Tabella 1 vengono elencati i software considerati e come questi rispondono ai requisiti emersi dall'analisi svolta da (Daquino et al. 2023) e riassumibili nei seguenti elementi di indagine: l'**usabilità** delle interfacce, tanto per gli utenti quanto per gli amministratori, ai fini di un'efficiente raccolta, validazione e pubblicazione di dati; la presenza di sistemi per la gestione e il salvataggio di informazioni sulla **provenienza** dei dati; l'integrazione nel software di *workflow* per la **gestione dei dati** (es. *versioning* di dati); il rispetto di criteri di **sostenibilità** e **riusabilità**.

Nome	Usabilità (utenti)	Usabilità (amministratori)	Gestione della Provenienza	Gestione Dati	Sostenibilità Riusabilità
Omeka S	✓	✓			✓
SMW	✓	✓	✓		✓
Sinopia	✓	✓			✓
ResearchSpace	✓		✓	✓	✓
Wikibase	✓	✓	✓	✓	
CLEF	✓	✓	✓	✓	✓

Tabella 1. Panoramica dei sistemi per la creazione collaborativa di Linked Data (Daquino et al. 2023).

¹² <https://www.w3.org/TR/prov-o/>.

Alla luce della comparazione emerge un quadro frastagliato, dove CLEF cerca di rispondere ai requisiti raccolti al momento del suo sviluppo. Al contempo CLEF non fornisce tutte le funzionalità di editing che le altre piattaforme invece offrono - le quali non sono discusse nelle analisi precedenti. Lungi dal mettere un punto sulla questione, consideriamo però CLEF come un punto di partenza per l'integrazione di requisiti (nuovi o già accolti da altre soluzioni) emersi in progetti recenti che hanno deciso di adottare CLEF per catalogare il patrimonio culturale digitale italiano.

3. Raccolta dei requisiti e analisi dei competitor

3.1 Raccolta dei requisiti

Tra le attività di ricerca che hanno spinto verso un aggiornamento dei sistemi e delle pratiche per la pubblicazione collaborativa di LOD, figura il caso di studio Global Education and Learning (GEL), un progetto concluso dedicato alla produzione e all'aggiornamento di una base di conoscenza bibliografica dedicata alla letteratura sui temi dell'educazione alla cittadinanza globale.

Più recenti sono gli altri due casi di studio, KNOT e ATLAS. Il primo è un progetto triennale pilota (2022-2025) dell'Università di Bologna sviluppato in collaborazione con l'Istituto Centrale per la Digitalizzazione del Patrimonio Culturale del Ministero della Cultura. Alla base di KNOT, vi è l'intento di valorizzare il patrimonio culturale digitale degli Atenei italiani, che comprende anzitutto progetti e prodotti della ricerca in ambito Digital Humanities. Uno dei primi obiettivi che KNOT intende raggiungere coincide con la creazione collaborativa di un *knowledge graph*. Nella stessa direzione va ATLAS, un progetto PRIN 2022 iniziato a ottobre 2023 e dedicato al censimento e l'estrazione di dati a partire da progetti di Digital Humanities aventi come oggetto il patrimonio culturale italiano. Fondamentale, in questo senso, è la definizione di un workflow per la raccolta, normalizzazione, bonifica e trasformazione dei dati in formato RDF in modo che tale processo non risulti essere alieno alla quotidiana attività editoriale di catalogazione, ma che sia invece integrato nel workflow editoriale.

Tra i progetti menzionati, KNOT e ATLAS presentano maggiori complessità, data l'eterogeneità delle risorse che si propongono di catalogare e il livello di interoperabilità semantica con risorse esterne (vocabolari e thesauri di dominio) che le descrizioni devono ottenere. Si parte infatti da edizioni critiche di testi, tradizioni testuali, corpora linguistici, per arrivare a tool più tecnici quali software di visualizzazione, servizi di raccomandazione e sistemi di analisi testuale. I *desiderata* e le questioni emersi dall'analisi preliminare e dal confronto con i committenti dei progetti menzionati hanno portato alla delineazione dei seguenti requisiti.

- **Entity Linking.** Uno dei principi cardine dietro la produzione di Linked Data riguarda l'opportunità e necessità di stabilire collegamenti con risorse ed entità presenti nel web per mezzo dei loro URI (Thanos 2017). Se da un lato queste connessioni consentono un notevole risparmio di risorse grazie al riutilizzo di informazioni già presenti in rete, recuperare le esatte entità di interesse potrebbe rivelarsi un compito gravoso, specie quando utenti con scarsa conoscenza dei mezzi a loro disposizione sono messi a confronto con enormi moli di dati. Non è da sottovalutare il rischio di errori, come ad esempio in casi di omonimie all'interno della stessa *knowledge base*, come avviene per gli autori descritti in GEL. La somma di queste complessità risulta dunque nella necessità di un rafforzamento degli strumenti

in grado di agevolare gli utenti nel recupero di entità già presenti in rete, es. generando suggerimenti automatici in tempo reale e metodi per la disambiguazione.

