

Inside the Meanings. The Usefulness of a Register of Ontologies in the cultural Heritage Sector

Chiara Veninata^(a)

a) Istituto Centrale per il Catalogo e la Documentazione (ICCD-MiBACT), <http://orcid.org/0000-0003-0981-1726>

Contact: Chiara Veninata, chiara.veninata@beniculturali.it

Received: 27 January 2020; Accepted: 18 February 2020; First Published: 15 May 2020

ABSTRACT

The article deals with the topic of the semantic web and the publication in linked open data of information relating to cultural heritage. In particular, the article analyzes the ontology registers, i.e. those tools that formally describe the ontological models available on the web and facilitate their retrieval and evaluation, encouraging their reuse and facilitating semantic alignment and interoperability processes. The ontology registers respond effectively to the absence of reference and orientation tools in the conceptual modeling processes of information and have been successfully tested in different domains, but are still unpublished in the cultural sphere.

The examination of the initiatives carried out in the last decade in the field of cultural heritage has clearly highlighted the lack of a consolidated epistemological structure in the conceptual modeling of information resources, despite the numerous ontologies created according to the multiple linked open data publication projects. Consequently, it is often difficult to fully understand all the ontologies available in relation to your area of interest and to obtain in a smooth and systematic way a reliable assessment of their representative capacity and their degree of semantic interoperability.

The analysis of the main registers of ontologies so far created outside the domain of cultural heritage has made it possible to identify and define the requirements of a register of ontologies for cultural heritage, and to elaborate the relative ontology. The clarification of the requirements also took into account a peculiar function that the registers could play in the cultural domain, as tools to support some features of a digital library of cultural heritage.

KEYWORDS

Ontology; Cultural heritage; Linked open data.

CITATION

Veninata, C. "Inside the Meanings. The Usefulness of a Register of Ontologies in the cultural Heritage Sector." *J LIS.it* 11, 2 (May 2020): 45–58. DOI: [10.4403/jlis.it-12624](https://doi.org/10.4403/jlis.it-12624).

Premessa

Il settore dei beni culturali rappresenta una delle aree più promettenti e stimolanti per quanto concerne l'applicazione degli standard e delle tecnologie del *semantic web* e, più in generale, la sperimentazione di nuove soluzioni in grado di connettere tipologie di informazioni intrinsecamente eterogenee. Istituzioni culturali quali biblioteche, archivi e musei, stanno rivolgendo una crescente attenzione alle nuove tecnologie del *web* per offrire all'utenza funzionalità di ricerca e di navigazione che, sfruttando le relazioni semantiche tra dati anche di diversa origine, permettono sia di migliorare l'interoperabilità tra sistemi differenti sia di accrescere le possibilità di integrazione, recupero e fruizione delle informazioni. L'utilizzo del linguaggio RDF¹ e la disponibilità di ontologie, costituiscono la base tecnologica necessaria per la pubblicazione sul *web* di dati secondo il paradigma dei *linked open data* (LOD), ovvero dati in formato aperto e collegati tra di loro. Le ontologie, intese come dizionari esterni rappresentativi ed esplicativi dei dati, consentono di rappresentare le risorse tramite la descrizione delle loro caratteristiche (o attributi), l'identificazione delle relazioni esistenti tra esse e della semantica che lega tali entità. Lo spazio di riferimento semantico, ovvero il settore in cui tali risorse e tali relazioni sono significative, è detto anche dominio dell'ontologia e tiene conto dello specifico contesto e dello specifico punto di vista sulla base dei quali la realtà è osservata. Il dominio dei beni culturali è, per sua natura, piuttosto complesso e l'esplicitazione semantica delle relazioni tra le sue numerose componenti (beni di natura biblioteconomica, archivistica e museale, fotografie, beni archeologici, beni immobili, beni etnoantropologici materiali e immateriali, architetture e poi artisti, autori, editori etc.) può essere più o meno significativamente espressa anche a seconda dell'ontologia che si prende a modello.

