



# Utilizzo del DOI (*Digital Object Identifier*) per la diffusione di progetti lessicografici digitali

Giovanni Salucci

Nei progetti lessicografici digitali viene consigliato di utilizzare gli Identificatori persistenti. In questo contributo si esplora l'opportunità di utilizzare il DOI (*Digital Object Identifier*) come strumento per la diffusione e promozione di un progetto lessicografico digitale, usando Crossref come agenzia di registrazione. Occorre registrare una serie di DOI, in corrispondenza dei vari livelli gerarchici con cui la banca-dati lessicografica è organizzata, prevedendo la compilazione di metadati di qualità e ricchi di informazioni, con l'obiettivo di identificare il sistema più ampio di metadati che possa favorire la diffusione del progetto e massimizzarne l'impatto. Nell'articolo viene quindi analizzato in dettaglio il tracciato di registrazione del DOI, mettendo in evidenza le informazioni necessarie e consigliate per la diffusione, esemplificando come collocarle nel sistema di tag previsti dallo schema di registrazione.

*In digital lexicographic projects, the use of persistent identifiers is recommended. This contribution explores the opportunity to adopt Digital Object Identifiers (DOIs) as a tool for the dissemination and promotion of a digital lexicographic project, utilizing Crossref as the registration agency. To achieve maximum dissemination, a series of DOIs need to be registered, corresponding to the various hierarchical levels through which the lexicon database is organized. This necessitates the compilation of high-quality metadata that is rich in information.*

*This article provides a detailed analysis of the DOI registration process, highlighting the necessary and recommended information for dissemination. It exemplifies how to incorporate this information into the tag system specified by the registration schema.*

**Parole chiave:** DH, Informatica umanistica, DOI, Metadati, Identificatori persistenti, Lessicografia

**Keywords:** DH, Digital Humanities, DOI, Metadata, Persistent IDentifiers, PID, Lexicography

**Sommario:** Introduzione - 1. Il tracciato di registrazione - 2. Come scegliere gli elementi necessari per ottenere metadati di qualità - 3. Le informazioni da gestire in un progetto lessicografico - 3.1 Livello del database - 3.2 Livello del *dataset* - 3.3 Livello del record - 4. Referenze bibliografiche - 5. Relazioni tra gli oggetti attraverso le relazioni tra i DOI - 6. Conclusioni e sviluppi futuri

## Peer review

Submitted 07/05/2023

Accepted 13/06/2023

Published 11/07/2023

## Open access

© 2023 | Attribution - Non commercial - Non derivatives (IT)

**Cita come** Giovanni Salucci, *Utilizzo del DOI (Digital Object Identifier) per la diffusione di progetti lessicografici digitali* in Rivista DILEF - III, 2023/3 (gennaio-dicembre), pp. 275-292.

**DOI**

10.35948/DILEF/2024.4327  
10.35948/DILEF/2024.4327

## Introduzione

In alcuni recenti progetti lessicografici nativi digitali, quali ad esempio il VoDIM (Biffi 2018) o il VoSCIP (Bertini 2019), l'attività lessicografica trae beneficio da numerose applicazioni software e procedure informatiche che supportano l'attività dei ricercatori nelle varie fasi di lavoro. Il risultato finale prevede la pubblicazione dei risultati nella forma di articoli in rivista e monografie, come accade in tutti i progetti di ricerca scientifica e nei progetti lessicografici tradizionali, ma oltre a questo sin dalla ideazione del progetto è previsto il rilascio di un sito internet e di una banca dati di schede lessicografiche interrogabili da un motore di ricerca più o meno complesso e strutturato. Infatti, nel caso di progetti lessicografici nativi digitali, è proprio l'insieme del motore di ricerca e della restituzione a video delle varie schede lessicografiche, organizzate secondo le specifiche di ciascuna impresa lessicografica, a costituire il risultato più rappresentativo dal punto di vista scientifico; purtroppo è difficoltoso collocare questo risultato digitale ai fini della valutazione scientifica e della progressione della carriera accademica in Italia e all'estero, in un contesto che privilegia articoli e monografie (Eve 2022). Il superamento di questo paradigma rappresenta una delle sfide ma anche opportunità più rilevanti della "Open Science" nel settore delle *digital humanities* (Longley 2021).

Il DOI (*Digital Object Identifier*) è un sistema interattivo per l'identificazione persistente e lo scambio interoperabile di informazioni in rete (Paskin 2010) e tra le varie opzioni di identificazione, il DOI è certamente la più diffusa (Klump 2017); nei progetti di *digital humanities* il DOI si può applicare a risorse di differenti tipologie, principalmente per dividerne le proprietà all'interno di un certo ambito scientifico, o per identificarle univocamente in vista della gestione della proprietà intellettuale (Salucci 2022).

I campi obbligatori (e quindi indispensabili) da compilare per la registrazione del DOI sono davvero pochi, e facilmente gestibili, e questo rappresenta un vantaggio in tutte quelle situazioni dove interessa registrare il DOI al solo fine della identificazione persistente, come in alcuni progetti di database genetici o biologici (Arend 2016). La situazione si complica, e non poco, qualora invece si voglia utilizzare il DOI non solo per "certificare" e rendere riconoscibile in modo univoco un progetto, ma si intenda davvero sfruttare le potenzialità offerte dagli aggregatori accademici e dalle banche dati specialistiche per diffonderlo. In questo caso, la completezza e ricchezza dei metadati con cui si registrano i DOI sono fondamentali, dato che gli aggregatori interrogano le *Application Programming Interface*<sup>1</sup> (API) del DOI per attingere le informazioni loro necessarie, prelevandole dai metadati di registrazione.

