

L'etnomusicologia nel web semantico

ILARIO MEANDRI, GIULIA FERDEGHINI, LIANNA D'AMATO

Abstract

Le tecnologie LOD hanno impiego in svariati progetti in ambito culturale a livello internazionale. Una prima sperimentazione di tali tecnologie nel campo dell'etnomusicologia è stata realizzata con il progetto SAMIC, che ha fornito le basi teoriche e tecnologiche per i progetti in *digital humanities* attualmente in corso. Nell'ambito del PRIN 2017, diretto da Giovanni Giuriati, l'Unità di Torino ha come obiettivi: 1) la realizzazione di un sistema catalografico per risorse etnomusicologiche; 2) la rappresentazione ontologica del dominio; 3) l'esposizione di dati catalografici in forma di Linked Open Data.

Lo scopo di questo articolo è fornire una panoramica generale sul tema dei LOD e sul loro utilizzo in ambito etnomusicologico e mostrare lo stato dell'arte del progetto e delle attività dell'Unità di Torino.

Ethnomusicology in the semantic web. *LOD technologies are used in a variety of projects in the cultural sphere at the international level. An initial experimentation of these technologies in the field of ethnomusicology was carried out with the SAMIC project, which provided the theoretical and technological basis for the current unito projects in digital humanities. As part of the PRIN 2017, directed by Giovanni Giuriati, the Turin Unit's objectives are: 1) the creation of a catalographic system for ethnomusicological resources; 2) the ontological representation of the domain; and 3) the exposure of catalographic data in the form of Linked Open Data. The purpose of this article is to provide a general overview of the topic of LODs and their use in ethnomusicology and to show the state of the art of the project and the activities of the Turin Unit.*

1.1 Scopo e limiti del presente lavoro¹

Nell'ambito del progetto PRIN «Patrimoni, festival, archivi: pratiche musicali e performative di tradizione orale nel XXI secolo», coordinato da Giovanni Giuriati, l'Unità di Torino ha tre obiettivi principali: 1) realizzare un sistema catalografico dedicato alle risorse di interesse etnomusicologico; 2) fornire rappresentazione ontologica del dominio etnomusicologico; 3) esporre i dati catalografici in forma di Linked Open Data (LOD).

Una prima sperimentazione delle tecnologie LOD in ambito etnomusicologico è stata realizzata nel progetto SAMIC (“Sound Archives and Musical Instruments Collections”, 2016-2019)² che ha costruito le premesse teoriche e tecnologiche del progetto PRIN in corso di attuazione. SAMIC ha realizzato il catalogo digitale LOD del Museo del Paesaggio Sonoro.³

Nelle battute iniziali del progetto, ci si trovava di fronte a un'alternativa: creare una *digital library* chiusa, basata su modelli catalografici “proprietary” o proporre un modello sperimentale di interazione tra *dataset* del dominio etnomusicologico, basato su tecnologie di interoperabilità già adottate in altri domini della conoscenza. L'impiego di sistemi proprietari chiusi ha una serie di inconvenienti manifesti. Il primo è l'obsolescenza tecnologica: una volta esaurita la fonte di finanziamento per realizzare un nuovo sistema lo stesso va in corso a un più o meno marcato invecchiamento, con l'esito – assai più frequente di quanto si possa immaginare – di una successiva scomparsa dal dominio digitale. Il secondo inconveniente è la dispersione digitale dei dati: ogni *data provider* adotta proprie regole di accesso e di definizione dei *record*: *dataset* che potrebbero o dovrebbero essere correlati esistono in domini digitali indipendenti. L'utente, sia esso un ricercatore o il pubblico generalista, non possiede in questo caso un punto di accesso unificato a dati catalografici che, se correlati, consentirebbero un notevole arricchimento reciproco. Le tecnologie LOD hanno ormai impiego in svariati progetti in ambito culturale. A livello internazionale, progetti LOD rilevanti sono stati intrapresi dalla Online Computer Library Center (OCLC) negli Stati Uniti, dalla Bibliothèque Nationale de France, dalla Biblioteca Nacional de España e da altre biblioteche nazionali europee. A livello europeo e in ambito museale i progetti LOD del British Museum e il progetto Europeana sono tra i più noti. In Italia, in ordine cronologico: nel 2013 l'ICAR ha esposto un *endpoint* SPARQL per gli archivi del Sistema

¹ Gli autori hanno elaborato questo testo in cooperazione e dunque ne condividono i contenuti, i metodi e le prospettive. La scrittura dei paragrafi è tuttavia stata divisa come segue: G. Ferdeghini ha scritto il § 1.4 dalle parole “L'idea di base delle ontologie” alle parole “superclasse presente nell'ontologia di dominio”, il § 1.6 e la Bibliografia/Sitografia. L. D'Amato ha scritto il § 1.5 dalle parole “Ciò consente ad esempio di concepire questa query SPARQL” al termine; I. Meandri ha scritto i restanti paragrafi.

² Cfr. Meandri e Ghirardini 2020.

³ Il catalogo digitale è reperibile al seguente indirizzo: <<http://outbrain.acusteme.org/>>.

Archivistico Nazionale; l'Archivio Centrale dello Stato e l'Istituto per i Beni culturali della Regione Emilia Romagna hanno realizzato ReLoad per sperimentare le metodologie del *semantic web*; l'ICCU ha pubblicato come dati aperti le informazioni dell'anagrafe delle biblioteche italiane; nel 2014 la Fondazione CDEC ha realizzato Open Memory Project per integrare e pubblicare in Linked Data i dati sulla Shoah in Italia provenienti da fonti archivistiche e bibliografiche. Nel 2016, il MIBAC ha pubblicato i dati relativi ai Luoghi della Cultura e agli Eventi culturali in Linked Data, che erano tradizionalmente gestiti tramite il DBUnico 2.0. Nel 2017, il Coordinamento delle Biblioteche Speciali e Specialistiche (CoBiS) di Torino ha pubblicato un portale nel quale convergevano inizialmente i cataloghi di sei biblioteche aderenti al progetto CoBiS LOD (attualmente diventate tredici), armonizzati dal punto di vista tecnico e semantico attraverso le tecnologie Linked Data e arricchiti dai dati provenienti da Wikidata, VIAF⁴ e altri dizionari biografici e bibliografici. Nel 2018, l'ICCD in collaborazione con il CNR ha avviato il progetto ArCo (Architettura della Conoscenza) che ha rilasciato ontologie OWL (Web Ontology Language)⁵ basate sul modello dei tracciati catalografici ICCD e la pubblicazione come Linked Open Data delle schede relative ai beni catalogati nel Sigec (Sistema Informativo Generale del Catalogo).⁶

Lo scopo di questo intervento è fornire al lettore/lettrice una panoramica generale sul tema dei Linked Data e sul loro utilizzo in relazione al dominio dell'etnomusicologia. I paragrafi introduttivi consentiranno anche a chi non possieda pregresse competenze in questo campo di comprendere lo stato del progetto e le attività dell'Unità di Torino. Utilizziamo qui concetti di larga applicazione nel campo delle *digital humanities*, esposti, per quanto possibile, con un linguaggio accessibile. Nel corso di questo lavoro entreremo talvolta nel dettaglio di alcune *query* in linguaggio SPARQL, esponendo i costrutti e fornendo il *link* per eseguirle in tempo reale. Sebbene la sintassi di queste *query* possa apparire complessa, va detto che la loro logica è ben accessibile e generalmente compresa dagli umanisti che d'abitudine si occupano di classi, sottoclassi e proprietà in modo assai più raffinato. In queste occasioni, riteniamo comunque utile focalizzarsi sui concetti teorici e applicativi che consentono di comprendere il lavoro che stiamo conducendo, senza la necessità di padroneggiare in dettaglio la sintassi delle *query* stesse.

1.2 Come funzionano i LOD? Un esempio a partire dai toponimi

Un esempio pragmatico può aiutare ad esporre il funzionamento dei LOD sia dal punto di vista tecnico che teorico. Consideriamo la necessità, comune a molte basi di dati, di localizzare geograficamente la provenienza di una risorsa. Di fatto nella localizzazione delle

⁴ VIAF (Virtual International Authority File): <<https://viaf.org>>. Tutti gli accessi a siti internet, qualora non diversamente specificato, si riferiscono alla data del 13 marzo 2023.

⁵ Web Ontology Language. Per una definizione cfr. *infra*.

⁶ <<http://wit.istc.cnr.it/arco>>.

risorse è sensato prevedere più strade. Ad esempio, per localizzazioni incerte, ma di cui si conosce grossomodo la provenienza geografica, sarebbe auspicabile poter georeferenziare l'informazione non solo per divisioni amministrative (comuni, città, regioni, stati ecc.) o per coordinata geografica, ma anche per areale. Per tenere la trattazione entro limiti ragionevoli in questo esempio consideriamo informazioni relative a toponimi riconducibili a divisioni amministrative di uno Stato. Una prima possibilità, seguita da molti sistemi, è quella di mantenere ogni informazione relativa ai luoghi geografici all'interno del proprio sistema catalografico. Sappiamo tuttavia che i toponimi esibiscono di regola varianti e, sebbene siano più stabili di altre categorie di metadati, non sono immuni da mutamenti nel tempo. Si rende perciò necessario un controllo a livello di *authority* di una lista di nomi accettati, di varianti, di toponimi non più utilizzati ma egualmente rilevanti sul piano storico. L'ipotesi alternativa è l'uso di campi di testo libero che, tuttavia, non disambiguando l'informazione geografica, rischiano di produrre una certa entropia nel proprio *dataset*. La stessa complessità esibita dai toponimi riguarda nella pragmatica ogni tipo di metadato catalografico: si pensi ai nomi degli autori o alle denominazioni locali di strumenti musicali. L'altissima mole di informazioni rende di fatto impraticabile la gestione interna d'*authority*. Ci si rivolge perciò a basi di dati specializzate (*authority services*) come VIAF per i nomi di autori o come Geonames⁷ per i toponimi. Geonames è un *database* geografico che contiene 25 milioni di nomi geografici ed è gestito da una comunità di utenti qualificati che contribuiscono con dati e informazioni al suo continuo aggiornamento. Ciò rende