Il processo di costruzione di una ontologia comporta un coinvolgimento di risorse (persone, tempi, costi) di solito piuttosto elevato. Solitamente si tratta di circoscrivere un determinato dominio della conoscenza, definirne i concetti (classi) e organizzarli in una gerarchia. Successivamente si definiscono gli attributi e le relazioni (proprietà o *slot*) tra concetti e le restrizioni su di essi. Infine, si individuano le istanze dei concetti, popolando l'ontologia.

La definizione dei concetti e delle relazioni tra concetti tra gli esperti di un determinato dominio richiede solitamente una enorme attività di negoziazione tra pari per la conciliazione di moltissime esigenze e punti di vista diversi (basti semplicemente pensare a tutti i punti di vista rappresentati dai diversi sistemi semantici associati ai vari sistemi linguistici).

Una pratica raccomandata da buona parte della letteratura sugli standard – in particolare in ambito *semantic web* e *linked open data* – è quella che prevede che, prima di creare una nuova ontologia, occorra valutare il riuso di ontologie disponibili per un determinato dominio di conoscenza facendo preferibilmente riferimento a ontologie standard o, in secondo luogo, a ontologie di dominio molto conosciute, documentate e mantenute da agenzie reputate affidabili. Raccomandazioni sul riuso di ontologie sono fornite anche dalle “Linee guida per l'interoperabilità semantica attraverso i *linked open data*” pubblicate nel novembre 2012 dall'Agenzia per l'Italia Digitale.²

¹ Cfr. Resource Description Framework, <https://www.w3.org/RDF/>.

² Cfr. https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf.

Ciononostante, il carattere distribuito dei processi di modellazione e pubblicazione di ontologie ha portato allo sviluppo di varie ontologie per domini molto simili. Questa tendenza, pur non costituendo un problema di per sé, rischia di indebolire le possibilità di realizzare l'interoperabilità semantica tra i dati, che è invece favorita dalla conoscenza completa dei modelli che stanno alla base della pubblicazione dei dati.

Linked open data e ontologie in ambito culturale

Prima di entrare nel merito di strumenti quali i registri di ontologie ed evidenziarne l'utilità nel web semantico, può risultare efficace partire dallo stato dell'arte dei progetti di pubblicazione di *linked open data* in ambito culturale e delle ontologie in essi utilizzate. Una prima analisi prende avvio con la *Review on linked open data sources*, effettuata nell'ambito del Progetto *Athena Plus* dell'ottobre 2013;³ essa fu successivamente integrata e aggiornata da una ricerca di OCLC condotta tra luglio e agosto 2014 e poi ripubblicata con alcune correzioni nel 2015.⁴ Quest'ultima riguardava un primo censimento di progetti per la pubblicazione di *linked open data* in ambito prevalentemente biblioteconomico. Le indagini di OCLC sono state effettuate diffondendo il *link* alla *survey* su molte *listserv* e su Twitter, contando quindi sulle potenzialità dei *social media* per raggiungere il maggior numero possibile di istituzioni.

I risultati della prima indagine del 2014, portata a termine nel 2017,⁵ sono stati successivamente confrontati con la revisione⁶ della medesima indagine condotta sempre da OCLC tra il 17 aprile e il 25 maggio 2018. All'indagine del 2018 hanno risposto 81 istituzioni che hanno segnalato un totale di 104 progetti, rispetto alle 71 istituzioni che ne avevano segnalati 112 nel 2015. Dei suddetti 104 progetti, solo 42 erano già stati descritti in precedenza. Il 75% delle 104 implementazioni sono attive e il 40% di queste è attiva da più di quattro anni. Dei 104 progetti descritti, 29 dichiarano di utilizzare *endpoint* SPARQL per consentire l'accesso diretto ai *linked data* e 33 dichiarano di fornire accesso ai *file dump*. Le istituzioni italiane coinvolte nell'indagine OCLC del 2018 sono solo 4: si tratta della Biblioteca della Camera dei deputati, di Casalini Libri (SHARE-VDE group), del Coordinamento delle Biblioteche Speciali e Specialistiche di Torino (CoBIS) e dell'Università degli Studi Roma TRE.