- **Datatype temporali.** Le complessità descrittive dei progetti catalogati in KNOT e ATLAS evidenziano la necessità di ampliare il numero di tipologie di dati (*datatype*) fruibili da parte dell'utente finale, specialmente quando si ha a che fare con date e periodi dai confini incerti.
- **Vocabolari controllati e thesauri SKOS.** Analogamente a quanto già espresso in merito al riutilizzo e alla riconciliazione di entità, una funzionalità di grande utilità nelle attività di catalogazione consiste nel recupero e riuso di nomenclature standardizzate. È il caso, ad esempio, dei vocabolari SKOS, delle risorse controllate i cui termini, dal significato comune e condiviso, risultano spesso fondamentali ai fini di una descrizione puntuale delle risorse in esame.
- **Multimedia.** Le possibilità descrittive e di arricchimento di una entità non si limitano ai soli dati testuali, ma possono richiedere l'ausilio di altri media, ovvero immagini, video e file audio, o anche anteprime di pagine web esterne. Come nel caso dei progetti catalogati in KNOT, l'introduzione di una preview delle risorse in rete descritte (tramite inclusione di un *iframe*) consentirebbe di completare la descrizione dell'oggetto in esame senza richiedere all'utente che naviga il catalogo di abbandonare la pagina web in cui queste sono citate.
- **Knowledge Extraction.** Un ultimo strumento riguarda una funzionalità avanzata in grado di supportare l'estrazione semi-automatica di entità chiave per la descrizione di un'entità a partire da una fonte di dati online (es. SPARQL endpoint, file statico, API). Questa funzionalità può prendere la forma di servizi di Named Entity Recognition da un testo pieno importato nella piattaforma di catalogazione, o può essere disegnata da un utente esperto proponendo una query personalizzata, basata sulle caratteristiche della fonte da interrogare.

3.2 Analisi comparata

Sebbene alcuni applicativi siano già in grado di far fronte ad alcune delle richieste emerse, a mancare è più spesso un'integrazione coerente e di facile utilizzo delle stesse. Per fornire un quadro più esaustivo possibile, di seguito vengono riportate le linee di lavoro individuate sulla base dei nuovi requisiti raccolti: per ciascuna di esse, vengono brevemente esposte le principali soluzioni esistenti.

Entity Linking. In CLEF, questo genere di funzionalità è già presente, seppure le entità siano al momento estraibili soltanto da Wikidata e dal catalogo stesso. Una soluzione analoga è stata proposta da Sinopia e Omeka S, che consentono ai propri utenti di ricevere dei suggerimenti da delle basi di dati predefinite. In ResearchSpace i creatori di Template possono specificare una query SPARQL da associare ad un campo per generare automaticamente dei suggerimenti, mentre il meccanismo di autocompletamento disponibile in Semantic MediaWiki consente il riuso delle entità presenti in Wikidata e di dati, semantici e non, estratti tramite API costruite *ad hoc* o attraverso un'apposita espansione.

Datatype temporali. All'interno di CLEF, le opzioni disponibili sono limitate a Literal (stringhe) e URI, escludendo di fatto altri tipi di dati fondamentali come le date e i periodi temporali. La definizione di datazioni è disponibile in Sinopia, mentre Omeka S dispone di un apposito modulo per la gestione di datatype numerici che consente di associare ad una voce di un Template un triplice campo di input (Anno, Mese, Giorno). A seconda dei dati forniti, il sistema salverà la datazione in input associandole un datatype (xsd:date, xsd:dateTime, xsd:gYear e xsd:gYearMonth). Un approccio del tutto analogo è proposto anche da Semantic MediaWiki, mentre molto più ricca è la scelta in ResearchSpace, dove a ciascun campo è associabile uno fra i datatype specificati in RDF 1.1.

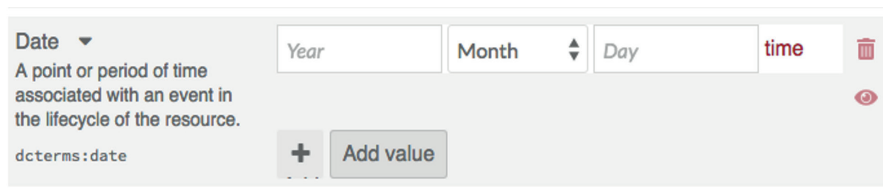


Figura 1. Un form per le date in Omeka S.

Vocabolari controllati e thesauri SKOS. Tra i vantaggi fondamentali dietro l'utilizzo di vocabolari controllati occorre menzionare a) l'opportunità di organizzare grandi quantità di dati mediante raggruppamenti e classificazioni, b) la riduzione delle ambiguità grazie all'impiego di termini dal significato condiviso, c) la semplificazione dei processi di interoperabilità dei dati e di integrazione di set di dati (Zaytseva e Āurĉo, 2020). Si pensi, nel caso di KNOT, all'esigenza di associare un progetto ad una disciplina o ambito di lavoro.

L'utilizzo di vocabolari controllati è una funzionalità integrata in Omeka S in seguito all'installazione del relativo modulo. Oltre ad una ricca scelta iniziale, agli utenti è concessa la possibilità di introdurre dei nuovi vocabolari specificando manualmente i termini. Similmente, in CLEF, gli utenti accreditati possono nativamente inserire liste di termini controllate nella forma <label, URI>. Tuttavia, importare liste di dimensioni più ampie può essere fonte di errori e ogni campo può essere associato ad un solo vocabolario. Sinopia offre un'ampia scelta di risorse controllate, ma non prevede alcuna funzione per l'integrazione di nuovi vocabolari, mentre più complete sono le soluzioni proposte da Semantic MediaWiki e ResearchSpace, che si avvalgono degli stessi strumenti già descritti in riferimento al task di Entity Linking per l'integrazione di risorse esterne.