Occorre quindi in partenza registrare il DOI prestando particolare attenzione alla definizione dei dati e alla loro compilazione, creando metadati di qualità. E in questo caso, anche la quantità di informazioni trattata diventa rilevante, visto che questi

metadati sono pubblicati ed accessibili con licenze CC0<sup>2</sup>, secondo i principi FAIR (Wilkinson et al. 2016). I metadati, diffusi o interrogabili attraverso le API, saranno proprio la base su cui lavoreranno le procedure di indicizzazione e ricerca dei vari aggregatori e motori di ricerca accademici, che utilizzeranno in modo particolare gli identificativi persistenti per collegare le differenti risorse (Ananthakrishnan 2020). Ricchezza, completezza e affidabilità sono dunque le parole chiave per garantire metadati di qualità, e da questi, per promuovere e diffondere le risorse collegate, facilitando la generazione di grafici che mappano le connessioni tra differenti entità all'interno della ricerca scientifica accademica (Macgregor 2023).

In questo contributo si propone l'adozione del DOI nei progetti lessicografici digitali, analizzando il tracciato di registrazione con l'obiettivo di identificare il sistema più ampio di metadati che possa favorire la diffusione del progetto e massimizzarne l'impatto. Questa proposta prende in carico le istanze delle raccomandazioni circa l'uso degli identificatori persistenti nei progetti di ricerca (McMurry 2017, Plomp 2020, Suominen 2021) fornendone un'applicazione operativa. Insieme alle valutazioni generali saranno introdotti alcuni esempi concreti, prestando particolare attenzione alle scelte tecnico-informatiche rispetto al tracciato XML; sarà analizzato nel dettaglio lo schema dei metadati per la registrazione del DOI suggerendo quali sezioni e *tag* utilizzare al fine della più ampia diffusione di un progetto di lessicografia digitale. Infine, per motivi di spazio, non saranno in questa sede presi in considerazione gli aspetti economici e le risorse necessarie per realizzare quanto proposto.

## 1. Il tracciato di registrazione

Per la registrazione dei DOI nei progetti lessicografici digitali, avendo l'intenzione di favorire la qualità e granularità dei dati, l'agenzia Crossref rappresenta la scelta migliore, soprattutto in presenza di un numero abbastanza ridotto di registrazioni (Salucci 2022). Inoltre, Crossref e Datacite (l'altra agenzia di registrazione DOI specializzata nella gestione dei *dataset*) collaborano fattivamente al progetto Scholix che prevede l'adozione di linee guida comuni per lo scambio link e l'interoperabilità tra risorse della letteratura scientifica e dataset della ricerca scientifica<sup>3</sup>.

Al momento della stesura del presente articolo, lo schema corrente è il 5.3.1 disponibile in rete sul sito di Crossref<sup>4</sup>.