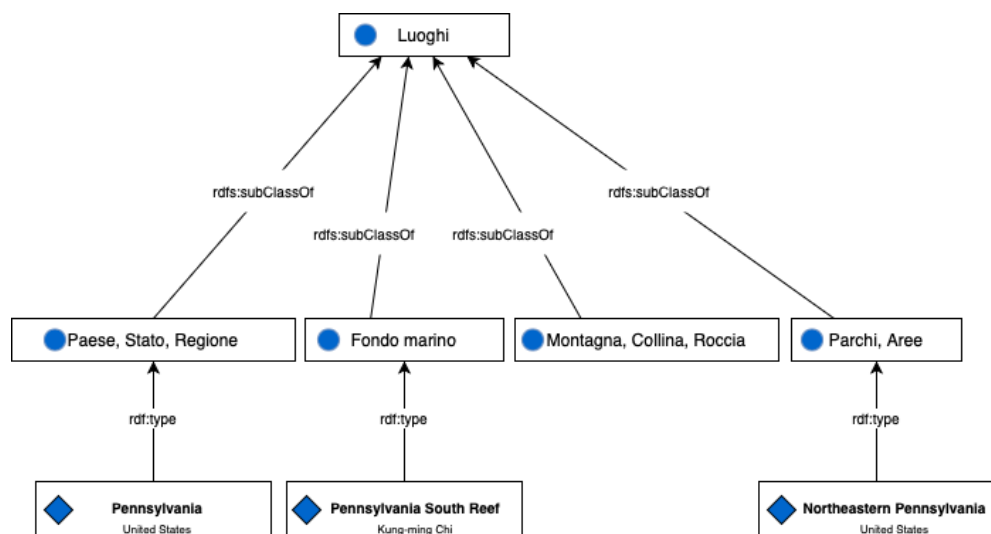


FIGURA 1. Classi di toponimi in una elementare rappresentazione ontologica.

⁷ Geonames: <<https://www.geonames.org>>.

possibile, ad esempio, la gestione delle varianti dei nomi (se ne contano attualmente circa 16 milioni), con un'accuratezza impensabile per un sistema proprietario e per una gestione centralizzata. Quando si rappresentano informazioni, come quelle geografiche, sorge il problema di differenziare le tipologie di toponimi. Ad esempio, la stringa "Pennsylvania" potrebbe virtualmente riferirsi a uno (o più) comuni, regioni, montagne, fiumi, luoghi sottomarini o anche luoghi immaginari. In questa semplice, per quanto indispensabile, divisione in classi incontriamo un concetto correlato ai LOD, ovvero quello di ontologia. Un'ontologia computazionale è la rappresentazione di un dominio della conoscenza attraverso un insieme strutturato di definizioni condivise di concetti e delle relazioni tra questi concetti, espressi in un linguaggio formale (ad esempio OWL, Web Ontology Language) e una sintassi (RDF, Resource Description Framework).⁸ Le ontologie computazionali sono utilizzate per rappresentare tassonomie in diversi domini di conoscenza e possono essere utilizzate per la ricerca semantica, l'integrazione e l'interoperabilità dei dati, l'elaborazione del linguaggio naturale e l'applicazione in sistemi che richiedano una comprensione semantica dei dati. Nello schema seguente (Fig. 1) si mostrano le possibili classi di toponimi fin qui prese in considerazione in una elementare rappresentazione ontologica (l'ontologia di Geonames è, come si può facilmente immaginare, ben più complessa).

Le informazioni geografiche di Geonames sono disponibili in diversi formati.⁹ Idealmente l'accesso a una fonte di Linked Open Data avviene attraverso un *repository/endpoint* di dati pubblicati in RDF, interrogabili tramite una *query* (interrogazione realizzata utilizzando il linguaggio SPARQL di cui si parlerà più avanti). RDF è uno standard del W3C utilizzato per rappresentare le informazioni sul web semantico in modo strutturato e interoperabile tra diverse applicazioni. La cosiddetta *tripla RDF* è la struttura di dati fondamentale di RDF. Una tripla è composta da tre parti: un soggetto (subject), un predicato (predicate) e un oggetto (object). Il soggetto rappresenta l'entità a cui si riferisce l'informazione, il predicato descrive la relazione tra il soggetto e l'oggetto, e l'oggetto rappresenta il valore associato al soggetto tramite il predicato, che può essere un'altra entità o una stringa alfanumerica (come un numero, un testo, una data). Nel primo esempio qui sotto l'entità "Po" è collegata al concetto/entità "fiume", mentre nel secondo si fornisce un dato numerico:

```
<Po> <è> <un fiume>
<Po> <è lungo> "652km"
```

Nel caso delle entità, il soggetto di una tripla RDF può diventare a sua volta l'oggetto di un'altra tripla, e viceversa. Questo è uno dei vantaggi del modello a grafo utilizzato in

⁸ Per un'introduzione a OWL e RDF si vedano: <<https://www.w3.org/OWL/>>, <<https://www.w3.org/RDF/>>; Allemang, Hendler and Gandon 2020.

⁹ Specifiche sulle diverse possibilità di accesso ai dati Geonames come dati semantici sono fornite qui: <<https://www.geonames.org/ontology/documentation.html>>.

RDF, nel quale le *entità* (elementi o concetti del mondo reale o di un sistema, rappresentati come nodi del grafo) sono connesse tra loro da relazioni semantiche (nel grafo rappresentate da archi) definite tramite triple. Ad esempio:

```
<Po> <si trova in> <Italia>
<Italia> <è> <una nazione>
```

Soggetti, predicati e oggetti, quando non si tratti di valori espressi attraverso stringhe alfanumeriche, sono definiti da URI (Uniform Resource Identifier) ovvero identificatori univoci e persistenti che vengono utilizzati per evitare ambiguità nella rappresentazione delle informazioni e per creare collegamenti tra le diverse risorse. Gli URI possono essere risorse del web, come pagine web o documenti (in questo caso rappresentano una sottospecie di URI, ovvero i comuni URL), oppure possono rappresentare concetti astratti, come le entità di un'ontologia. Nel web semantico, gli URI vengono utilizzati dunque anche per l'identificazione di concetti, classi, proprietà, relazioni e istanze. Il processo di trasformazione di dati strutturati in una rappresentazione di dati collegati si definisce "triplicazione" e si realizza utilizzando lo standard RDF. La triplicazione consiste nel rappresentare ogni informazione contenuta in un insieme di dati strutturati come triple RDF, ciascuna delle quali composta, come nell'esempio precedente, da un soggetto, un predicato e un oggetto. Un insieme di triple genera un grafo RDF.

Ad esempio, nel nostro progetto, si parte da dati catalografici inseriti in un sistema di *back end* (immissione dati) che verranno poi triplicati e pubblicati come LOD in un *endpoint* RDF.¹⁰ L'operazione di triplicazione è essenziale per l'esposizione di dati LOD.

SPARQL (SPARQL Protocol and RDF Query Language) è un linguaggio di interrogazione per la *semantic web*, utilizzato per estrarre informazioni da grafi RDF.¹¹ SPARQL consente di effettuare ricerche semantiche, permettendo di recuperare dati specifici da diversi *dataset* e di combinarne i risultati. Il linguaggio supporta questo tipo di interrogazione attraverso la definizione di *pattern* di ricerca e l'utilizzo di regole di inferenza che consentono di derivare nuove informazioni a partire dai dati esistenti. SPARQL utilizza una sintassi simile a SQL (Structured Query Language), ma con alcune differenze significative per gestire le peculiarità del web semantico e delle ontologie. Il linguaggio è basato su un modello a grafo e consente di impostare *query* che si basano sulle relazioni tra i dati e sulle ontologie di riferimento. Una *query* SPARQL consente perciò di configurare l'interrogazione all'archivio di dati aperti con notevole flessibilità rispetto alle ricerche tradizionali poiché consente di combinare diverse triple, di eseguire *query* complesse basate su relazioni e operazioni di aggregazione, raggruppamento e ordinamento dei risultati. Uno dei punti di forza di SPARQL è la capacità di navigare il grafo RDF esplorando le relazioni tra i dati. Per i nomi geografici,

¹⁰ A partire dal progetto SAMIC abbiamo adottato Virtuoso, un popolare sistema per gestione basi di dati RDF e LOD.

¹¹ <<https://www.w3.org/TR/rdf-sparql-query/>>.

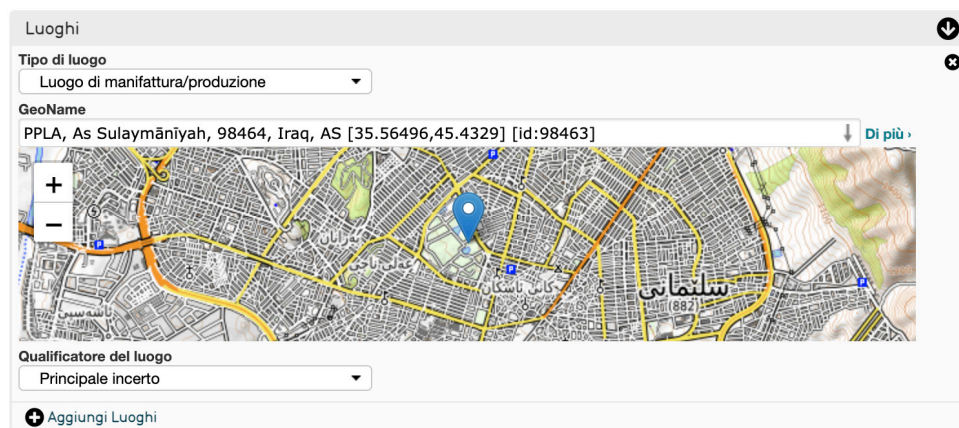


FIGURA 2. Correlazione di due informazioni geografiche LOD.

per tornare all'esempio precedente, è possibile configurare la *query* per ottenere informazioni specifiche relative alle classi dell'ontologia Geonames che rappresentano toponimi corrispondenti a divisioni amministrative (comune, provincia, regione), escludendo ad esempio i luoghi sottomarini o i fiumi o altre classi geografiche ove li si consideri non pertinenti per la ricerca. L'interrogazione al *repository* LOD avviene in tempo reale, dunque ogni cambiamento nell'informazione si riflette su una nuova *query*. Come Geonames rappresenta toponimi, altre fonti di dati aperte consentono l'accesso a informazioni geografiche. OpenStreetMaps (OSM),¹² ad esempio, fornisce l'accesso a mappe disponibili come informazioni semantiche (sorrette da una specifica ontologia) e interrogabili attraverso *query* SPARQL. Partendo da un toponimo, che nel grafo Geonames è collegato a una (o più) coordinate geografiche, è possibile interrogare simultaneamente OSM chiedendo di fornire una mappa (satellitare, urbanistica, orografica, idrografica... dipende dalla classe pertinente alla ricerca) relativa alle coordinate geografiche corrispondenti al toponimo scelto.