Mentre sul fronte internazionale il presente contributo non aggiunge molto al sondaggio di OCLC, salvo una verifica puntuale sulla reale disponibilità degli *endpoint*, essa arricchisce notevolmente l'indagine sul fronte italiano. Per quanto riguarda l'indagine OCLC, in questa ricerca si tiene conto solo dei progetti che hanno riguardato la produzione di dati descrittivi del patrimonio culturale⁷ e che hanno dichiarato la presenza di un *endpoint* SPARQL o di *file dump*, ovvero di quelli in cui sia stato possibile verificare l'utilizzo delle ontologie dichiarate e la conseguente modellazione dei dati,

³ Disponibile all'indirizzo <http://www.athenaplus.eu/getFile.php?id=190> (consultato il 19/12/2019).

⁴ L'indagine del 2014 non considerava alcune realtà rilevanti come le biblioteche nazionali di Francia e Germania. Così il sondaggio è stato ripetuto tra il 1° giugno e il 31 luglio 2015.

⁵ Cfr. Smith-Yoshimura, *Analysis of International Linked Data Survey for Implementers*, D-Lib Magazine, 2017, 22 (7/8), pp. 141-167, disponibile all'indirizzo <http://doi.org/10.1045/july2016-smith-yoshimura>.

⁶ Disponibile all'indirizzo <https://www.oclc.org/research/themes/data-science/linkddata/linked-data-survey.html>.

⁷ Si escludono dunque progetti, come quello della Charles University di Praga (cfr. <https://etl.linkedpipes.com>) o della Cornell University (cfr. <http://ld4p.org/>) o della North Rhine-Westphalian Library Service Center (cfr. <http://lobid.org>), che riguardano modalità innovative di produzione automatica di *linked open data*.

escludendo quindi i progetti che dichiarano di utilizzare *linked data* solo nel *back end*⁸ o quelli in cui l'accesso ai dati è ancora dichiarato privato per questioni tecniche⁹ o risulta non raggiungibile alla data del 31/12/2019¹⁰ o, infine, i progetti di digitalizzazione in cui la messa a disposizione di LOD è di fatto delegata al portale.

L'indagine evidenzia 57 progetti di pubblicazione di *linked open data* riconducibili all'ambito culturale.¹¹ Tra questi, i 15 progetti italiani censiti risultano tutti mantenuti e dotati di un *endpoint* SPARQL funzionante.

⁸ È il caso dei progetti di biblioteca digitale della Agencia Española de Cooperación Internacional para el Desarrollo (AECID), o del catalogo di Anythink Libraries, o della Digital Public Library of America (cfr. <http://dp.la/>), della biblioteca del Ministero della Difesa spagnolo ecc.

⁹ Si vedano ad esempio i progetti della George Washington University, il progetto della National Diet Library riguardante la "Nippon Decimal Classification" (NDC), della National Library of Medicine (USA), della University of South Florida St. Petersburg, della Memorial University of Newfoundland, della National Library of Portugal, della National Library of Scotland, della National Library of Wales.

¹⁰ È il caso della Australian National University.