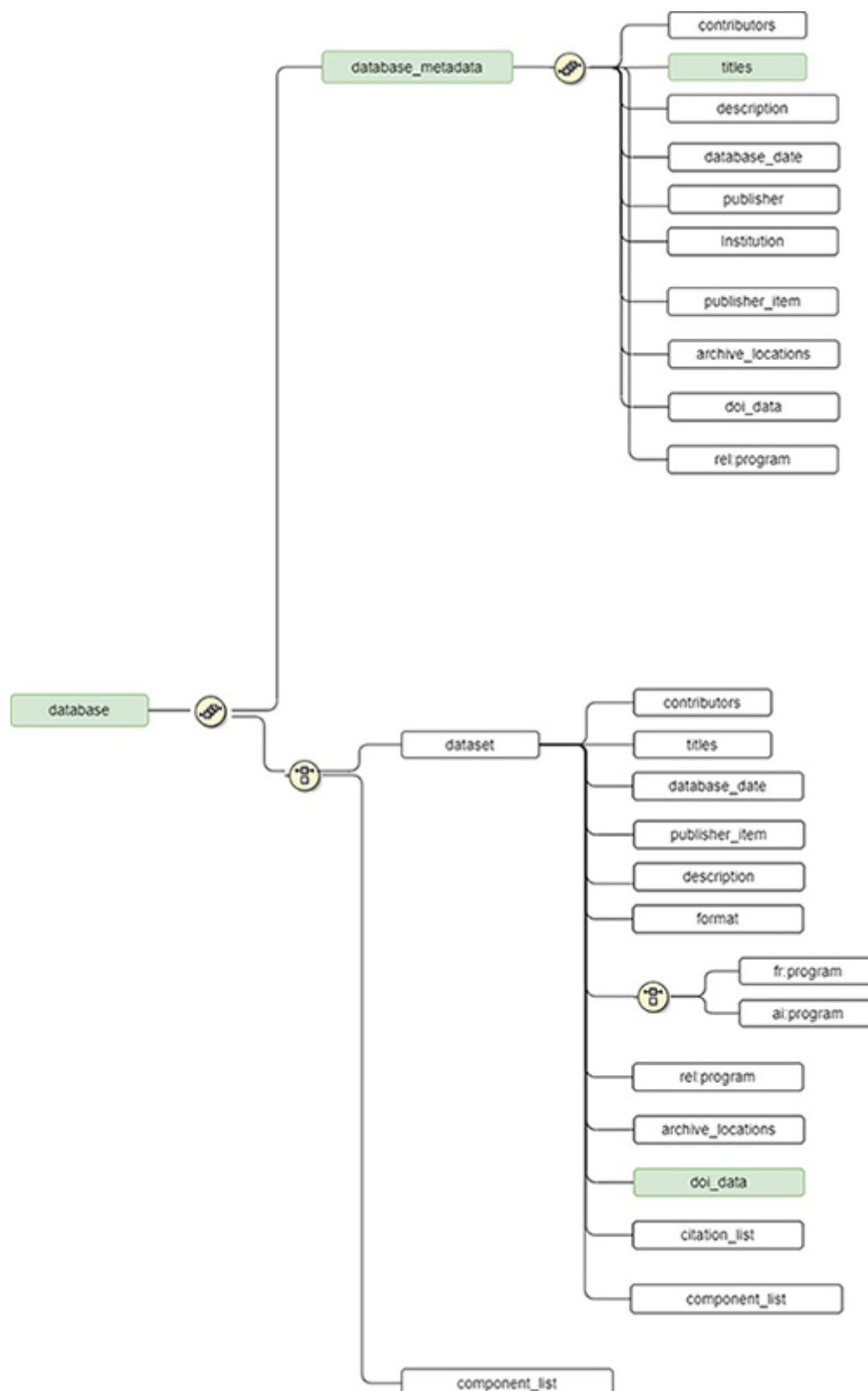


Fig. 1: la struttura degli elementi del tracciato dei metadati per il **database**.

Come noto, Crossref consente la registrazione del DOI per diverse tipologie di materiali. Nel caso di progetti lessicografici, la tipologia di registrazione da utilizzare è

il “**dataset**”: sul sito web è pubblicata una pagina<sup>5</sup> con alcune indicazioni operative per la compilazione dei metadati, e un file XML di esempio. In linea con quanto indicato nella documentazione ufficiale, con questa tipologia di registrazione è possibile descrivere le informazioni riguardo a uno o più *record* di database, o collezioni di *record*; ovviamente il deposito del “**dataset**” in Crossref fa riferimento solamente al deposito dei metadati, e non al contenuto del singolo *record* o dell’intero database a cui si fa riferimento; questo vale in generale per tutte le agenzie del DOI, e non solo per Crossref: le agenzie di registrazione del DOI non si occupano della archiviazione o preservazione delle risorse digitali, ma solamente di attribuire un identificatore persistente, attraverso il deposito dei metadati previsti dai differenti tracciati di registrazione.

Analizzando dunque le caratteristiche fondamentali del tracciato Crossref di registrazione, si può notare che lo schema XML per la registrazione di qualunque DOI è organizzato nei seguenti blocchi:

<doi\_batch> è il contenitore per la trasmissione di tutta la registrazione; qui si specificano anche i *namespaces* dei vari schemi (o profili) XML utilizzati.

All’interno di questo blocco sono presenti:

- <head> contiene le informazioni generali ad uso tecnico per la registrazione, comuni a tutte le registrazioni DOI con Crossref;
- <body> è il contenitore di tutti i metadati identificativi.

All’interno del <body> sono presenti differenti strutture di *tag* proprio per descrivere le diverse tipologie di oggetti per i quali sono ammesse le registrazioni. Nel caso dei progetti di lessicografia, il *tag* contenitore da utilizzare è “**database**”; nello schema sottostante (Fig. 1) viene rappresentata la struttura logica del tracciato dei metadati per il **database**. Si noti che in verde sono indicati i *tag* obbligatori per la registrazione.

- <database> contiene le informazioni generali sulla banca dati, o sull’insieme di più collezioni: agisce quindi come eventuale raccoglitore logico di più collezioni, o gruppi di schede;
- <database\_metadata> contiene i metadati descrittivi dell’intero database;
- <dataset> è il blocco di informazioni descrittive della *Collezione di record* o del singolo *record*.

È un blocco ripetibile; il valore assegnato all’attributo “type\_collection” ne caratterizza il livello di descrizione. Se il DOI si riferisce all’intera raccolta o ad un gruppo di schede, occorre assegnare il valore “**collection**”; se fa riferimento ad una singola scheda, esso assume il valore “**record**”. In entrambi i casi, ciascun

blocco `<dataset>` contiene un'ampia raccolta di metadati, in quanto è proprio a questo blocco che corrisponde la registrazione del DOI.

`<component_list>` è un blocco di informazioni strutturate per un gruppo di componenti, solitamente utilizzato per descrivere e identificare materiali a corredo, quali figure, tabelle o altre parti o informazioni che siano rilevanti per la citabilità. Questo blocco può essere inserito in varie posizioni nello schema, a seconda del livello di annidamento della entità a cui fanno riferimento.

## 2. Come scegliere gli elementi necessari per ottenere metadati di qualità

Analizzando l'unico esempio di registrazione di un **“dataset”** presente sul sito di Crossref in formato XML fornito insieme alla documentazione ufficiale, si nota che esso contiene una quantità non troppo estesa di informazioni; tutte le indicazioni minime obbligatorie previste dal tracciato sono ovviamente rispettate, ma questo esempio non può essere preso a modello per gli scopi di promozione e diffusione, proprio perché si tratta di un modello generico, e quindi mancano indicazioni su come coprire tutta la casistica articolata che si presenta in un progetto di banca dati lessicografica.

Osservando lo schema presentato in precedenza (Fig. 1) si può subito notare a colpo d'occhio come i campi obbligatori, e quindi necessari per la registrazione del DOI, siano davvero pochi (titolo, dati per il DOI e qualche altro dato tecnico; nella figura sono evidenziati in verde chiaro) e facilmente gestibili; non è quindi difficile organizzarsi per registrare DOI con i dati minimi. Passare dai dati obbligatori a quelli “raccomandati” per la valorizzazione è una operazione più complessa, visto che anche il tracciato fornito di esempio non è esaustivo. Per effettuare questa operazione di individuazione dei metadati di qualità, occorre quindi tornare ad analizzare nel dettaglio il tracciato di registrazione, per valutare quale potrebbe essere lo schema da raccomandare (e quindi ben più esteso sia di quello obbligatorio sia di quello presente nel file XML di esempio) per la registrazione di DOI da assegnare ad un progetto lessicografico digitale: un database al quale hanno lavorato più collaboratori, con differenti ruoli e responsabilità, realizzando un certo numero di schede lessicografiche, raggruppate in alcune collezioni.