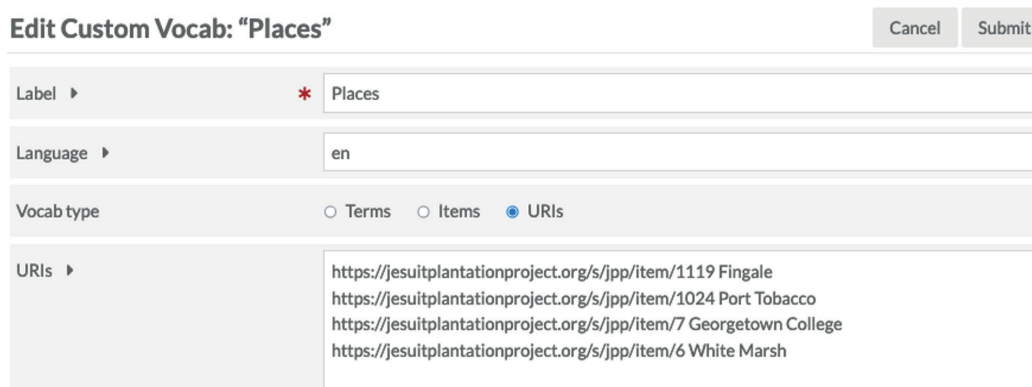


Figura 2. Un form per l'inserimento manuale di vocabolari controllati in Omeka S.

Multimedia. L'utilizzo di risorse multimediali è presente in tutti i sistemi tranne Sinopia e CLEF. In sistemi come ResearchSpace e Omeka S è inoltre possibile integrare strumenti più avanzati come un visualizzatore IIIF. Meno frequente è invece il ricorso agli iframe, disponibili comunque in Omeka S e in Semantic MediaWiki, a cui sono più spesso preferiti dei semplici link alla risorsa.

Knowledge Extraction. Come già anticipato, lo sviluppo degli applicativi per la creazione di LOD auspica la ricerca di un compromesso tra due opposte esigenze: da un lato fornire strumenti in grado di supplire alle carenze tecniche degli utenti meno esperti, dall'altro, promuovere una progressiva automatizzazione di attività più complesse. Si è visto ad esempio il ruolo determinante dei servizi di Entity Linking. Tuttavia, in alcuni casi risulta conveniente, se non addirittura necessario, operare un recupero di un'ampia mole di dati (es. la lista di tutte le persone menzionate in un corpus linguistico descritto in ATLAS). In simili scenari, sarebbe auspicabile mettere a disposizione dei collaboratori più esperti un nuovo strumento per l'estrazione semiautomatica di dati tramite interrogazioni mirate. Attualmente, nessuno dei sistemi esaminati prevede un servizio di questo tipo. La sua implementazione non dovrebbe limitarsi al solo recupero di dati RDF, bensì andrebbe estesa anche all'estrazione di informazioni ed entità chiave da altri servizi e risorse (es. API, file statici, ...).

	Entity linking	Datatype temporali	Import automatico di nuovi thesauri	Multimedia	Knowledge Extraction
CLEF	✓				
OmekaS	✓	✓		✓	
Sinopia	✓	✓			
SMW	✓	✓	✓	✓	
ResearchSpace	✓	✓	✓	✓	
Wikibase	✓	✓		✓	

Tabella 2: funzionalità emerse e soddisfacimento nei principali CMS.

In Tabella 2 vengono sintetizzati i nuovi requisiti emersi e la capacità dei software esistenti di rispondere a tali esigenze. Rispetto all'analisi iniziale vediamo come CLEF soffra di alcune lacune a cui altri *competitor* hanno invece già saputo rispondere. Non di meno, l'assenza totale di soluzioni per knowledge extraction semi-automatica, ci fa comunque valutare positivamente la possibilità di espandere CLEF rispetto alle altre soluzioni data la modularità e sostenibilità del codice e la possibilità di effettuare tali integrazioni in un sistema che sia effettivamente nativo LOD. Considerando i requisiti identificati e le analisi proposte, un ulteriore obiettivo nello sviluppo di CLEF 2.0 è comprendere come ottimizzare i meccanismi e le funzionalità esistenti all'interno di un sistema unificato e coerente. In tal senso, CLEF ha dimostrato la sua capacità di fornire un ambiente di lavoro favorevole all'integrazione di nuovi strumenti, preservando le caratteristiche della sua versione originale.

4. CLEF: il sistema collaborativo di catalogazione nativo Linked Open Data

4.1 CLEF 1.0

In sintesi, CLEF si basa sulla suddivisione dei compiti tra membri accreditati e contributori. Il ruolo dei primi consiste nella definizione dei parametri fondamentali del progetto in costruzione (e.g. nome del catalogo, l'endpoint), gestire l'organizzazione dei dati attraverso la definizione di Template e validare i dati inseriti per la pubblicazione. La creazione di un Template consiste nella combinazione di molteplici campi di input tra quelli messi a disposizione dalla piattaforma, che nella versione originaria di CLEF include le seguenti opzioni:

Textbox. Un campo di testo, destinato ad accogliere alternativamente una fra le seguenti tipologie di valori: brevi stringhe di testo, stringhe riconciliate ad URI di entità provenienti da Wikidata e dal catalogo stesso, stringhe riconciliate a località geografiche di GeoNames. Questa funzionalità è presente anche in Omeka S.