¹¹ Si fornisce di seguito l'elenco dei progetti: American Numismatic Society (<http://numismatics.org/archives> e <http://numismatics.org/authorities>), Amsterdam museum (<https://www.amsterdammuseum.nl/open-data>), Archaeology Data Service (<http://archaeologydataservice.ac.uk/research/stellar/>), Archivio Centrale dello Stato (<http://dati.acs.beniculturali.it>), Archivio storico della Presidenza della Repubblica italiana (<https://archivio.quirinale.it/aspr/redazione/linked-open-data>), Bavarian State Library (lod.b3kat.de), Biblioteca de Galicia (http://biblioteca.galiciiana.gal/gl/datos_abiertos/datos_abiertos.cmd), Biblioteca Nacional de España (<http://datos.bne.es/>), Bibliotheque Nationale de France (<http://data.bnf.fr/>), Biblioteca Nazionale Centrale di Firenze, Soggettario (<http://thes.bncf.firenze.sbn.it/thes-dati.htm>), British Library, British National Bibliography (<http://bnb.data.bl.uk>), British Museum (<http://collection.britishmuseum.org>), Camera dei deputati (<http://dati.camera.it/>), Carnegie hall (<http://data.carnegiehall.org>), Casalini Libri (www.share-vde.org), Centro di Documentazione Ebraica Contemporanea CDEC (<http://dati.cdec.it/>), Claros Project (<https://clarosdata.wordpress.com/>), CoBIS - Coordinamento delle Biblioteche Speciali e Specialistiche di Torino (<https://dati.cobis.to.it/>), Consiglio Nazionale delle Ricerche (<http://data.cnr.it/>), Corago LOD (<http://www.disit.org/corago/>), Cultura Italia (<http://dati.culturaitalia.it/>), Data Archiving and Networked Services (DANS), Royal Netherlands Academy of Arts and Sciences (<http://www.cedar-project.nl/>), Deutsche Nationalbibliothek (<http://www.dnb.de/EN/lids>), Europeana Foundation (<http://europeana.eu/>), Fondazione Zeri (<http://data.fondazionezeri.unibo.it/>), Fundación Ignacio Larramendi (http://www.larramendi.es/i18n/cms/elemento.cmd?id=estaticos/paginas/Biblioteca_Virtual_Ignacio_Larramen.html), German National Library (<http://www.dnb.de/lids>), Getty Vocabularies (<http://vocab.getty.edu/>), Goldsmiths' College (<http://slickmem.data.t-mus.org/>), Hellespont Project (<http://hellespont.dainst.org/startpage/index.html>), Historic Environment Scotland (<http://heritagedata.org/live/schemes/scapa.html>), Istituto Centrale per gli Archivi – Sistema Archivistico Nazionale (<http://san.beniculturali.it/web/san/dati-san-lod>), Istituto Centrale per il Catalogo e la Documentazione – Catalogo generale dei beni culturali (<http://dati.beniculturali.it/arco/>), Istituto per i beni artistici culturali e naturali (IBACN) della Regione Emilia-Romagna (<http://ibc.regione.emilia-romagna.it/servizi-online/lod>), Library of Congress (<http://id.loc.gov/> e <http://bibframe.org/>), Linked Jazz (<https://linkedjazz.org/>), Linking Lives (<http://archiveshub.ac.uk/linkinglives/>), linked open data for Academia (LOD.AC) project (<http://lod.ac/>), Ministero per i beni e le attività culturali e per il turismo (<http://dati.beniculturali.it>), National Diet Library (<http://id.ndl.go.jp/auth/ndla>), National Library of Finland (<http://data.nationallibrary.fi>), National Library of Medicine (<https://id.nlm.nih.gov/mesh/>), National Széchényi Library (<http://v.mek.oszk.hu/FlintSparqlEditor/index-mek.html>), Nomisma (<http://nomisma.org>), North Carolina State University Libraries (<http://www.lib.ncsu.edu/ld/onld/>), NTNU (Norwegian University of Science and Technology) University Library (<http://www.ntnu.no/ub/digital/>), OCLC (<http://www.worldcat.org/> e <http://viaf.org>), Oslo Public Library (<http://data.deichman.no/>), Progetto Reload (<http://labs.regesta.com/reloadProject/client/>), Rijksmuseum (<https://datahub.io/rasvaan/201805-rma-collection>), Russian Linked Culture Cloud (<http://culturecloud.ru/>), Springer Nature (<http://www.springernature.com/scigraph>), University of Alberta Libraries (<http://canlink.library.ualberta.ca> e <http://dx.doi.org/10.7939/DVN/URXSGC>), University of California - Los Angeles

Volendo classificare a grandi linee la tipologia dei dati descrittivi del patrimonio culturale riconducendoli a specifici ambiti disciplinari, sul totale dei progetti censiti, 34 pubblicano dati bibliografici, 30 *authority file*, 13 pubblicano dati su opere d'arte, 11 pubblicano dati archivistici. Le altre tipologie di dati pubblicati sono dati archeologici (6), musicologici (2), vocabolari (9), dati biografici (9), storici (8) e geografici (4).