In Fig. 2 è riprodotto un semplice collegamento tra due informazioni geografiche – toponimi e mappe – reperite da due diversi *repository* esterni LOD che servono centinaia di applicazioni sul web.¹³

Si è finora insistito su un solo dominio, quello dei luoghi geografici. Possiamo chiederci se anche altri domini di conoscenza, similmente a toponimi e mappe, siano rappresentati in LOD. La galassia formata da tutte le informazioni sui dati collegati è chiamata Linked Data Cloud, ovvero l'insieme di dati interconnessi, rappresentati utilizzando gli standard del web semantico. Questi dati sono pubblicati sul web e collegati tra loro attraverso URI comuni, che formano una vasta rete di informazioni interconnesse. La Linked

¹² <<https://www.openstreetmap.org/>>.

¹³ Questa correlazione usa funzionalità Linked Data native del CMS Collective Access e, in particolare, un *metadata element* di tipo GeoNames.

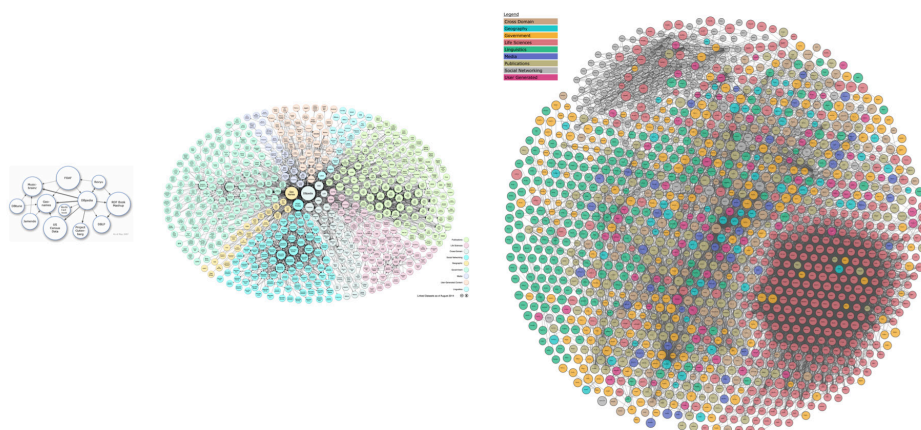


FIGURA 3. Evoluzione della LOD Cloud (primo diagramma: plot al 1/5/2007; secondo diagramma: 30/08/2014; terzo diagramma: 3/9/2023).

Data Cloud si basa su un principio fondamentale del web semantico, ovvero l'idea che i dati debbano essere resi disponibili in un formato standardizzato e strutturato, in modo che possano essere facilmente condivisi e collegati tra loro. Le informazioni contenute nella Linked Data Cloud possono essere utilizzate da applicazioni e servizi con una flessibilità impensabile se i dati fossero presentati in formati proprietari o isolati.

LOD Cloud Diagram¹⁴ monitora i *dataset* pubblicati in forma di Linked Data. Nelle istantanee riprodotte in Fig. 3, ogni specifico dominio di conoscenza è rappresentato con un cerchio (come i toponimi di Geonames, gli autori di VIAF, le mappe di OSM).

Le tre immagini mostrano l'evoluzione della Linked Data Cloud (Fig. 3), a partire dal 2007 e consentono di cogliere a colpo d'occhio l'espansione esponenziale dei *dataset* LOD.

Come si può evincere dalla figura, la Linked Data Cloud è composta da migliaia di fonti di dati, tra cui *dataset* governativi, dati scientifici, cataloghi di biblioteche, informazioni geospaziali e *dataset* derivanti da diversi domini della conoscenza. È legittimo chiedersi quali siano i vantaggi dell'interconnessione. Le risposte sono molteplici. La prima, pragmatica, è la capacità di rappresentare informazioni in un modo più accurato di quel che si potrebbe fare con proprie risorse. Se le mappe, o i toponimi, o i nomi di autori – per fermarsi agli esempi sin qui trattati – non sono più gestiti a livello di un singolo *resource center* ma la loro rappresentazione è demandata a basi di dati specialistiche costruite in modo collaborativo, la qualità e la quantità dell'informazione aumentano. Secondo questo paradigma, un *database* locale delega competenza su alcuni domini della conoscenza e ha il compito di organizzare le informazioni che già esistono nel web semantico.

¹⁴ <<https://lod-cloud.net>>.



FIGURA 4. Quagliere (Museo del Paesaggio Sonoro).

Un secondo aspetto più sottile e decisivo può spiegare perché la Linked Data Cloud si stia popolando così rapidamente. La questione fondamentale riguarda l'arricchimento: se i dati sono interconnessi, si può pensare alle relazioni tra i dati in modo nuovo. Il web semantico può essere utilizzato come uno strumento di scoperta (*discovery tool*) perché consente di collegare e integrare dati provenienti da diverse fonti. L'adozione di tecnologie semantiche è in grado di offrire numerosi vantaggi nell'ambito della gestione dei cataloghi digitali, non solo perché *apre i cataloghi* rendendoli accessibili in rete come dati semantici, ma perché consente l'integrazione di dati provenienti da fonti eterogenee, anche inerenti ad altri domini della conoscenza. Grazie alla creazione di collegamenti semantici tra le diverse fonti di informazione i cataloghi si arricchiscono diventando strumenti capaci di supportare analisi più approfondite. Si promuove il riuso e la vitalità dei dati catalografici in contesti diversi, con possibilità e scenari imprevedibili: un portato inevitabile, per così dire, naturale della filosofia alla base dei LOD. Inoltre, utilizzando i principi del web semantico, le informazioni possono essere strutturate in modo da renderle interpretabili da parte delle macchine, consentendo alle applicazioni di scoprire, comprendere e utilizzare i dati in modo automatizzato.

Consideriamo ora un esempio teorico, nel quale un portale culturale dedicato al dominio dell'etnomusicologia venga strutturato attraverso la concezione di *query* SPARQL che integrino dati provenienti da vari domini della conoscenza.

In Fig. 4 è riprodotto un quagliere conservato al Museo del Paesaggio Sonoro.¹⁵ Supponiamo che un'ontologia definisca l'entità "Strumento musicale", la quale può avere alcune proprietà, come un *taxon* Hornbostel-Sachs, una definizione tipologica e più denominazioni locali. Supponiamo inoltre che "Europa" sia un toponimo definito semanticamente, collegato a un determinato areale, e che anche "dipinto" sia un elemento definito semanticamente come sottoclasse delle "fonti iconografiche" o delle "arti visive". Se questi concetti sono rappresentati ontologicamente ed esistono relative fonti di dati aperti (*endpoint* SPARQL), è possibile concepire una *query* per reperire tutti i "dipinti" in

¹⁵ Questo quagliere (<<https://outbrain.acusteme.org/entity/0330SM>>) ed altri richiami possono essere visualizzati nel catalogo digitale del museo: <https://outbrain.acusteme.org/?q=&def_tipologica=quaglieri>.

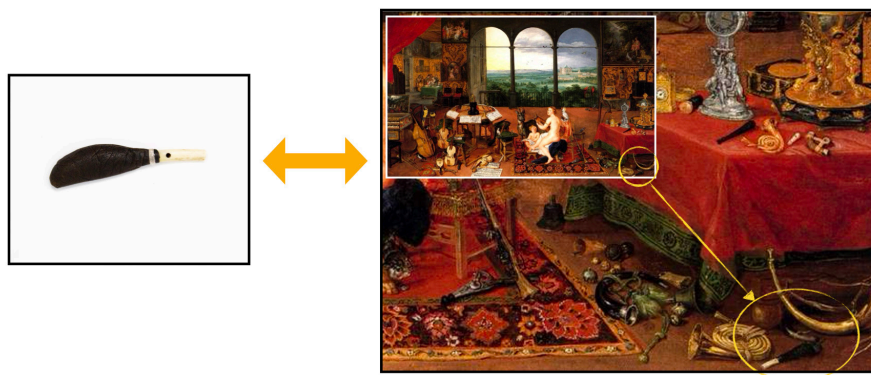


FIGURA 5. Una possibile correlazione LOD tra entità di due domini della conoscenza (un quagliere del Museo del Paesaggio sonoro; un quagliere nel noto dipinto di Rubens e Brueghel il Vecchio – *The Sense of Hearing* 1617-1618).

“Europa” che raffigurano “quaglieri” o strumenti musicali il cui *taxon* sia “421.231.123.1-62 Flauti a fessura interna diritti, singoli, a fessura e finestra, semichiusi, privi di fori digitali, con serbatoio d’aria flessibile”.