Textarea. Un campo di input per descrizioni testuali estese, dalle quali è possibile estrarre entità nominate riconciliate a Wikidata, tramite un sistema di Named Entity Recognition e Data Reconciliation. Funzione non disponibile in nessuno dei software analizzati.

Dropdown. Menù a tendina popolato con una serie di etichette associate a URI, tra cui l'utente finale sarà chiamato a selezionare un solo valore. Questa funzionalità si ritrova anche negli altri competitor di CLEF.

Checkbox. Segue un funzionamento del tutto analogo a quello di un Dropdown pur consentendo all'utente finale di selezionare più di un valore tra quelli proposti. Questa funzionalità non è presente in altri competitor.

4.2 CLEF 2.0

CLEF 2.0 interviene sulla lista di opzioni menzionata sopra per arricchire il sistema con funzionalità che rispondono ai requisiti raccolti. Ad essere aggiornati non sono soltanto le possibilità di interazione degli utenti, ma anche il sistema di gestione che consente la generazione, serializzazione e salvataggio di Linked Data, oltre alla visualizzazione finale dei dati.

Entity Linking. CLEF integra già una funzionalità di suggerimento automatico per assistere gli utenti durante la compilazione di una textbox. Nella lista di suggerimenti compaiono entità provenienti da Wikidata e dal catalogo stesso, nel caso in cui l'entità cercata non esista in Wikidata, associati ad una breve descrizione per facilitare la disambiguazione.

Wikidata rappresenta una risorsa di eccezionale importanza nel panorama dei Linked Data, ma nonostante alcune recenti proposte di farne l'*authority hub* per eccellenza (Van Veen 2019), altre risorse nel panorama del Semantic Web continuano ad offrire maggiore copertura in settori più vicini al patrimonio culturale. Tra queste c'è VIAF¹³ (Virtual International Authority File), uno strumento fondamentale per l'identificazione di entità legate all'universo bibliografico (Angeli et al. 2014). La versione 2.0 di CLEF integra VIAF nella lista di suggerimenti qualora una stringa di input non dovesse restituire risultati in Wikidata (Figura 3).

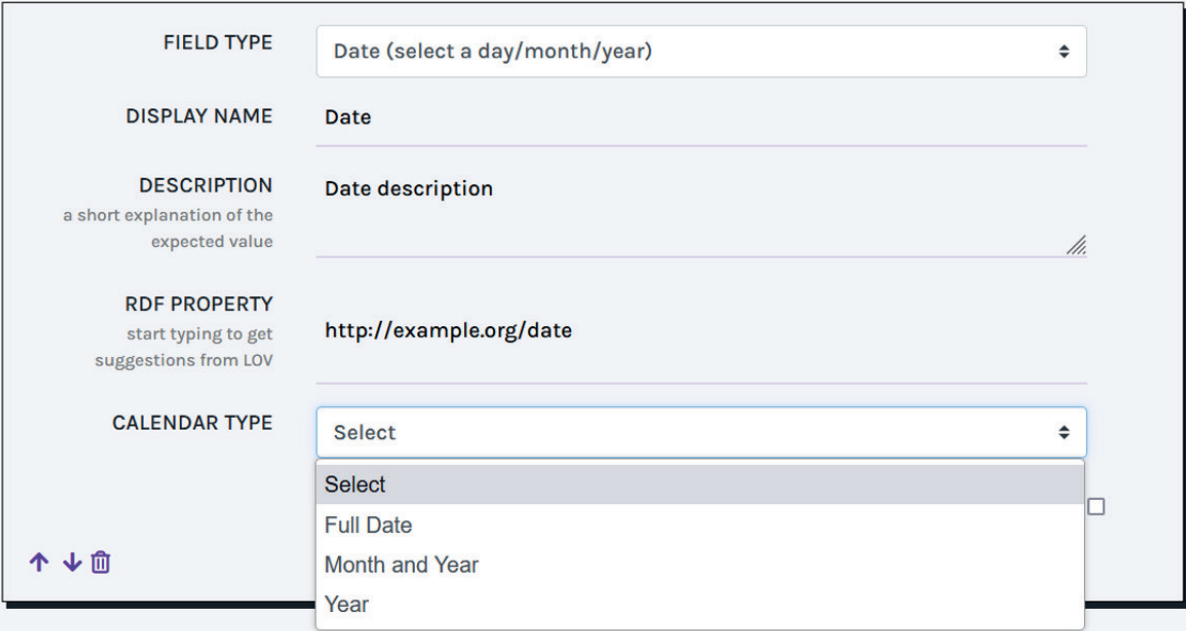
¹³ <https://viaf.org/>

The image shows two parts of the CLEF 2.0 interface. The top part is a configuration panel for an RDF property. It includes a text input field with the URL 'http://schema.org/byArtist', a 'VALUE TYPE' dropdown menu currently set to 'Entity (URI from Wikidata, VIAF, or catalogue)', and a 'PLACEHOLDER' field. The dropdown menu is open, showing options: 'Select', 'Free text (Literal)', 'Entity (URI from Wikidata, VIAF, or catalogue)' (highlighted), 'Location (from geonames)', and 'URL'. The bottom part shows a record preview for 'NOME OPERA *'. The 'O.T.' field is empty. The 'ARTISTA' field contains 'saribat' and a dropdown menu with suggestions: 'Šaribat, Šaribat Ahmad 1957-', 'Saribatur, Zekeriya', 'Saribatur, Zekerya 1958-', and 'Saribatur Yaman, Zeynep Gözen'. To the right, a larger view of the record shows 'O.T.' as empty, 'NOME OPERA' as empty, 'ARTISTA' as 'Saribatur, Zekeriya', and 'DIMENSIONI OPERA' as '100x150 cm'.