In questa analisi, verrà seguito un approccio non convenzionale: invece che partire dal tracciato, dai vari *tag* disponibili e presentare lo scopo di ciascun *tag*, si preferisce partire dalle esigenze di “promozione” e di “diffusione” per arrivare a identificare nel tracciato dove inserire queste informazioni. Si dà quindi la priorità alle informazioni da trasmettere, evidenziando se e come quella informazione sia stata prevista nello schema e come debba essere gestita. Questo approccio ha anche il vantaggio che può essere facilmente adattato a vari contesti, esteso a situazioni di banche dati differenti



da quelle analizzate in questo articolo, oppure anche ridotto per renderlo più sostenibile qualora la proposta fosse ritenuta troppo dispendiosa o non adatta al contesto specifico.

### 3. Le informazioni da gestire in un progetto lessicografico

Un progetto lessicografico che arriva a costruire una raccolta di schede può essere rappresentato come una realtà almeno a due livelli:

1. il livello generale dell'intero progetto e quindi il livello del **database**;
2. il livello di dettaglio, rappresentato dalla singola scheda (e quindi il livello del **record**, cioè della singola scheda lessicografica).

Ma più frequentemente, nei progetti di ricerca finanziati dal MUR, collaborano più gruppi di lavoro, più sedi, più unità di ricerca; in questo caso è necessario introdurre un terzo livello, intermedio tra i due indicati in precedenza, costituito dalla raccolta di schede predisposte da ciascuna singola unità: si definisce quindi il livello intermedio della **collezione** di *record*.

Quindi lo scenario a tre livelli è quello più comune, così rappresentato gerarchicamente:

1. il livello del **database**;
2. il livello della **collezione**, e quindi il livello della raccolta di schede lessicografiche;
3. il livello del **record**, cioè della singola scheda lessicografica.

Questo scenario viene illustrato nella Figura sottostante (Fig. 2)

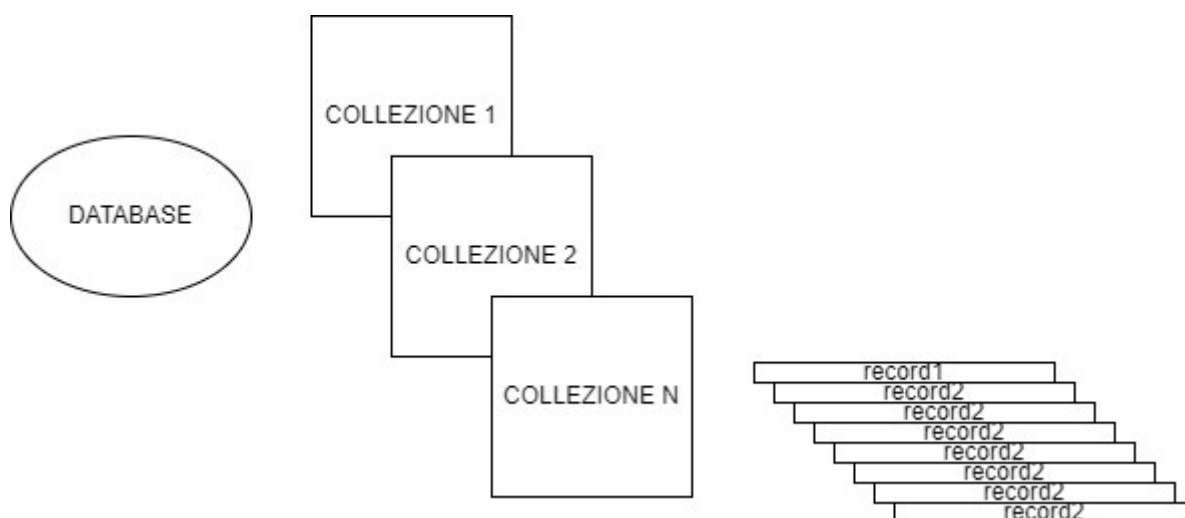


Fig. 2: la struttura tipica di un progetto lessicografico, a tre livelli di descrizione: il livello



globale del **database**, il livello intermedio rappresentato dalle **collezioni**; il livello di dettaglio costituito dalla singola scheda lessicografica (**record**).

Qualora ci siano tutti e tre questi livelli, è evidente che per la identificazione e assegnazione dei DOI si dovrebbe prevedere:

- un primo DOI (e relativi metadati) per il **database** nella sua interezza;
- un certo numero di DOI (e relativi metadati) per le varie collezioni, uno per ciascuna **collezione** (che corrisponde di solito alla raccolta di schede prodotte da ciascuna unità / gruppo di lavoro);
- un numero maggiore di DOI (e relativi metadati) uno per ciascuna scheda (**record**).

I metadati da assegnare e compilare per ciascun DOI possono quindi essere di volta in volta riempiti con informazioni (metadati) specifiche per quel particolare DOI, oppure provenire da scelte effettuate in precedenza, cioè essere “ereditati” dai livelli superiori; questo perché, come visto in precedenza, esiste una struttura gerarchica tra **database** / **collezione** / **record**.