I risultati di questa correlazione sono molteplici. Da un lato i LOD possono essere strumenti per la scoperta (*discovery tool*), per la ricerca in dati strutturati (e.g. per la conduzione di ricerche in iconografia musicale); dall’altro, sono strumenti che consentono di arricchire i propri portali, poiché una parte di dati è gestita direttamente dal proprio *resource center*, una parte proviene, invece, in tempo reale da altri *repository* di LOD. L’arricchimento è reciproco poiché la correlazione è naturalmente simmetrica. Un *repository* specializzato in dipinti (e.g. la *digital library* di un museo) può chiedere, ad esempio, quali musei o istituzioni conservino oggetti (appartenenti a una specifica classe di una rappresentazione ontologica) simili a quello rappresentato in una determinata risorsa iconografica interrogando simultaneamente più *endpoint* SPARQL.

Naturalmente, una serie di precondizioni è necessaria affinché la *query* possa essere eseguita restituendo risultati qualitativamente rilevanti. La condizione fondamentale è definire semanticamente i concetti necessari alla rappresentazione di un dominio della conoscenza, ovvero fornire una rappresentazione ontologica di questo dominio. Nel nostro specifico esempio lo si è premesso: strumenti musicali, toponimi e luoghi geografici, dipinti. Si dovrebbero poi combinare più ontologie per integrare informazioni appartenenti a domini diversi. Nel nostro specifico esempio si potrebbe utilizzare un’ontologia relativa agli strumenti musicali per descrivere strumenti o oggetti sonori presenti in un quadro, in modo da sapere che quel dipinto contiene richiami di caccia. Questo permetterebbe, ad esempio, di chiedere quali dipinti, in una determinata epoca e regione geografica, raffigurano richiami da caccia in generale o richiami da caccia per le quaglie e in particolare un quagliere, come nel celebre esempio di Rubens e Brueghel il Vecchio. Nella realtà dei fatti,

poiché le competenze di dominio sono fondamentali per poter identificare e classificare uno strumento, basterebbe che l'istituzione che possiede o descrive digitalmente un quadro identifichi la presenza di uno strumento musicale in un'opera¹⁶ (utilizzando cioè la classe generale di un'ontologia relativa agli strumenti musicali) consentendo poi che un altro dominio della conoscenza (o più d'uno) producano altre asserzioni semantiche su quel determinato dipinto. Similmente al principio illustrato nell'accostamento tra strumento musicale/rappresentazione iconografica è possibile pensare liberamente ad altre correlazioni tra i dati, a seconda del proprio specifico ambito di ricerca o delle strategie di arricchimento progettate da un portale. In questi anni di lavoro dedicati al tema dell'esposizione dei dati in LOD, ci sembra di aver compreso che parte dell'attività di ricerca su archivi e digitalizzazione si configurerà sempre di più come campo inter e trans-disciplinare nel quale il ricercatore/la ricercatrice gioca un ruolo fondamentale nell'individuare e sperimentare interconnessioni LOD sensibili rispetto al proprio dominio di specializzazione. In una parola, è (o può essere) importante imparare a navigare nella Linked Data Cloud, un'attitudine icasticamente riassunta dall'esergo che traiamo da Forster, il celebre "Only connect". L'attitudine della *community* al riuso di informazioni già pubblicate è peraltro testimoniata dalla seguente *query* <<https://w.wiki/6SMC>>, nella quale vengono visualizzati su una mappa geografica le localizzazioni di dipinti conservati in Europa per i quali gli utenti Wikidata hanno utilizzato *taxa* della classificazione HS nella revisione Guizzi. Non è una *query* attualmente utilizzata nel nostro CMS o nel *design* del nostro futuro portale di navigazione nei dati catalografici – le informazioni restituite sono per il momento poche e sembrano perlopiù orientate all'identificazione di flauti di pan – ma è un esempio eloquente di come, a partire dalla pubblicazione di dati semantici, altri dati possano nel tempo convergere configurando un progressivo arricchimento dell'informazione.

1.3 Un'ontologia per il dominio etnomusicologico

Da quanto abbiamo esposto dovrebbe apparire chiara la necessità di provvedere a una rappresentazione ontologica per il dominio dell'etnomusicologia. Un'ontologia che organizzi le risorse di questo specifico dominio della conoscenza consentirebbe infatti la pubblicazione di dati in LOD, ma anche il riuso di tale rappresentazione a vantaggio di altri domini della conoscenza (la descrizione strumenti musicali rappresentati in dipinti valga come esemplificazione di questa possibilità).

¹⁶ Senza voler sminuire la complessità dell'attribuzione di *aboutness*, e nello specifico, di un soggetto a un'opera d'arte figurativa – operazione per la quale sono stati elaborati complessi cataloghi come ad esempio Iconclass (<<http://www.iconclass.org/help/outline>>) o Art & Architecture Thesaurus Online (AAT – <<https://www.getty.edu/research/tools/vocabularies/obtain/download.htm>> [Guerrieri 2022]) – si sottolinea la capitale importanza della competenza del catalogatore in fase di catalogazione di una risorsa, soprattutto quando, come nel caso qui presentato, essa possa intersecare domini della conoscenza diversi e apparentemente lontani. Qualora, come plausibile, la competenza esperta di un catalogatore non possa spaziare in una pluralità di campi diversi con la medesima profondità e accuratezza è comunque auspicabile una collaborazione interdisciplinare per la quale i LOD possono costituire un *framework* possibile di collaborazione.

Come si è detto, una prima sperimentazione delle ontologie e delle tecnologie LOD in ambito etnomusicologico è stata realizzata nel progetto SAMIC, che aveva l'obiettivo iniziale di rendere fruibile in rete, come *digital library*, la collezione di strumenti musicali e oggetti sonori del Museo del Paesaggio Sonoro di Riva Presso Chieri. L'analisi dello stato dell'arte ha portato all'individuazione di due progetti digitali, uno italiano e uno europeo, ovvero la normativa SM dell'ICCD (ICCD-SM 4.0)¹⁷ e l'ambizioso progetto di interoperabilità tra musei organologici del consorzio europeo MIMO (Musical Instruments Museums Online).¹⁸ SAMIC ha condotto una sperimentazione costruita entro un dominio ristretto, con due obiettivi principali:

- 1) Utilizzare un proprio *data model* per la catalogazione di strumenti musicali;
- 2) Garantire l'interoperabilità dei/tra i propri dataset con/e progetti nazionali e internazionali nel dominio degli strumenti musicali.

Perché un progetto di ricerca in questo ambito? Perché non “accontentarsi”, diciamo così, di una già esistente normativa italiana dedicata agli strumenti musicali, corredata da un relativo software di catalogazione (il SIGECweb) che consente peraltro l'esposizione dei dati catalografici raccolti nel Catalogo Generale dei Beni Culturali. La risposta a questa obiezione è complessa e ha orientato la proposta di ricerca del progetto PRIN:

- 1) La ricerca etnomusicologica si svolge in contesti internazionali rispetto ai quali è utile possedere CMS (Content Management Systems) – ovvero di interfacce di immissioni dati¹⁹ – interoperabili ma indipendenti rispetto ai sistemi informativi nazionali: se l'interoperabilità può essere garantita, la parte del *dataset* patrimonializzata in un ente conservatore, soggetta alle norme nazionali, può essere riversata nel catalogo nazionale mentre la parte del *dataset* non soggetta a tali norme può essere catalogata, utilizzata a fini di ricerca, ma non riversata nel catalogo nazionale o resa disponibile in LOD in modo da garantire la possibilità, a livello internazionale, di eseguire una ricerca integrata (tramite interrogazione simultanea a più endpoint SPARQL o riversamento entro un singolo endpoint SPARQL);

- 2) I *corpora* documentali di interesse etnomusicologico posseggono un elevato grado di complessità. Per la ricerca sono rilevanti sia risorse materiali (strumenti musicali, oggetti e paraphernalia di interesse etnografico-musicale, partiture, parti, appunti di campo, libri, documenti audiovisivi e fotografici su qualsiasi supporto e formato, più in

¹⁷ Al tempo in cui il progetto SAMIC fu concepito lo standard SM 1 era in sperimentazione. Rilasciata poi nel 2016 è oggi giunta alla versione 4.0 (<http://www.iccd.beniculturali.it/ricercanormative/64/sm-strumenti-musicali-4_00>).

¹⁸ <<https://mimo-international.com/MIMO/>>.

¹⁹ Un Content Management System (CMS) è un software che consente di organizzare e pubblicare contenuti digitali, come testi, immagini, video, audio e altri tipi di file. Un CMS permette di creare, modificare e pubblicare contenuti attraverso una interfaccia *web-based*. Il progetto SAMIC, così come il progetto PRIN, utilizzano come piattaforma di *data entry* il CMS Collective Access (<<https://www.collectiveaccess.org>>), ampiamente utilizzato da musei e archivi a livello nazionale e internazionale, con un profilo creato ad hoc per il dominio etnomusicologico, attualmente in sperimentazione. Il sistema di data entry e l'ontologia verranno rilasciati in libero accesso al termine del progetto.

generale sia risorse pubblicate che non pubblicate), sia beni immateriali (il contenuto di o associato ad alcune di queste risorse). Questa tipologia di risorse può far parte di beni patrimonializzati oppure no. Di fatto nella ricerca si correlano spesso beni patrimonializzati a beni non patrimonializzati la cui rappresentazione in dominio digitale è comunque auspicabile. Nelle norme nazionali non esiste al momento una definizione unificata di bene etnomusicologico, esistono di conseguenza più normative (peraltro in competizione) e più sistemi informativi di riferimento.