Figura 3. In alto, il dropdown con i vari *value type* disponibili per il campo di tipo *textbox*. In basso, un esempio di selezione di un'entità VIAF tramite suggerimenti automatici (a sinistra) e la relativa visualizzazione all'interno del record prodotto (a destra).

Nel caso di studio KNOT risulta fondamentale anche la possibilità di menzionare l'URL di risorse web collegate. Gli URL vengono gestiti diversamente dalle entità recuperate da Wikidata, VIAF, catalogo e GeoNames, per le quali CLEF prevede un'opzione specifica. Per supportare questo tipo di input, CLEF 2.0 integra un *value type* da associare a *textbox* in fase di creazione di un Template (Figura 3). Anche in questo caso, la risorsa indicata è immediatamente raggiungibile dalla scheda catalografica di riferimento.

Datatype temporali. Nella sua prima versione, CLEF prevede solo l'input di URI e valori aventi datatype `xsd:string`. Per far fronte ad esigenze di maggiore accuratezza descrittiva, l'aggiornamento introduce tre nuovi datatype per la definizione di datazioni: `xsd:Date` (YYYY-MM-GG), `xsd:gYearMonth` (YYYY-MM) e `xsd:gYear` (YYYY) (Figura 4).



The screenshot shows a configuration form for a 'Date' field in CLEF. The form is organized into several sections:

- FIELD TYPE:** A dropdown menu with the selected value 'Date (select a day/month/year)'.
- DISPLAY NAME:** A text input field containing 'Date'.
- DESCRIPTION:** A text input field containing 'Date description'.
- RDF PROPERTY:** A text input field containing 'http://example.org/date'.
- CALENDAR TYPE:** A dropdown menu with 'Select' selected. A dropdown menu is open below it, showing the following options: 'Select', 'Full Date', 'Month and Year', and 'Year'.

At the bottom left of the form, there are three small icons: an upward arrow, a downward arrow, and a trash can icon.

Figura 4. creazione di un Date field in CLEF.

Vocabolari controllati e thesauri SKOS. CLEF 2.0 ha introdotto la possibilità di creare campi dedicati all’inserimento di termini provenienti da vocabolari controllati selezionando il vocabolario di riferimento da una lista limitata di risorse già integrate¹⁴, fondamentali per la descrizione di progetti di Digital Humanities all’interno del progetto KNOT. Inoltre, viene data la possibilità di aggiungere nuovi termini effettuando una query SPARQL ad un servizio online. In fase di definizione del template si inserisce l’URL della risorsa, l’endpoint e una query SPARQL per estrarre termini sulla base di una stringa di input. In fase di inserimento dati, all’utente viene mostrata una lista di suggerimenti automatici in maniera del tutto analoga al meccanismo previsto per le entità estratte da Wikidata, VIAF e catalogo (Figura 5). Tuttavia, poiché l’autocompletamento non garantisce una visuale completa delle risorse terminologiche disponibili, agli utenti vengono forniti dei collegamenti rapidi ai vocabolari SKOS selezionati, agevolandone la consultazione.

¹⁴ Vocabolario delle Licenze (<https://schema.gov.it/lodview/controlled-vocabulary/licences>); authority tables Data Theme, File Type, Access Right, Language, and Frequency (<https://op.europa.eu/en/web/eu-vocabularies/authority-tables>); Taxonomy of Digital Research Activities in the Humanities (<https://vocabs.dariah.eu/tadirah/en/>).

The screenshot is divided into four quadrants. Top-left: A list of default vocabularies with checkboxes: TADIRAH, FREQUENCY, LANGUAGE, FILE-TYPE, DATA-THEME, ACCESS-RIGHT, LICENSES, and an 'ADD A NEW VOCABULARY +' button. Top-right: A form for adding a new vocabulary with fields for 'label for the new vocabulary', 'vocabulary's webpage', 'SPARQL query to get a label...', and 'SPARQL query endpoint'. A 'Licenses' section contains a SPARQL query: `PREFIX skos: <http://www.w3.org/2004/02/skos/core#> select distinct ?label ?uri where { ?uri skos:inScheme <https://w3id.org/italia/controlled-vocabulary/licences>. ?uri skos:prefLabel ?label.}` and an 'Add Vocabulary' button. Bottom-left: A project creation form for 'ALCIDE (Analysis of Language and Content In a Digital Environi...)' with an 'AMBITO DI RICERCA' section showing a search for 'an' and suggestions for 'Analyzing - TADIRAH' and 'Annotating - TADIRAH'. Bottom-right: A final visualization of the 'ALCIDE' project with 'ANALYZING' selected under 'AMBITO DI RICERCA'.