Occorre fare una precisazione: il fatto che esista una gerarchia e che quindi sia possibile in certi casi ereditare certe informazioni dai livelli superiori non significa che nella compilazione dei metadati si possa tralasciare quelle informazioni. Ciascun set di metadati che si riferisca ad una certa entità, di qualunque livello sia, deve essere riempito completamente, in modo da essere totalmente autonomo. Nell'insieme dei metadati compilati è bene che contenga informazioni per collegarlo ad altre entità (le cosiddette relazioni) laddove abbia senso e/o sia opportuno. Anche per questo motivo, predisporre metadati ricchi di qualità è una attività dispendiosa, sia che sia effettuata manualmente, sia che preveda anche l'utilizzo di sistemi informatici che riducano il *data-entry*, grazie ad una gestione automatizzata dei valori ereditati.

### 3.1 Livello del database

Per il livello di **database** le informazioni rilevanti sono: titolo, informazioni descrittive, finanziatori, autori e curatori, copyright, licenze e riutilizzo, data di pubblicazione, editore e eventuali codici assegnati. Occorre individuare dove collocare queste informazioni nel tracciato. Il *tag* principale contenitore globale è `<database_metadata>` per il quale potrà essere specificato l'attributo della lingua.

All'interno di questo contenitore si potrà utilizzare:

- `<title>` per definire il titolo del progetto / banca dati;

- `<subtitle>` per gestire il sottotitolo o la seconda parte del titolo;
- `<description>` per una descrizione che si applichi all'intero progetto.

Per la datazione, è possibile usare il gruppo di date `<database_date>`.

In questo blocco, è possibile inserire tre differenti date, che si applicano all'intero database:

- `<creation_date>` la data di inizio del progetto;
- `<publication_date>` la data di prima pubblicazione (messa online) del progetto;
- `<update_date>` la data di ultima modifica effettuata in qualunque parte o *record* del database.

Per l'editore, esiste il blocco `<publisher>` al cui interno poter specificare:

- `<publisher_name>` il nome dell'editore;
- `<publisher_place>` il luogo di edizione.

Per le responsabilità (autori, curatori, ecc..) a questo livello di descrizione ci si riferisce all'intervento di chi ha progettato e definito il database, e a chi ne ha la responsabilità scientifica. Il blocco da utilizzare è `<contributors>`.

Ogni volta che viene inserito un soggetto (utilizzando il *tag* `<contributor>`), occorre specificarne l'ordine ("**first**" o "**additional**") e il ruolo, attraverso l'attributo obbligatorio "**contributor\_role**". I ruoli non sono arbitrari, è possibile sceglierli solamente da una lista normalizzata che contiene i seguenti valori: "**author**", "**editor**", "**chair**", "**reviewer**", "**review-assistant**", "**stats-reviewer**", "**reviewer-external**", "**reader**", "**translator**".

Ciascun soggetto può essere un Ente o una Persona fisica; nel primo caso si usa il blocco `<organization>` e se ne specifica il nome; nel caso di persona fisica si utilizza il blocco `<person_name>` e la situazione è più complessa e articolata; i dati da inserire sono organizzati in *tag* distinti, che prevedono:

- `<given_name>` il nome;
- `<surname>` il cognome;
- `<ORCID>` il codice ORCID che identifica il ricercatore.

Inoltre è anche obbligatorio inserire informazioni sulla affiliazione del ricercatore, utilizzando il blocco `<affiliations>` che deve contenere almeno un blocco `<institution>` a sua volta così organizzato:

- `<institution_name>` il nome intero della organizzazione di affiliazione;

- `<institution_id>` l'eventuale codice identificativo della organizzazione di affiliazione;
- `<institution_place>` la città dell'organizzazione di affiliazione;
- `<institution_department>` l'eventuale dipartimento di affiliazione.

Ovviamente, nella scelta della realizzazione dei metadati, le affiliazioni da inserire nei metadati sono tutte e sole quelle pertinenti in qualche modo al progetto; questo vale soprattutto nel caso di un nominativo che lavorasse su più sedi o collaborasse con varie organizzazioni.

Per quanto riguarda le informazioni riguardo ai finanziamenti ricevuti, alle licenze e al copyright è molto importante che queste informazioni siano inserite nei metadati; il tracciato prevede che questi dati stiano nel blocco del *dataset* e quindi saranno descritte in seguito.

Un'altra informazione utile potrebbe essere quella di indicare la presenza di un sistema di preservazione digitale per la banca dati; in questo caso si utilizza il blocco `<archive_locations>` che consente di utilizzare uno o più elementi `<archive>` corrispondenti a uno o più tra i *repository* ammissibili certificati per la preservazione a lungo termine. Al momento quelli ammessi sono: "CLOCKSS", "LOCKSS", "Portico", "KB", "Internet Archive", "DWT".

### 3.2 Livello del *dataset*

Potranno essere presenti da 1 a n *dataset*, sulla base della organizzazione del gruppo di lavoro e della banca dati realizzata. Ad ogni *dataset* dovrebbe corrispondere una "collezione" o "raccolta" omogenea di schede, che hanno in comune alcuni processi e/o un tema o soggetto. In molti casi le due cose coincidono, ad esempio quando le diverse unità partecipanti al progetto si dividono le schede secondo criteri cronologici e/o tematici; in questo caso è consigliabile predisporre un *dataset* per ciascuna unità di lavoro.

Per il livello di **dataset** le informazioni rilevanti sono: titolo, informazioni descrittive, finanziatori, autori e curatori, copyright, licenze e riutilizzo, data di pubblicazione. Occorre individuare dove collocare queste informazioni nel tracciato. Il *tag* principale contenitore globale è `<dataset>` per il quale dovrà essere specificato l'attributo **"dataset\_type"** con valore uguale a **'collection'**.