Un sistema unificato specificamente rivolto alla ricerca, che possieda capacità interoperative e consenta il riversamento dei corpora patrimonializzati verso i cataloghi digitali “normati”, laddove gli stessi lo consentano, può in teoria tenere insieme questa complessità. I *corpora* di risorse del Museo del Paesaggio Sonoro sono un buon esempio di questa commistione:

1) Esiste un fondo di strumenti musicali, patrimonializzato;

2) Esiste una rete informale di utensili e oggetti (e.g. utilizzati nei laboratori didattici) che non è patrimonializzata o di manufatti non patrimonializzati investigati in sede di ricerca sul campo;

3) Esistono fondi di partiture edite e inedite di interesse etnomusicologico;

4) Esiste una rete informale di documentazione audiovisiva posseduta dai ricercatori e dalle ricercatrici che ruotano attorno al museo, non patrimonializzata o solo in parte patrimonializzabile, ma comunque correlabile ai fondi museali.

In generale il Museo, nella sua fondamentale interazione con la ricerca scientifica, funziona come aggregatore continuo di risorse, formali e informali, utilizzate a fini di ricerca o di didattica, ma nella maggior parte dei casi non patrimonializzate e in taluni casi non patrimonializzabili. Si tratta di una condizione assai diffusa che riguarda, peraltro, anche enti che per statuto e missione non avrebbero scopo di enti conservatori (come le biblioteche, per le quali le norme catalografiche si stanno espandendo al fine di includere la possibilità di descrivere beni non librari, fondi archivistici, oggetti). Un esempio concreto: una risorsa pubblicata (una musica per bandella edita a fine Ottocento) è correlabile a documentazione audiovisiva raccolta sul campo di repertori di tradizione orale per quintetto di fiati, a strumenti musicali e oggetti sonori, ad appunti di ricerca sul campo, a trascrizioni, interviste, documenti audiovisivi o fonti iconografiche già presenti in rete, a musica a stampa e manoscritta non pubblicata inerente le stesse pratiche. Non si tratta di esempi di specie, ma della concreta complessità documentale con cui si confronta quotidianamente l'etnomusicologo.

È possibile tenere tutto insieme nel dominio digitale? Inoltre, l'esito della sperimentazione sull'interoperabilità consente di proporre una metodologia di condivisione dei dati che consenta, sia a livello nazionale che internazionale, di contribuire al superamento della frammentazione della rappresentazione del dominio etnomusicologico in rete? La complessità del dominio “scavalca” per così dire il quadro normativo attuale, ormai troppo restrittivo per l'etnomusicologia, così come per altri ambiti di ricerca.

Ci si può lamentare delle conseguenze che derivano da questa assenza di “recinzioni”, per così dire; oppure se ne possono comprendere le ragioni nella crescita di complessità delle risorse stesse (o nelle pratiche di conservazione, patrimonializzazione e riuso). Noi preferiamo considerare le possibilità che si aprono a partire dal lavoro su questo stato entropico a valle del problema e non a monte. È facilmente intuibile la direzione di una fusione della prospettiva archivistica, bibliografica e di quelle dedicate al patrimonio materiale e immateriale in una serie di descrizioni integrate sorrette da altrettante rappresentazioni ontologiche che tendono, nel tempo, a convergere (progressive revisioni dei domini rappresentati dalle ontologie possono esibire tendenze centripete: la recente *release* dell’ontologia RiC-O è in questo senso un esempio eclatante).²⁰ Il corollario di questa convergenza in atto è che entro non molto tempo possano diventare superflue operazioni che anche SAMIC, nell’ambito complesso delle risorse di tipo etnomusicologico, aveva considerato fondamentali, come ad esempio il riversamento verso *digital libraries* nazionali attraverso le opportune ontologie (e.g. i libri verso ICCU, gli strumenti verso ICCD o MIMO).

Come si è già detto, SAMIC ha reso interoperabili i propri *dataset* relativi a strumenti musicali, pubblicandoli in forma di LOD, allineandoli a una prima proposta ontologica, ma limitata al dominio dell’organologia. Diversi passaggi sono necessari perché i dati catalografici relativi agli strumenti possano essere pubblicati in forma di LOD. Abbiamo iniziato rendendo disponibile la classificazione Hornbostel-Sachs come dato semantico, utilizzando una base di conoscenza comunemente usata per rilasciare dati in LOD, ovvero Wikidata. L’albero HS, nella revisione curata da Febo Guizzi, è stato rappresentato su Wikidata in italiano e in inglese e può dunque essere interrogato in tempo reale per reperire *taxon* specifici con cui metadattare uno strumento musicale.

Come si evince dalla Fig. 6, il CMS di *data entry* accede in tempo reale, tramite *query* SPARQL, all’albero HS (vengono in questo caso filtrati gli elementi pertinenti della HS sulla base della stringa “Flauti a fessura interna diritti”). Ciò consente di metadattare semanticamente i nostri strumenti. L’albero HS è al contempo accessibile da altri applicativi. Il rilascio sotto forma di LOD può essere potenzialmente importante anche per un altro aspetto dell’arricchimento che non è stato discusso finora. Supponiamo che altri progetti internazionali intendano fare uso della classificazione HS e necessitino di una sua traduzione francese. Se la stessa è reperibile in Wikidata è probabile che un progetto non costruisca competitivamente un nuovo albero HS da zero, ma si basi, collaborativamente, su rappresentazioni esistenti, aggiungendo una traduzione francese alle nostre versioni in italiano e inglese (lo stesso ragionamento vale per le altre lingue). L’esistenza di dati può cioè fungere da calamita affinché, con l’attitudine collaborativa di una *community* di utenti specializzati, la rappresentazione sia nel tempo più completa. È forse superfluo

²⁰ RiC-O (versione 0.2, 2021) L’ontologia RiC-O è sviluppata dall’International Council on Archives (ICA) per fornire un modello concettuale per descrivere record di diversi tipi e formati, inclusi i record archivistici e culturali. <https://www.ica.org/standards/RiC/RiC-O_v0-2.html>.

The screenshot shows the Acusteme web application interface. At the top, there is a navigation bar with the Acusteme logo and the text 'NUOVO CERCA GESTISCI IMPORTA CRONOLOGIA'. Below this, there is a search bar and a 'Cerca' button. The main content area is divided into a left sidebar and a main panel. The sidebar contains a menu with categories like 'INVENTARIO', 'DATI AMMINISTRATIVI', 'TUTELA', 'IDENTIFICAZIONE', 'COLLOCAZIONE', 'CLASSIFICAZIONE', 'FABBRICAZIONE', 'USO/CONTESTO', 'DESCRIZIONE', 'CONSERVAZIONE/RESTAURI', 'MATERIALI/TECNICHE', 'MISURE', 'ESPOSIZIONI', 'MEDIA', 'MODELLI 3D', 'BIBLIOGRAFIA/SITOGRAFIA', 'ANNOTAZIONI', 'DATI SCHEDA', and 'SOMMARIO'. The 'CLASSIFICAZIONE' category is highlighted. The main panel shows a search bar with the text 'Flauti a fessura interna diritti' and a list of classification items. The items are listed in a hierarchical manner, starting with '421.231.1 Flauti a fessura interna diritti' and continuing down to '421.231.133 Flauti a fessura interna diritti, in serie, a fessura e finestra, chiusi'. At the bottom of the main panel, there is a search bar with the text 'keyword MIMO all languages' and a button to 'Aggiungi keyword MIMO all languages'.

FIGURA 6. L'albero HS come dato semantico.

a questo punto della trattazione ribadire che più progetti su strumenti musicali che utilizzino un medesimo albero HS (o anche versioni diverse ma semanticamente correlate di una classificazione HS)²¹ possono consentire di navigare nei propri cataloghi a partire da un dato semantico relativamente stabile: è ovvero possibile concepire *query* SPARQL che interrogano simultaneamente più *endpoints* chiedendo, ad esempio, di reperire tutti gli strumenti musicali appartenenti al *taxon*, poniamo, “421.231.123.1-62”.

Come si può facilmente immaginare la porzione di ontologia che qui si sta esponendo (con riferimento alla figura precedente, il fatto che esista un oggetto fisico che può essere uno “Strumento musicale”, che può avere un “Taxon HS”) è un pezzo importante ma pur tuttavia piccolo rispetto al dominio etnomusicologico che si intende rappresentare. Anche rimanendo al solo sottodominio degli strumenti musicali molte altre entità e relazioni tra entità contribuiscono a costruire l'ontologia proposta da SAMIC, che qui non commenteremo ulteriormente.

Il PRIN estende a un nuovo livello di complessità l'esperimento SAMIC. La domanda principale è: di quali elementi si compone il dominio etnomusicologico? Le risposte,

²¹ MIMO e SAMIC utilizzano ad esempio due diverse versioni della classificazione HS (revisione MIMO e revisione Guizzi), entrambe rappresentate su Wikidata. Per la revisione MIMO si vedano Birley, Myers and Shepherd 2020. Per la revisione Guizzi si vedano Sachs e Guizzi 2020a Sachs e Guizzi 2020b.

naturalmente, possono variare notevolmente a seconda della prospettiva e tradizione analitico-scientifica dalla quale si risponde. Il nostro approccio è orientato (come d'altronde lo è il lavoro ontologico) al pragmatismo e a una costruzione di tipo *bottom-up*. In altre parole, una possibile risposta non è tanto cosa può esistere teoricamente, ma cosa esiste fattualmente nelle rappresentazioni digitali di un dominio. Siamo perciò partiti dal nostro Museo e da una serie limitata di altri *repository* digitali etnomusicologici esistenti in Italia, per mappare le risorse catalogate e i modelli di catalogazione. Naturalmente la collazione tra *data models* porta alla luce una certa entropia, per così dire, un certo disordine. Nel lavoro di rappresentazione ontologica è perfettamente normale che ciò accada. Una risposta di primo livello (ovvero che renda conto delle *core entities* che in genere vengono considerate da più di un archivio digitale parte integrante del dominio) può essere quella che segue. Il dominio etnomusicologico può comporsi di:

1) Materiale di ricerca sul campo in qualsiasi formato e supporto, compresi appunti, fotografie, filmati; Strumenti musicali;

2) Materiale pubblicato (*book materials* e non *book materials*) in qualsiasi formato e supporto;

3) Descrizione delle campagne di ricerca sul campo (date, luoghi geografici, agenti);

4) Supporto alla descrizione degli agenti (persone, famiglie, istituzioni);

5) Descrittori specifici pertinenti al dominio etnomusicologico, che interessano ogni tipo di risorsa fin qui elencata ne consentono descrizione catalogografica o annotazione, ovvero:

a) Informazioni sugli antroponimi (gruppi etnici, demonimi, nazionalità, aree geo-culturali);

b) Informazioni relative ai fenomeni musicali (scale, modi, nomi locali di voci, generi musicali, ensemble vocali o strumentali, ecc.);

c) Informazioni relative ai fenomeni coreutici (nomi locali delle danze, elementi coreografici che compongono una danza, ecc.);

d) Informazioni sugli oggetti e sui *paraphernalia* che sono parte del rito o che sono utensili in qualche modo connessi a pratiche o saperi musicali.