Figura 5. In alto, l'elenco dei vocabolari disponibili di default in CLEF 2.0 (a sinistra) e il form per l'aggiunta di nuovi thesauri per la produzione di suggerimenti automatici (a destra). In basso, suggerimenti automatici durante la creazione di un record (a sinistra) e relativa visualizzazione finale (a destra).

Multimedia. CLEF 2.0 introduce due nuove funzionalità (Multimedia e Preview) che consentono agli utenti di inserire il riferimento URL a file multimediali presenti in rete (immagini, audio o video) e preview di pagine web esterne. In CLEF non è possibile importare file, ma è necessario garantirne la conservazione altrove. Tale restrizione dipende da una scelta di design: CLEF si propone come strumento agile per la catalogazione nativa Linked Data, non per la creazione di cataloghi multimediali. Per ovviare a tale limitazione, CLEF consente nativamente di selezionare quali URL si riferiscono ad una risorsa digitale online per la quale non si ha a disposizione una strategia di preservazione a lungo termine. In tali casi, il sistema provvede a inviare una richiesta di deposito di uno snapshot della risorsa web in questione su Internet Archive. In questo modo viene garantita agli utenti la possibilità di linkare qualsiasi file, senza porre limiti di durata, dimensione o preservazione a lungo termine.

Knowledge extraction. La creazione di Linked Data si rivela un processo dispendioso quando l'inserimento dei valori va svolto manualmente (es. tutte le persone menzionate in un documento). I tool esistenti si sono occupati solo parzialmente di offrire un vero e proprio strumento per l'estrazione automatica di dati. La principale novità in CLEF 2.0 coincide quindi con l'introduzione

di un prototipo di Knowledge extraction, progettato per l'estrazione automatica di entità a partire da una varietà di fonti online: API, endpoint SPARQL e file statici in formato JSON e CSV. Questa funzionalità ha richiesto lo sviluppo di funzioni anche complesse per gestire correttamente le possibili casistiche. Al momento, si è scelto di limitare l'utilizzo di questo strumento all'estrazione di liste di URI e relative *label*, utilizzando una proprietà di default (schema:keywords) per relazionare il record alle entità estratte. L'interrogazione a un servizio API richiede tre input da parte dell'utente: l'URL dell'API, i parametri della richiesta AJAX e percorso per identificare URI ed etichette nella risposta (JSON) fornita dall'API. Più semplice è invece l'interrogazione di un endpoint SPARQL, che necessita soltanto di un URL e di una query da eseguire (Figura 6). Per interrogare i file statici, CLEF 2.0 ricorre all'utilizzo di SPARQL Anything (Daga et al. 2021). SPARQL Anything funziona come un tipico endpoint SPARQL, con la particolarità di poter generare dati RDF a partire da una vasta gamma di fonti di dati non nativamente RDF. L'utente deve specificare l'URL della risorsa desiderata e una query SPARQL valida (Figura 7).

THANKS FOR HELPING CLEF TO GROW!

EXTRACTOR TYPE: Static File

FILE URL: sparqlJSONjson

SPARQLANYTHING QUERY: SELECT ?uri ?label WHERE { SERVICE <x-sparql-anything:sparqlJSONjson> { ?value xyz:label ?label_info . ?value xyz:uri ?uri_info . ?label_info xyz:value ?label . ?label_info xyz:xml%3Alang "en" . ?uri_info xyz:value ?uri . } }

Buttons: Back, Next

Figura 6. Esempio di interrogazione SPARQL per l'estrazione di entità.

THANKS FOR HELPING CLEF TO GROW!

EXTRACTOR TYPE: SPARQL

SPARQL endpoint: https://schema.gov.it/sparql

QUERY: PREFIX skos: <http://www.w3.org/2004/02/skos/core#> select distinct ?label ?uri where { ?uri skos:inScheme <https://w3id.org/italia/controlled-vocabulary/licences> . ?uri skos:prefLabel ?label. FILTER (lang(?label) = "en"). FILTER(REGEX(?label, "c", "i")) } LIMIT 10

Buttons: Back, Next

Figura 7. Esempio di query SPARQL Anything per il recupero di entità da un file JSON.

Una volta estratti dei risultati, la lista di URI e label viene proposta all'utente, il quale può curarla manualmente e scegliere se eliminare alcune istanze. Nuove estrazioni possono essere effettuate, anche selezionando fonti diverse: per ciascuna di esse, il catalogo crea un nuovo grafo RDF in cui conservare i risultati ottenuti, l'autore e il metodo utilizzato per l'estrazione. Lo stesso grafo viene poi collegato a quello del record.

5. Discussione

L'aggiornamento di CLEF alla sua versione 2.0 ha tentato di rispondere ai requisiti dei casi di studio KNOT, ATLAS e GEL. Sebbene diverse problematiche avessero già trovato risposta in altri sistemi, il presente progetto vuole rafforzare la base offerta da CLEF, proponendo un sistema unitario, leggero e di facile utilizzo, in grado di soddisfare i requisiti essenziali nel panorama delle attività di catalogazione nel Web Semantico.

La più importante novità introdotta da CLEF 2.0 coincide con l'integrazione di uno strumento per la Knowledge Extraction. Si tratta senza dubbio di una funzionalità più avanzata rispetto alle altre qui illustrate, che cerca di soddisfare esigenze più complesse, vicine ad un pubblico esperto, come quello che si accinge a catalogare i progetti di Digital Humanities relativi al patrimonio culturale italiano (KNOT e ATLAS). La possibilità di estrarre dati da file statici è al momento il miglior compromesso per consentire ad utenti meno esperti di lavorare su risorse che richiedono meno competenze tecniche rispetto a servizi API ed endpoint SPARQL.