Per quanto riguarda titoli, descrizioni, autori e curatori valgono esattamente le stesse considerazioni e indicazioni precedentemente illustrate; in questo caso si tratta semplicemente di applicarle al livello delle collezione oggetto di descrizione.

Il livello *dataset* di collezione è pensato come un livello che corrisponde ad una pubblicazione a sé stante, e quindi è proprio nel blocco dataset che è possibile aggiungere informazioni sui finanziamenti, sulle condizioni di accesso e riutilizzo,

copyright. In dettaglio, i blocchi da utilizzare per registrare queste informazioni sono così rintracciabili: per il finanziamento c'è il modulo **fundref** che viene aggiunto allo schema base Crossref per tracciare i dati, utilizzando il blocco `<fr:program>` dentro cui sono inseriti i dati dei finanziatori usando `<fr:assertion name="fundgroup">` così strutturato:

- `<fr:assertion name="funder_name">` il nome della organizzazione che finanzia;
- `<fr:assertion name="funder_identifier">` il codice identificativo del finanziatore;
- `<fr:assertion name="award_number">` il nome o codice identificativo del programma di finanziamento.

Quest'ultimo è un dato molto importante, laddove comunicato dal finanziatore. L'inserimento di questo dato nella registrazione del DOI crea infatti la premessa per un tracciamento automatico del finanziamento collegato ai prodotti della ricerca, necessario alla rendicontazione amministrativa del finanziamento.

Per le licenze di accesso si deve utilizzare il modulo **accessIndicators** che verrà quindi aggiunto anch'esso allo schema base di Crossref.

Attraverso il blocco `<ai:program name="AccessIndicators">` è possibile indicare le modalità di accesso, se libero utilizzando il tag `<ai:free_to_read/>` altrimenti, usando il tag `<ai:license_ref>` è possibile indicare il link alla licenza scelta; il tag è ripetibile in quanto si possono definire più licenze, una per ciascuna "versione" usando l'attributo **"applies\_to"** e/o specificando la data di validità della applicabilità della licenza, attraverso l'attributo **"start\_date"**. Può essere possibile quindi stabilire che fino ad una certa data l'accesso sia limitato, mentre poi la risorsa diventi ad accesso libero a partire da una certa data.

### 3.3 Livello del record

Certamente saranno presenti, nel database, numerose schede. A ciascuna scheda corrisponderà un singolo DOI e quindi un proprio intero set completo di metadati, per la registrazione e diffusione.

Al livello di **record** si potrà optare per un insieme non eccessivamente ampio di metadati, visto che un'ampia descrizione del progetto e delle responsabilità è già stata definita nella registrazione della **collection** e del **database** cui il record appartiene. Tutte le informazioni ereditate dai livelli gerarchicamente superiori possono essere in qualche modo recuperati inserendo una **relazione**, come indicato di seguito, nello specifico paragrafo 5 dedicato alle relazioni.

Per il livello di **record** le informazioni rilevanti sono: titolo, informazioni descrittive, autori e curatori, licenze e riutilizzo, data di pubblicazione. Il *tag* principale contenitore globale è `<dataset>`, ma stavolta l'attributo “*dataset\_type*” avrà valore uguale a ‘**record**’.

Per quanto riguarda titoli, descrizioni, autori e curatori valgono esattamente le stesse considerazioni e indicazioni precedentemente illustrate; in questo caso si tratta semplicemente di applicarle al livello del singolo record, oggetto di descrizione.

#### 4. Referenze bibliografiche

Per avvantaggiarsi a pieno delle potenzialità offerte dalla registrazione del DOI, viene consigliato di provvedere alla predisposizione della bibliografia del progetto, intesa come l'elenco delle pubblicazioni da citare collegate in qualche modo al progetto, e di inserirla nei metadati. Nella predisposizione dell'elenco risulta fondamentale rintracciare (se presenti) i DOI delle singole referenze, proprio per garantire quei processi di circolarità della informazione e aggancio citazionale così utile per la diffusione e la tracciabilità del progetto stesso.

Il blocco per gestire le citazioni è `<citation_list>` che raggruppa tutte le referenze; ciascuna citazione è inserita in un *tag* `<citation>` a cui è attribuita una key univoca, e che contiene:

- `<doi>` l'identificativo del DOI di quella citazione
- `<unstructured_citation>` la citazione completa, scritta tutta di seguito secondo un certo formato omogeneo (regole editoriali per le citazioni).

Un aspetto interessante delle citazioni è che al momento della trasmissione dei metadati, Crossref effettua una verifica dei DOI inseriti nelle citazioni, e blocca la registrazione se uno o più DOI risultano non esistenti; d'altro lato, se il *tag* `<doi>` è stato lasciato vuoto, Crossref prova ad assegnarlo lui in automatico; utilizzando alcuni algoritmi e processi di intelligenza artificiale che scompongono la stringa inserita nel *tag* `<unstructured_citation>`, Crossref ricava i dati essenziali dell'opera e da questi prova ad agganciare il DOI che corrisponda a quei metadati con un buon grado di affidabilità.

Il blocco `<citation_list>` può essere inserito all'interno di `<dataset>`, sia nel caso di una *collection* che di un singolo *record*; evidentemente quali referenze bibliografiche inserire dipenderà dal livello di descrizione a cui faranno riferimento. A livello di *record*, infatti, occorrerà inserire esclusivamente referenze che trattano direttamente dell'oggetto collegato a quel *record*.