È probabile che per quanto comprensivo questo elenco non soddisfi tutti i palati, o forse sì, ma partiamo, a livello di *design*, dal presupposto che possa non farlo. È importante sottolineare che l'elenco non ha velleità teoriche o normative, non descrive un *dover essere*, ma uno stato delle cose. Di fronte all'obiezione "manca un x" – cioè a dire, lo schema proposto manca di un pezzo di rappresentazione di dominio – la soluzione pragmatica è trovare la risposta alle seguenti domande:

- in quale *dataset* è rappresentata l'informazione "x" mancante, e attraverso quale *data model*?

- come è possibile rappresentare l'informazione "x" mancante a livello ontologico e come è possibile integrarla nel sistema?

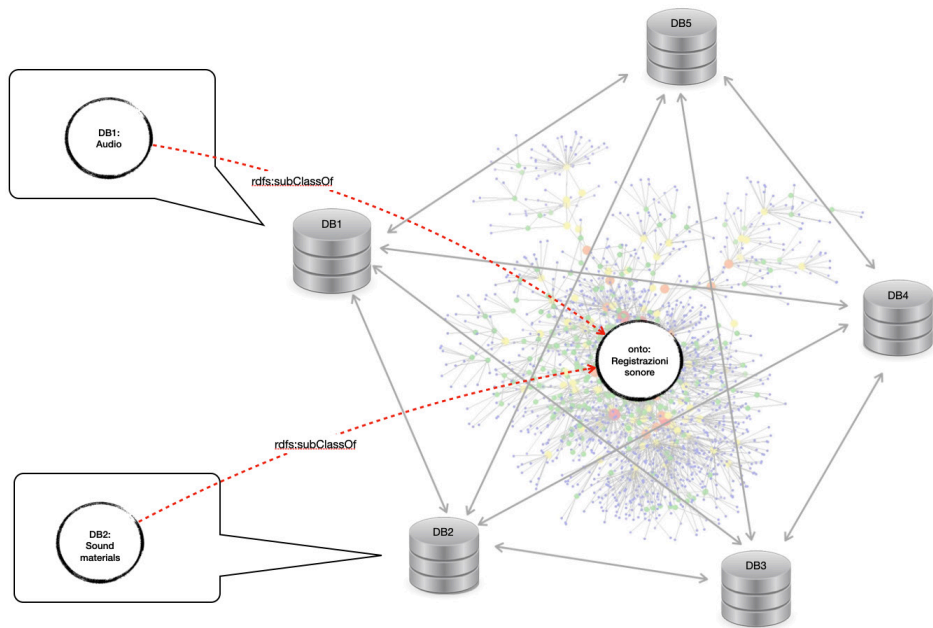


FIGURA 7. *Resource center* locali e ontologia, un possibile modello di interoperabilità.

1.4 Il principio del riuso

In un dominio di tale complessità è chiaro che la rappresentazione ontologica intreccia spesso rappresentazioni già esistenti in altri domini della conoscenza. Una buona pratica nel *design* di ontologia è il riuso di ontologie già esistenti. Se descrivo un agente, supponiamo, un costruttore di strumenti musicali, o un restauratore, o un musicista, è auspicabile il riuso di rappresentazioni ontologiche già collaudate nel campo della rappresentazione degli agenti. Similmente, se devo descrivere un materiale bibliografico è opportuno che si tenga conto di rappresentazioni ontologiche già pubblicate per questo campo. Il principio della tripla AAA (*Anyone can say Anything about Any topic*) che orienta il *design* ontologico, prescrive che si possano (o che si debbano) estendere ontologie computazionali già esistenti con ulteriori specificazioni (asserzioni) derivanti da un dominio della conoscenza specializzato. Ad esempio, i materiali audiovisivi possono essere descritti utilizzando in toto o in parte ontologie già pubblicate, ma i descrittori che riguardano elementi peculiari, come il fatto che in una determinata risorsa audiovisiva compaiano strumenti musicali che hanno un “taxon HS”, una “definizione tipologica” e una “denominazione locale”, sono specifici del dominio etnomusicologico e pertanto devono essere ontologizzati per essere resi interoperabili. L’ontologia risultante – anche fermandosi a questo specifico esempio – è il prodotto di un’aggregazione di rappresentazioni ontologiche esistenti con

rami specifici nella prospettiva di un singolo dominio. Questo lavoro di razionalizzazione ha conseguenze desiderabili anche sul piano dell'interoperabilità tra archivi digitali che utilizzano *data model* diversi (su diversi domini o su uno stesso dominio).

L'idea di base delle ontologie è, come abbiamo visto, rappresentare classi e relazioni tra classi. In particolare è nostro interesse mostrare come concetti analoghi, espressi con termini diversi in *database* distinti, possano essere messi tra loro in relazione e i dati possano essere resi interoperabili.

Con riferimento alla figura che precede consideriamo, ad esempio, le registrazioni sonore contenute in due distinti *database* DB1 e DB2. È molto plausibile che gli elementi specifici, le singole registrazioni, non siano esattamente le stesse in ciascuno dei *database*: potranno esserci casi di coincidenza perfetta, casi di elementi simili, ma catalogati con titoli diversi o, con probabilità ancora maggiore, elementi del tutto distinti. Tuttavia, è facilmente intuibile che alla base delle due catalogazioni locali vi sia il "concetto" di "registrazione sonora" e che, sebbene la definizione che ciascun *database* ne potrebbe fornire possa essere specifica e non necessariamente coincidente con l'altra, entrambi i *database* si riferiscano a insiemi che contengono elementi analoghi (registrazioni sonore). Per poter collegare nell'ambiente del *semantic web* tali elementi, è necessario dichiarare questo rapporto di analogia. Ciò è reso possibile dalla creazione di un'ontologia, che è un'astrazione semantica e logica coerente e formalizzata, afferente a un dominio specifico della conoscenza. Nell'ontologia sarà necessario creare una classe gerarchicamente alta, una astrazione concettuale, che possa comprendere una grande varietà di elementi accomunati da caratteristiche generiche, ma esaustive, che si riferisca sia ai materiali contenuti in DB1 sia a quelli contenuti in DB2. Potremmo per esempio ipotizzare una classe "registrazioni sonore", la cui definizione contempla il fatto di essere registrazioni di suoni, indipendentemente dal mezzo con cui sono state realizzate, in quale formato, da chi o come e da quale ne sia il contenuto. A questo punto i singoli *database*, valutata la coerenza concettuale tra il proprio insieme di elementi e la definizione che ne esprime la classe dell'ontologia di dominio, potranno scegliere di "allineare" la propria classe locale a quella ontologica.

Per esemplificare: se DB1 definisce le proprie registrazioni sonore "Audio" e il DB2 "Sound Materials" (ma sempre di registrazioni sonore si tratta), entrambi i *database* possono concordare sul fatto che i propri elementi possono essere considerati parte della classe "Registrazioni sonore", in quanto ne rispettano i requisiti anche se localmente sono denominati in modo diverso ("audio" o "sound materials"). In un linguaggio più tecnico si dice che le due classi sono sottoclassi (*rdfs:subClassOf*) della superclasse presente nell'ontologia di dominio.

La metodologia di rappresentazione ontologica che abbiamo fin qui esemplificato ha conseguenze sul modo in cui si progettano standard descrittivi? A nostro parere sì e si tratta di un sostanziale cambiamento di paradigma. L'esistenza di norme a monte è sempre importante, e costruisce un orizzonte di riferimento. Si accetta tuttavia l'esistenza di diversi modelli concettuali (e relativi *data model*), sia a livello nazionale che a livello internazionale. È l'annoso tema della dispersione digitale affrontato in apertura, cui si

dà però una risposta concreta, ovvero che è possibile ottenere l'interoperabilità e l'aggregazione di dati a valle, analizzando in primo luogo ciò che esiste, e fornendone poi una rappresentazione ontologica: si riconoscono, una volta per tutte, l'esistenza di diversi modelli "indigeni" e attraverso i Linked Open Data consentiamo un allineamento ad ontologie "ecumeniche" che consentono l'interoperabilità.

A cosa serve questa impresa certo di non facile realizzazione? L'obiettivo è, di nuovo, l'arricchimento. Lo scopo è porre le basi per concepire un punto di aggregazione specifico di dominio che, attraverso *query* SPARQL condotte su più *endpoint* di dati LOD allineati a un'ontologia, consentano di interrogare simultaneamente diversi *resource center*, indipendentemente dal modello di dati locale.

1.5 Stato dell'arte

Il lavoro di *design* dell'ontologia e dell'interoperabilità prosegue attualmente su più direttrici. La prima è *organizzare l'informazione già presente in forma di LOD*. Qualche esempio può essere d'aiuto per comprendere come stiamo strutturando il nostro lavoro. Consideriamo il caso di risorse pubblicate di musica a stampa, per le quali esistono già norme internazionali, tradizioni descrittive e persino ontologie. Consideriamo un dettaglio, ovvero la descrizione di una presentazione musicale. Si tratta naturalmente di presentazioni della risorsa musicale applicabili in ambito euroculto e solo parzialmente in ambito etnomusicologico (dove è necessario introdurre specifici descrittivi). Ma per tenere l'esempio in limiti ragionevoli supponiamo di voler descrivere una musica per *bandella* che è a tutti gli effetti una risorsa pubblicata e che, almeno a livello della presentazione musicale, può essere descritta secondo le norme bibliografiche. Si prescrivono in questi casi termini tratti da un vocabolario standard, come "Cartina", "Conduttore", "Intavolatura", "Parte", ecc.