L'attuale implementazione richiede la conoscenza di SPARQL per estrarre i dati. Sebbene queste criticità possano apparire in aperto contrasto rispetto agli obiettivi di usabilità alla base di CLEF, i vantaggi del ricorso a SPARQL Anything rimangono preponderanti. In primo luogo, questo sistema consente l'utilizzo di uno strumento unico per l'analisi di più formati di file, evitando lo sviluppo *ad hoc* di nuove soluzioni ogni qualvolta un nuovo formato venga inserito. Al tempo stesso, l'impiego di un motore di query SPARQL trasferisce sull'analisi di file statici tutti i vantaggi di un linguaggio di interrogazione, garantendo un efficiente meccanismo per il filtraggio dei dati, fondamentale in presenza di documenti di grandi dimensioni. Inoltre, dal punto di vista dell'usabilità del sistema, SPARQL Anything richiede ai propri utenti la conoscenza di un unico linguaggio per l'interrogazione, semplificando i requisiti di conoscenza pregressa.

Un altro aspetto cruciale nell'aggiornamento proposto da CLEF 2.0 riguarda il ricorso a risorse multimediali online. Accade spesso che i documenti e i servizi richiesti non siano gestiti direttamente dagli utenti, sollevando quindi importanti riflessioni circa la preservazione in rete delle risorse in questione (Regino e Dos Reis 2022). Sulla base di queste considerazioni, appare logico sottolineare nuovamente l'importanza riservata da CLEF ad una strategia trasversale per la preservazione a lungo termine di risorse online.

6. Conclusioni

L'introduzione al presente articolo ha tentato di delineare le caratteristiche fondamentali dell'attuale scenario della creazione di collezioni di LOD, anticipando l'esigenza di un rinnovamento dei sistemi esistenti sulla base di nuovi requisiti. Dall'analisi dei principali sistemi per la catalogazione

collaborativa di Linked Data emergono i punti di forza di CLEF, l'applicazione web scelta come punto di partenza del presente lavoro. Sulla base dei casi di studio KNOT, ATLAS e GEL sono stati definiti i nuovi requisiti per l'aggiornamento di CLEF 2.0: per ciascuna delle nuove funzioni richieste è stata proposta una breve analisi delle soluzioni già esistenti sviluppate dagli altri principali CMS e si è dato risalto all'esigenza di rafforzare gli strumenti per il riutilizzo di entità estratte da basi di dati strutturate (es.: VIAF, thesauri SKOS) e non strutturate (SPARQL Anything), che costituiscono attualmente una delle sfide più importanti nel panorama LOD. Al tempo stesso, particolare attenzione è stata posta sulla necessità di ampliare le tipologie di dati e le risorse multimediali, ponendo l'accento sulla loro preservazione a lungo termine.

Sebbene alcune criticità siano emerse nello sviluppo di soluzioni per Knowledge Extraction, si è cercato di analizzare i limiti, discutere i vantaggi e pensare a futuri sviluppi, mirati alla costruzione, empirica ed imperfetta, di un software trasparente per rispondere alle esigenze in evoluzione del patrimonio culturale digitale italiano. L'insieme delle criticità, dei traguardi e delle limitazioni emerse prepara di fatto il terreno per ulteriori analisi e nuovi sviluppi. La realizzazione di un'applicazione sempre più completa e di facile utilizzo resta l'obiettivo primario per tentare di portare i Linked Open Data “nelle tasche di tutti” (Pellegrino, Scarano, e Spagnuolo 2022).

7. Ringraziamenti

Questo progetto ha ricevuto finanziamento dall'Unione europea – Next Generation EU (PRIN2022, 20227M8RS7) e dal programma European Union's Horizon 2020 research and innovation (GA 101004746).

Riferimenti bibliografici*

Angjeli, Anila, Andrew MacEwan, e Vincent Boulet. 2014. “ISNI and VIAF – Transforming Ways of Trustfully Consolidating Identities.” <https://doi.org/10.13140/RG.2.1.1350.8640>.

Davis, Edie, e Bahareh Heravi. 2021. “Linked Data and Cultural Heritage.” *Journal on Computing and Cultural Heritage* 14 (2): 1-18. <https://doi.org/10.1145/3429458>.

Daga, Enrico, Luigi Asprino, Paul Mulholland, e Aldo Gangemi. 2021. “Facade-X: An Opinionated Approach to SPARQL Anything.” *Studies on the Semantic Web* 53: 58-73. <https://doi.org/10.3233/ssw210035>.

Daquino, Marilena. 2021. “Linked Open Data Native Cataloguing and Archival Description.” *JLIS.it* 12 (3): 91-104. <https://doi.org/10.4403/jlis.it-12703>.

Daquino, Marilena, Mari Wigham, Enrico Daga, Lucia Giagnolini, and Francesca Tomasi. 2023. “CLEF. A Linked Open Data Native System for Crowdsourcing.” *Journal on Computing and Cultural Heritage*. Association for Computing Machinery (ACM). <https://doi.org/10.1145/3594721>.