## 5. Relazioni tra gli oggetti attraverso le relazioni tra i DOI

L'ultimo argomento importante da considerare è quello delle relazioni. Come descritto in precedenza, ogni singolo DOI fa riferimento ad una distinta entità, che può corrispondere ad un singolo **record**, o una **collezione**, o addirittura ad un intero **database**. I metadati devono essere completi, ma allo stesso tempo non dovrebbero essere ridondanti oppure contenere informazione inutile o fuorviante.

Ad esempio, nel caso del DOI di un *record* del database, l'autore di quel *record* va inserito nel blocco <contributors>, ma come regolarsi per quanto riguarda il curatore dell'intero database? In generale, per non perdere le informazioni descrittive presenti ai livelli superiori, nel caso di registrazione di DOI dei singoli record, si può utilizzare una relazione, estendendo lo schema di registrazione con il blocco <rel:>. Collegando attraverso una relazione i due DOI, si collegheranno automaticamente i due oggetti cui i DOI riferiscono; dalla tipologia di relazione si potranno quindi desumere le proprietà da ereditare. Nel caso specifico, la relazione tra il *record* e il *database* a cui appartiene, sarà una relazione di appartenenza.

Per registrare queste relazioni, definite solitamente nei metadati delle entità di livello inferiore e che puntano ad entità di livello superiore, si utilizzerà l'elemento <rel:inter\_work\_relation> per il quale occorre specificare gli attributi **relationship-type** e **relations\_type**.

Il primo attributo (da scegliersi tra una lista controllata), serve ad indentificare il tipo di relazione.

I tipi di relazione (valori ammessi) sono: **isDerivedFrom**, **hasDerivation**, **isReviewOf**, **hasReview**, **isCommentOn**, **hasComment**, **isReplyTo**, **hasReply**, **basedOnData**, **isDataBasisFor**, **hasRelatedMaterial**, **isRelatedMaterial**, **isCompiledBy**, **compiles**, **isDocumentedBy**, **documents**, **isSupplementTo**, **isSupplementedBy**, **isContinuedBy**, **continues**, **isPartOf**, **hasPart**, **references**, **isReferencedBy**, **isBasedOn**, **isBasisFor**, **requires**, **isRequiredBy**, **finances**, **isFinancedBy**.

Il secondo attributo (anch'esso da scegliersi tra una lista controllata) serve per identificare il tipo di entità collegata dalla relazione.

I tipi di entità collegata sono: "doi", "issn", "isbn", "uri", "pmid", "pmcid", "purl", "arxiv", "ark", "handle", "uuid", "ecli", "accession", "other".

Si potrebbe dunque creare la relazione di appartenenza di un **record** alla **collezione** cui appartiene, attraverso una relazione così costruita:

```
<rel:program xmlns="http://www.Crossref.org/relations.xsd">
<rel:related_item>
<rel:description>Il record aaaaaaa appartiene alla collezione bbbbbbb
</rel:description>
```



```

<rel:inter_work_relation      relationship-type="isPartOf"      identifier-
type="doi">10.xxxxx/yyyyyy</rel:inter_work_relation>
</rel:related_item>
</rel:program>

```

Nell'esempio sopra indicato le stringhe evidenziate con i colori verde, celeste e giallo sono dei segnaposti e devono essere sostituite dai dati veri, rispettivamente: il titolo (nome) del *record*, il titolo (nome) della collezione e il DOI assegnato alla collezione. Analogamente si potrà definire una relazione tra la collezione e il database di appartenenza. In questo modo, nella creazione delle relazioni, basterà collegarsi al livello immediatamente superiore, e sfruttare le catene di relazioni che appunto possono indicare le tipologie di gerarchie attraverso l'attributo "relationship-type".

La presenza, all'interno del blocco relazioni, di un DOI collegato al DOI che si deve registrare, verrà utilizzato dagli aggregatori proprio per considerare collegate le due entità, secondo la tipologia di relazione indicata, alimentando quel sistema di collegamenti e puntamenti che favoriscono la rintracciabilità delle risorse, intesa contemporaneamente come identificazione univoca e reperibilità: grazie alla presenza dei due DOI sarà possibile passare dall'una all'altra semplicemente con un click. Questo è il grande vantaggio dell'utilizzo del sistema del DOI che è un sistema identificativo persistente interattivo.

## 6. Conclusioni e sviluppi futuri

In questo articolo sono state introdotte e analizzate le basi teorico-pratiche per l'adozione del DOI nei progetti digitali di redazione di schede lessicografiche.

Il DOI è un sistema di identificazione persistente degli oggetti digitali, fisici o logici, molto diffuso nell'editoria accademica e nell'archiviazione di *dataset* di esperimenti scientifici, ma scarsamente utilizzato sinora nei progetti di *digital humanities* con scopi di promozione e diffusione.

L'adozione del DOI utilizzando il tracciato minimo obbligatorio proposto dalle agenzie di registrazione è relativamente semplice; lo scenario si complica notevolmente se invece viene scelto di adottare la politica della predisposizione dei metadati di qualità; in questi casi occorre predisporre metadati accurati e più ampi possibile, che richiedono un considerevole sforzo concettuale e di gestione.

Nel presente articolo sono state introdotte gradualmente le informazioni rilevanti, ai vari livelli descrittivi, identificando nella documentazione ufficiale di Crossref quali elementi e attributi sono previsti per la predisposizione di un XML valido ai fini della registrazione ma che sia anche vantaggioso ai fini della diffusione e promozione; si rimanda ad un prossimo studio l'analisi delle implicazioni pratiche e dei risultati



---

misurabili che derivano dall'adozione di queste metodologie, con un'analisi dei costi-benefici di tale intervento.