Nel CMS sarebbe auspicabile che il campo "presentazione musicale" (per musica euroculto) fosse collegato, già in fase di immissione dei dati, a un termine definito semanticamente. Esistono più fonti di LOD per le presentazioni musicali. Una fonte di comune accesso è Wikidata.²² Wikidata è una *knowledge base* che rappresenta, allo stato attuale, più di 108 milioni di entità relative a ogni dominio della conoscenza. Tali entità, identificate in modo univoco da un codice formato da una Q seguita da cifre, sono accessibili tramite un *endpoint* SPARQL. Una ricerca libera tra le entità rappresentate in questo enorme *repository* di dati semantici porta però a una sicura entropia. Una ricerca non qualificata, cioè eseguita tramite *label* come ad esempio "Partitura" (italiano), restituisce molti risultati con codici diversi, tra i quali potrebbero comparire delle P, ovvero codici che identificano le proprietà:

Partitura (Q23862474), un articolo nell'enciclopedia di Otto;

Partitura (Q18719616), opera di Granados;

²² <<https://www.wikidata.org/>>.


Freebase ID	/m/0nkr2 ↳ 1 reference
GitHub topic	sheetmusic ↳ 0 references
	sheet-music ↳ 0 references
Google Knowledge Graph ID	/g/1213jb18  ↳ 0 references
Gran Enciclopèdia Catalana ID	0129981 ↳ 0 references
Gran Enciclopèdia de la Música ID	2859 ↳ 0 references
Great Russian Encyclopedia Online ID (old version)	2322186 subject named as ПАРТИТУРА ↳ 0 references
IFLA value vocabularies ID	tos#a ↳ 0 references
JSTOR topic ID	sheet-music ↳ 0 references
KBpedia ID	SheetMusic ↳ 1 reference
Library of Congress Genre/Form Terms ID	gf2014027077 subject named as Scores ↳ 0 references
LoC and MARC vocabularies ID	mmusicformat/score ↳ 0 references

FIGURA 8. Parte delle asserzioni sull'elemento Q187947.

Partitura (Q187947), partitura musicale;
 Partitura (P3030), file digitale che contiene una partitura;
 Lineární partitura/Linear Score (Q76348952), una litografia di Grygar;
 ecc.

Il nostro termine partitura (Q187947), pertinente rispetto al lessico prescritto di ambito bibliografico, è disperso, per così dire, tra più di 1000 entità Wikidata. Una *query* SPARQL che interroghi l'*endpoint* Wikidata va di conseguenza raffinata al fine di poter ottenere risultati qualificati e stabili. Per nostra fortuna – più che una fortuna si tratta di una tendenza in atto che facilita enormemente il lavoro di qualificazione delle *query* – tra gli enti internazionali che convergono su Wikidata anche IFLA (International Federation of Library Associations and Institutions), la Library of Congress e molti altri enti, pubblici e privati, stanno riversando gran parte dei loro vocabolari su questa popolare *knowledge base*. L'elemento Q187947 (Partitura musicale) ad esempio, è oggetto di una serie di asserzioni (*statement*):

Tra le asserzioni mostrate alla figura precedente due sono particolarmente pertinenti per qualificare la nostra *query*, ovvero IFLA value vocabularies ID²³ e LoC and MARC vocabularies ID.²⁴ Entrambe le asserzioni collegano termini di vocabolario (gestiti da IFLA, MARC, LoC o di altri enti che producono *statement* simili). Nella maggior parte dei casi questi stessi termini sono pubblicati in forma di LOD e possiedono uno specifico URI. Il Q187947 corrisponde ad esempio al vocabolario di termini UNIMARC/IFLA Type of Score: <<https://www.iflstandards.info/unimarc/terms/tos#a>>.

La stessa tipologia di *statement* è prodotta per altre voci pertinenti (come ad esempio “Intavolatura” Q143217).²⁵ Wikidata, con le sue 100 milioni di entità, può essere utilizzata come *knowledge base* e fonte di LOD nel nostro progetto a patto che voci create da diversi utenti, per le ragioni più disparate, e per diversi domini della conoscenza, vengano qualificate e disambiguate attraverso una serie di *statement*. Ciò consente ad esempio di concepire questa *query* SPARQL:

```
SELECT DISTINCT ?item ?itemLabel
WHERE {
  ?item wdt:P9112 ?obj .
  FILTER(REGEX(STR(?obj), “^tos”))
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language “en” .
  }
}
```

²³ <<https://www.wikidata.org/wiki/Property:P9112>>.

²⁴ <<https://www.wikidata.org/wiki/Property:P4801>>.

²⁵ <<https://www.wikidata.org/wiki/Q143217>>.

Per eseguirla in tempo reale: <<https://w.wiki/6SXA>>. La *query*, il cui costrutto è spiegato in nota,²⁶ restituisce i termini prescritti dalle norme biblioteconomiche internazionali, ma reperiti in LOD come entità semanticamente definite. Una *query* interna al CMS

The screenshot shows a web interface for metadata management. At the top, there are navigation links: NUOVO, CERCA, GESTISCI, IMPORTA, CRONOLOGIA, and a search box. Below this is a control bar with 'Salva' (Save) and 'Annulla' (Cancel) buttons, and a 'Cancella' (Delete) button. The main content area is divided into sections. The first section is 'Presentazione musicale (forma codificata) (LoC da Wikidata)'. It contains a dropdown menu with 'partitur' selected. Below the dropdown are two links: 'partitura [http://www.wikidata.org/entity/Q187947] [http://id.loc.gov/vocabulary/mmusicformat/score] (rappresentazione grafica di una composizione che richiede l'esecuzione : partitura da tavolo [http://www.wikidata.org/entity/Q7673115] [http://id.loc.gov/vocabulary/mmusicformat/tablebk] (partitura che prevede una disposizione delle parti con orient esecutori posti attorno a un tavolo)'. The second section is 'Termine relativo al sistema di notazione (risorse etnomusicologiche)'. It contains a dropdown menu with 'Not set' selected. At the bottom of this section is a button 'Aggiungi Termine relativo alla notazione'.

FIGURA 9. Metadattazione con le presentazioni musicali in LOD (il dato semantico viene reperito in fase di catalogazione).

(in realtà poco più complessa dell'esempio precedente)²⁷ consente di collegare semanticamente una presentazione musicale in fase di immissione dati:

Quello che abbiamo presentato è l'esempio di una qualificazione delle *query* piuttosto semplice da gestire. Altri tipi di raffinamento richiedono un lavoro più intensivo (che spesso peraltro stimola riflessioni teorico-metodologiche). Il caso degli etnonimi e dei demonimi è certamente più spinoso. Limitiamoci in questa sede al caso dei demonimi. Per impostare una *query* che risponda a queste specificità, è necessario svolgere una ricerca preliminare per individuare come la *community* ha descritto gli elementi inerenti ai demonimi.

Esistono più strade da seguire, vanno quindi valutati i risultati per decidere quale intraprendere in base alle esigenze descrittive delle risorse catalografiche.

Se lo scopo fosse lo studio dei termini linguistici, si potrebbe impostare la seguente

²⁶ La clausola "SELECT DISTINCT ?item ?itemLabel" definisce le variabili che saranno restituite come risultato della *query*. In questo caso, la variabile "?item" rappresenta gli elementi di Wikidata che soddisfano i criteri della *query*, mentre la variabile "?itemLabel" rappresenta l'etichetta (in inglese) di ciascun elemento. La clausola "WHERE" specifica le condizioni che devono essere soddisfatte per selezionare gli elementi di Wikidata da restituire. In questo caso, la condizione è che ogni elemento debba avere una proprietà wdr:P9112 (IFLA value vocabularies ID) che abbia un valore che inizia con "tos" (il valore del *namespace* <http://iflstandards.info/ns/unimarc/terms/tos>). La clausola "FILTER" viene utilizzata per filtrare solo gli elementi di Wikidata che soddisfano la condizione specificata nella clausola "WHERE". La clausola "SERVICE wikibase:label" viene utilizzata per recuperare le etichette degli elementi di Wikidata in inglese (che è la lingua specificata nel parametro "wikibase:language"). La *query* restituirà le etichette degli elementi in inglese.

²⁷ La maggiore complessità della *query* è dovuta al fatto che si tengono in considerazione, nel nostro caso, i termini prescritti dalle norme italiane: <https://norme.iccu.sbn.it/index.php?title=Guida_musica/Appendici/Appendice_VII>.

query per recuperare i demonimi delle nazioni, ricorrendo ai dati lessicografici disponibili su Wikidata:

```
SELECT distinct ?dem ?demLabel ?statoSovranoLabel
WHERE {
  ?dem wdt:P31 wd:Q33829.
  ?L wdt:P6271 ?statoSovrano .
  ?L wdt:P5137 ?dem .
  ?statoSovrano wdt:P31 ?ist .
VALUES ?ist {wd:Q3624078 wd:Q6256}
SERVICE wikibase:label { bd:serviceParam wikibase:language "it, en". }
}
ORDER BY (?demLabel)
```

Nella *query* chiediamo di cercare tra i lessemi (?L) disponibili su Wikidata, tramite la specifica *proprietà* descrittiva “demonimo di” – P6271, tutti gli elementi (?statoSovrano) collegati che sono istanza di (P31) paese (Q6256) o stato sovrano (Q3624078), e tramite la proprietà “elemento per questo senso” – P5137, tutti gli elementi (?dem) corrispondenti al lessema, che sono istanza di popolazione (Q33829).