Diefenbach, Dennis, Max De Wilde, e Samantha Alipio. “Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph.” In *The Semantic Web – ISWC 2021*, a cura di Andreas Hotho, Eva Blomqvist, Stefan Dietze, Achille Fokoue, Ying Ding, Payam Barnaghi, Armin Haller, Mauro Dragoni, e Harith Alani, 12922: 631–47. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-88361-4_37.

Hawkins, Ashleigh. “Archives, Linked Data and the Digital Humanities: Increasing Access to Digitised and Born-Digital Archives via the Semantic Web.” *Archival Science* 22 (3): 319–44. <https://doi.org/10.1007/s10502-021-09381-0>.

Hermon, Sorin, e Franco Niccolucci. 2021. “FAIR Data and Cultural Heritage Special Issue Editorial Note.” *International Journal on Digital Libraries* 22 (3): 251-55. <https://doi.org/10.1007/s00799-021-00309-8>.

Marden, Julia, Carolyn Li-Madeo, Noreen Whysel, e Jeffrey Edelstein. “Linked Open Data for Cultural Heritage: Evolution of an Information Technology.” In *Proceedings of the 31st ACM International Conference on Design of Communication*, 107–12. Greenville, North Carolina (USA): Association for Computing Machinery, 2013. <https://doi.org/10.1145/2507065.2507103>.

Mitchell, Sonia Natalie, Andrew Lahiff, Nathan Cummings, Jonathan Hollocombe, Bram Boskamp, Ryan Field, Dennis Reddyhoff, Kristian Zarebski, Antony Wilson, Bruno Viola, Martin Burke, Archibald Blair, Paul Bessell, Richard Blackwell, Lisa A. Boden, Alys Brett, Sam Brett, Ruth Dundas, Jessica Enright, Alejandra N. Gonzalez-Beltran, Claire Harris, Ian Hinder, Christopher David Hughes, Martin Knight, Vito Mano, Ciaran McMonagle, Dominic Mellor, Sibylle Mohr, Glenn Marion, Louise Matthews, Iain J. McKendrick, Christopher Mark Pooley, Thibaud Porphyre, Aaron Reeves, Edward Townsend, Robert Turner, Jeremy Walton, e Richard Reeve. 2022. “FAIR Data Pipeline: Provenance-Driven Data Management for Traceable Scientific Work-

* Tutti i siti web hanno come data di ultima consultazione il 7 maggio 2024.

flows.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 380 (2233): 1-23. <https://doi.org/10.1098/rsta.2021.0300>.

O’Hara, Kieron, Harith Alani, Yannis Kalfoglou, e Nigel Shadbolt. 2004. “Trust Strategies for the Semantic Web.” In *Workshop on Trust, Security, and Reputation on the Semantic Web, 3rd International Semantic Web Conference (ISWC04), 7-11 November 2004, Hiroshima Prince Hotel, Hiroshima, Japan*. <https://ceur-ws.org/Vol-127/paper5.pdf>.

Oldman, Dominic, e Diana Tanase. 2018. “Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace.” In *The Semantic Web – ISWC 2018*, a cura di Denny Vrandečić, Kalina Bontcheva, Mari Carmen Suárez-Figueroa, Valentina Presutti, Irene Celino, Marta Sabou, Lucie-Aimée Kaffee, e Elena Simperl, 11137: 325–40. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-00668-6_20.

Pellegrino, Maria Angela, Vittorio Scarano, e Carmine Spagnuolo. 2022. “Move Cultural Heritage Knowledge Graphs in Everyone’s Pocket.” *Semantic Web* 14 (2): 323-59. <https://doi.org/10.3233/sw-223117>.

Poblet, Marta, Pompeu Casanovas, e Víctor Rodríguez-Doncel. 2019. *Linked Democracy: Foundations, Tools, and Applications*. Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-030-13363-4>.

Regino, André Gomes, e Julio Cesar Dos Reis. 2022. “Leveraging Linked Open Data: A Link Maintenance Framework.” In *Anais Estendidos Do XXVIII Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia 2022)*, 15–18. Curitiba: Sociedade Brasileira de Computação. https://doi.org/10.5753/webmedia_estendido.2022.225651.

Sarasua, Cristina, Elena Simperl, Natasha F. Noy, Abraham Bernstein, e Jan Marco Leimeister. 2015. “Crowdsourcing and the Semantic Web: A Research Manifesto.” *Human Computation* 2 (1): 3-17. <https://doi.org/10.15346/hc.v2i1.2>.

Thanos, Costantino. 2017. “Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies.” *Publications* 5 (1): 1-19. <https://doi.org/10.3390/publications5010002>.

Tomasi, Francesca. 2023. “Archival Finding Aids in Linked Open Data Between Description and Interpretation.” *JLIS.it* 14 (3): 134-46. <https://doi.org/10.36253/jlis.it-557>.

Van Veen, Theo. 2019. “Wikidata: From ‘an’ Identifier to ‘the’ Identifier.” *Information Technology and Libraries* 38 (2): 72–81. <https://doi.org/10.6017/ital.v38i2.10886>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C ’t Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wit-

tenburg, Katherine Wolstencroft, Jun Zhao, e Barend Mons. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1): 1–9. <https://doi.org/10.1038/sdata.2016.18>.

Zaytseva, Ksenia, e Matej Ďurčo. 2020. *Controlled Vocabularies and SKOS*. <https://campus.dariah.eu/id/D8d6OrLdpLlGRqBSQDVN0>.