---

## Note

1. Usualmente note tramite l'acronimo API, si tratta di un set di procedure software disponibili in un sistema informatico per dialogare con l'esterno; le API consentono generalmente di interrogare e manipolare i dati.
2. La licenza CC0 disponibile all'indirizzo <https://creativecommons.org/publicdomain/zero/1.0/> attesta il pubblico dominio, ovvero l'assenza di diritti d'autore.
3. Gli scopi del progetto Scholix e la sua descrizione sono reperibili all'indirizzo <http://www.scholix.org/>; all'indirizzo <https://www.eventdata.crossref.org/guide/app-scholix/> è disponibile la documentazione del tracciato di interrogazione e di scambio dati messo a disposizione da Crossref tramite le API relativo al progetto Scholix.
4. All'indirizzo [https://data.crossref.org/reports/help/schema\\_doc/5.3.1/index.html](https://data.crossref.org/reports/help/schema_doc/5.3.1/index.html) è disponibile la documentazione ufficiale dello schema dei metadati nella versione 5.3.1.
5. La documentazione ufficiale che presenta le linee guida e alcuni esempi per la registrazione del dataset con CROSSREF è disponibile al seguente indirizzo:  
<https://www.crossref.org/documentation/schema-library/markup-guide-record-types/datasets/>.

## Bibliografia

- Ananthakrishnan 2020 = Rachana Ananthakrishnan, Kyle Chard, Mike D'Arcy, Ian Foster, Carl Kesselman, Brendan McCollam, Jim Pruyne, Philippe Rocca-Serra, Robert Schuler, and Rick Wagner. 2020. *An Open Ecosystem for Pervasive Use of Persistent Identifiers*. In *Practice and Experience in Advanced Research Computing* (PEARC '20). ACM, 2020. DOI: <https://doi.org/10.1145/3311790.3396660>
- Arend 2016 = Daniel Arend, Astrid Junker, Uwe Scholz, Danuta Schuler, Juliane Wylie and Matthias Lange, *PGP repository: a plant phenomics and genomics data publication infrastructure*. Database Vol. 2016, (2016). DOI: <https://doi.org/10.1093/database/baw033>
- Bertini 2019 = Patrizia Bertini Malgarini, Marco Biffi, Ugo Vignuzzi, *Dal "Vocabolario storico della cucina italiana postunitaria" (VoSCIP) al "Vocabolario Dinamico dell'Italiano Moderno" (VoDIM): riflessioni di metodo e prototipi*. Studi di lessicografia italiana XXXVI:(2019), pp. 341-366.
- Biffi 2018 = Marco Biffi, *Strumenti informatico-linguistici per la realizzazione di un dizionario dell'italiano post-unitario*, in JADT '18. Proceedings of the 14th international conference on statistical analysis of textual data, a cura di Domenica Fioredistella Iezzi, Livia Celardo, Michelangelo Misuraca, Roma, Universitalia, vol. 1, pp. 99-107. ISBN: 978-88-3293-137-2
- Eve 2022 = Eve, Martin Paul, *Open Access in the Humanities Disciplines*. In: O'Sullivan, James (ed.) *The Bloomsbury Handbook of Digital Humanities*. London: Bloomsbury, (2022), pp. 223-231.
- Klump 2017 = Jens Klump e Robert Huber, *20 Years of Persistent Identifiers – Which Systems are Here to Stay?* Data Science Journal, 16: 9, 2017, pp. 1–7, DOI: <https://doi.org/10.5334/dsj-2017-009>
- Longley 2021 = Paul Longley Arthur , Lydia Hearn, *Toward Open Research: A Narrative Review of the Challenges and Opportunities for Open Humanities*, Journal of Communication, Volume 71, Issue 5, October 2021, pp. 827–853. DOI: <https://doi.org/10.1093/joc/jqab028>
- Macgregor 2023 = Macgregor, G., Lancho-Barrantes, B. S., & Rasmussen Pennington, D., *Measuring the concept of PID literacy: user perceptions and understanding of persistent identifiers in support of open scholarly infrastructure*. Open Information Science, 2023 (in press).
- McMurry 2017 = Julie A. McMurry, Nick Juty, Niklas Blomberg, Tony Burdett, Tom Conlin, Nathalie Conte et al., *Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data*. PLoS Biol 15(6): e2001414, 2017. DOI: <https://doi.org/10.1371/journal.pbio.2001414>
- Paskin 2010 = Norman Paskin, *Digital Object Identifier (DOI) System*, in *Encyclopedia of Library and Information Sciences*, Third Edition. Taylor & Francis, 2010.
- Plomp 2020 = Esther Plomp, *Going Digital: Persistent Identifiers for Research Samples, Resources and Instruments*. Data Science Journal, 19: 46, 2020 pp. 1–8. DOI: <https://doi.org/10.5334/dsj-2020-046>
- Salucci 2022 = Giovanni Salucci, *Utilizzo del DOI (Digital Object Identifier) nei progetti di digital humanities* in DILEF. Rivista digitale del Dipartimento di Lettere e Filosofia - 2 (2022), pp. 308-319. DOI: <https://doi.org/10.35948/DILEF/2023.4307>
- Suominen 2021 = Tommi Suominen, Joonas Nikkanen, Tuomas J. Alaterä, and Toni Sissala, *Linking SSH research publications, datasets and infrastructures in research.fi*. Proceedings of the ICTeSSH 2021 Conference. <https://doi.org/10.21428/7a45813f.28b75760>

- 
- Wilkinson et al. 2016 = Mark D. Wilkinson et al., *The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 3, 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>