I risultati mostrano gli elementi (Q) corrispondenti ai demonimi (?demLabel) associati allo ?statoSovrano di riferimento, come mostra la *query*: <[https://w.wiki/8\\$W2](https://w.wiki/8$W2)>.

Un caso ulteriore, che richiede molto meno lavoro di qualificazione della *query* ma al contempo un ingente lavoro di *data curation*, è quello relativo alla classificazione HS. Avendo noi rappresentato l'albero HS su Wikidata è facile concepire questa *query*: <<https://w.wiki/6SYx>>. La stessa *query* è fondamento di un'altra, analoga solo leggermente più complessa, utilizzata per generare i risultati alla precedente Fig. 6 in fase di metadattazione dello strumento musicale. Tuttavia, come si può facilmente immaginare, è stato necessario un lungo lavoro, solo parzialmente automatizzato, per l'inserimento dell'intero albero italiano e inglese su Wikidata e quindi renderlo disponibile come dato semantico.

1.6 Test su diversi contesti etnografici

Gli esempi che abbiamo fin qui considerato insistono su piccole porzioni di un *data model*. Sia per la costruzione dell'ontologia, sia per il test del sistema catalografico, si sta attualmente procedendo al test del sistema di immissione dati e dell'ontologia su materiali il più eterogenei possibile e provenienti da diversi contesti etnografici. Questo lavoro più lungo e complesso riguarda l'intero *data model* e la sua flessibilità.

Procediamo prima di tutto alla catalogazione di materiali campione, i quali fungono da test per lo sviluppo ulteriore dell'ontologia. Per quanto il *data model* abbia superato una fase di progettazione teorica, gli errori concettuali sono all'ordine del giorno (ed è

normale che sia così se i test coinvolgono prospettive e contesti eterogenei e debbono, almeno a livello ontologico, poter integrare *data model* di altri *dataset*). Di tanto in tanto i test costringono a tornare indietro e ristrutturare il nostro modello di dati con l'obiettivo, che può sembrare paradossale, di diminuire la precisione del modello.

Risultati migliori possono talvolta essere ottenuti con ontologie che hanno un ampio potere euristico – un modello deve cioè essere in grado di descrivere in modo coerente e non ambiguo – ma mantengono equilibrio tra gli opposti rappresentati a un estremo dall'eccesso di dettagli e all'altro da una descrizione talmente generale da risultare inservibile. Descrivere i singoli *mudra* di un *kathakali* è forse eccessivo: le informazioni diventano troppo specialistiche e riproducono gli stessi inconvenienti di standard catalografici del passato, ove la compilazione di un record si trasformava in una lunga e laboriosa attività, per certi versi paragonabile a quella che impegna lo studioso nella scrittura di un libro o di un articolo. All'altro estremo, dire che si tratta semplicemente di una danza non ha alcun potere euristico per classificare, poniamo, un documento audiovisivo di *kathakali*. Come si è detto, il lavoro di raffinamento dell'ontologia prevede fasi di ripensamento nelle quali si adottano posizioni intermedie tra l'eccesso specialistico e la descrizione troppo generica e questo è certamente l'equilibrio più difficile da mantenere in fase di *design* di un'ontologia.

Il rilascio di una prima *draft* dell'ontologia è previsto al termine del 2024. Contestualmente l'Unità di Torino rilascerà in accesso aperto il sistema di *data entry* (basato su Collective Access).

Riferimenti

- Allemang, Dean, Jim Hendler, e Fabien Gandon
 2020 *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL*, New York, Association for Computing Machinery.
- Berners-Lee, Tim
 2001 *L'architettura del nuovo web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*, in collaborazione con Mark Fischetti, Milano, Feltrinelli.
- Biagetti, Maria Teresa
 2016 "Un modello ontologico per l'integrazione delle informazioni del patrimonio culturale: CIDOC-CRM", *JLIS.it*, VII/3: 43-77.
- Bizer, Christian, Tom Heath Tom e Tim Berners-Lee
 2009 "Linked Data – The Story So Far", *International Journal on Semantic Web and Information Systems*, VI/3: 1-22.
- Birley, Margaret, Arnold Myers, e Rupert Shepherd
 2020 "The revision of the Hornbostel-Sachs classification in 2011 by the MIMO consortium", in Cristina Ghirardini (a cura di), *Reflecting on Hornbostel-Sachs's Versuch a century later. Proceedings of the international meeting Venice, 3-4 July 2015*, Venezia, Edizioni Fondazione Levi: 167-180.

- Crupi, Gianfranco
2013 “Beyond the Pillars of Hercules: Linked data and Cultural heritage”, *JLIS.it*, IV/1, gennaio: 25-49.
- Della Valle, Emanuele, Irene Celino, e Dario Cerizza
2009 *Semantic web. Dai fondamenti alla realizzazione di un'applicazione*, Milano, Pearson.
- Di Noia, Tommaso, Roberto De Virgilio, Eugenio Di Sciascio, e Francesco M. Donini
2013 *Semantic web: tra antologie e open data*, Milano, Apogeo.
- DuCharme, Bob
2013 *Learning SPARQL, querying and updating with SPARQL 1.1*, Sebastopol, O'Reilly Media Inc.
- Gruber, Tom R.
1993 “A translation approach to portable ontologies”, *Knowledge Acquisition*, V/2: 199-220.
- Gruber, Tom R.
<<http://www.ksl.stanford.edu/kst/what-is-an-ontology.html>>.
- Guercio, Maria
2010 *Archivistica informatica. I documenti in ambiente digitale*, Roma, Carocci.
- Guerrieri, Claudia
2022 “Linked Open Data e rappresentazione del patrimonio culturale: un caso applicativo per diffondere la conoscenza dei beni culturali ecclesiastici nel web semantico”, *Digitalia*, 1: 184-202.
- Guerrini, Mauro
2015 “Classificazioni del sapere: Web semantico, Linked data e ontologie. Il ruolo rinnovato delle biblioteche nella trasmissione della conoscenza registrata”, in Fiammetta Sabba (a cura di), *Noetica versus informatica: le nuove strutture della comunicazione scientifica: atti del convegno internazionale, Roma, Tempio di Adriano, 19-20 novembre 2013*: 145-155.
- Guerrini, Mauro, e Tiziana Possemato
2012 “Linked data: un nuovo alfabeto del web semantico”, *Biblioteche oggi*, aprile: 7-15.
2015 *Linked data per biblioteche, archivi e musei*, Milano, Editrice Bibliografica.
- Hornbostel, Erich M. von, Curt Sachs, e Febo Guizzi
2020a “Classification of musical instruments”, in Cristina Ghirardini (a cura di), *Reflecting on Hornbostel-Sachs's Versuch a century later. Proceedings of the international meeting Venice, 3-4 July 2015*, Venezia, Edizioni Fondazione Levi: 227-280.
2020b “Sistematica degli strumenti musicali”, in Cristina Ghirardini (a cura di), *Reflecting on Hornbostel-Sachs's Versuch a century later. Proceedings of the international meeting Venice, 3-4 July 2015*, Venezia, Edizioni Fondazione Levi: 281-341.
- Hooland, Seth van, e Ruben Verborgh
2014 *Linked data for libraries, archives and museums. How to clean, link and publish your metadata*, Londra, Facet.
- International Council on Archives (ICA)
2021 *Record in Contexts. A Conceptual Model For Archival Description. Consultation Draft v.0.2*, <https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf>.
- Lyons, Bertrand, e Kara Van Malssen
2016 *BIBFRAME AV Assessment: Technical, Structural, and Preservation Metadata*, Library of Congress, <<https://www.weareavp.com/wp-content/uploads/2019/03/bf-avtechstudy-01-04-2016.pdf>>.
- Martinelli, Luca
2016 “Wikidata: la soluzione wikimediana ai linked open data”, *AIB Studi*, LVI/1: 75-85.

- Mazzini, Silvia, e Francesca Ricci
 2011 “EAC-CPF Ontology and Linked Archival Data”, in *Proceedings of the 1st International Workshop on Semantic Digital Archives, Berlin 29/9/2011 (SDA 2011)*: 72-81, <<http://ceur-ws.org/Vol-801/paper6.pdf>>.
- Meandri, Ilario e Cristina Ghirardini
 2020 (a cura di) *SAMIC. Sound Archives and Musical Instruments Collections. Documentazione del progetto e guida al sistema di catalogazione*, Lucca, LIM.
- Osorio-Kupferblum, Naomi
 2016 “Aboutness”, *Analysis*, LXXVI/4: 528-546.
- Rondeau, Sophie
 2014 “The Life and Times of Aboutness: A Review of the Library and Information Science Literature”, *Evidence based library and information practice*, IX/1: 14-35.
- Signore, Oreste e Ufficio W3C
 2015 “RDF per la rappresentazione della conoscenza” <https://www.researchgate.net/publication/228596094_RDF_per_la_rappresentazione_della_conoscenza>.
- Solodovnik, Iryna
 2011 “Comunicazione e ricerca semantica di contenuti informativi: tra metadati, Linked open Data e Ontologie”, *Bibliotime*, XIV/2, <<http://eprints.rclis.org/15966/1/Metad-LOD-Ontologie.pdf>>.
- Tomasi, Francesca
 2013 “Le edizioni digitali come nuovo modello per dati di autorità concettuali”, *JLIS.it*, IV/2.
 2017 “Archivi di persona in linked open data: il modello concettuale come strumento di integrazione nei GLAMs”, *Rivista di biblioteconomia e scienze dell'informazione*, LVII/2 <<http://aibstudi.aib.it/article/view/11647/10983>>.
- Tomasi, Francesca, e Marilena Daquino
 2015 “Modellare ontologicamente il dominio archivistico in una prospettiva di integrazione disciplinare”, *JLIS.it*, VI/3: 13-40.
- Valacchi, Federico
 2016 “Pezzi di cose nel mondo. Il processo di integrazione delle descrizioni archivistiche nei sistemi interculturali”, *JLIS.it*, VII/2.