Se si passa all'analisi dei modelli ontologici sottesi alla descrizione del patrimonio culturale, dei 57 progetti censiti 38 dichiarano di usare le ontologie Dublin Core e DC Terms, 14 l'ontologia CIDOC-CRM, 12 la Bibliographic ontology, 9 BIBFRAME, 6 l'ontologia EDM, 7 la Biographical Ontology, 10 RDA; ben 27 progetti hanno invece sviluppato proprie ontologie.

Dunque l'analisi rivela una realtà composita e interessante, segno anche della ripresa di un vivace dibattito non solo relativo alle potenzialità connesse con le nuove tecnologie ma che coinvolge anche i modelli descrittivi, con dei ripensamenti in chiave critica circa la relativa produzione scientifica e tecnica dei decenni precedenti.

Fino alla fine degli anni Novanta il dibattito scientifico sui modelli descrittivi è stato infatti caratterizzato dalla ricerca e pubblicazione di tracciati dati ritenuti adeguati – in una determinata comunità di riferimento – alla descrizione del patrimonio culturale e specifici per la descrizione bibliografica, per la descrizione archivistica e per la descrizione di oggetti cosiddetti museali.

Frutto di tale dibattito è stata la produzione di numerosi modelli concettuali: una ricerca condotta nel 2009 da Jenn Riley dell'Indiana University Libraries White Professional Development Award ne contava circa 105 tra quelli maggiormente diffusi a livello internazionale. Jenn Riley ne fornisce una spettacolare rappresentazione grafica,¹² evocativa della complessità della situazione, organizzando i modelli a seconda dei domini specifici (biblioteche, archivi, musei), delle comunità di riferimento, della funzione svolta (modelli concettuali, vocabolari controllati), delle tecnologie utilizzate (schemi XSD, ontologie, thesauri SKOS) e dello scopo per cui sono stati concepiti (metadati tecnici, descrittivi, strutturali, metadati per la conservazione a lungo termine etc.). A livello nazionale tale complessità si complica ulteriormente: basti pensare che solo l'Istituto Centrale per il Catalogo e la Documentazione fornisce ulteriori 30 tracciati di metadati descrittivi per la catalogazione dei beni archeologici, storico-artistici, architettonici etc.¹³ e che, in generale, gli Istituti centrali del MiBAC hanno spesso adottato profili nazionali che si presentano come adattamenti dei tracciati internazionali (ad esempio, i tracciati del catalogo del Sistema archivistico nazionale – CAT-SAN – basati sullo standard internazionale EAD o lo standard MAG per la metadattazione delle immagini digitali pubblicato da ICCU e basato sullo standard internazionale METS).

Tale proliferazione nasce indubbiamente dal desiderio di esprimere con la massima analiticità possibile tutto il potenziale informativo dei dati raccolti nella fase di schedatura e di catalogazione dei

(<http://link.library.ucla.edu>), University of Nevada, Las Vegas (<https://www.library.unlv.edu/linked-data>), University of Oxford (oxlod.eng.ox.ac.uk), Yale Center for British Art (<http://britishart.yale.edu/collections/using-collections/technology/linked-open-data>).

¹² Cfr. Jenn Riley, *Seeing standards, A Visualization of the Metadata Universe*, disponibile all'indirizzo <http://jennriley.com/metadatamap/>.

¹³ Cfr. Maria Letizia Mancinelli, *Gli standard catalografici dell'Istituto Centrale per il Catalogo e la Documentazione*, in Roberta Tucci, *Le voci, le opere e le cose. La catalogazione dei beni culturali demioetnoantropologici*, Roma, Istituto centrale per il catalogo e la documentazione - Ministero dei beni e delle attività culturali e del turismo, 2018, pp. 279–302 disponibile all'indirizzo <http://www.iccd.beniculturali.it/getFile.php?id=6670>.

beni culturali senza i vincoli, spesso restrittivi, imposti dall'adesione a tracciati *standard* esterni. Ciò è spesso avvenuto anche al costo della perdita del potenziale di interoperabilità sintattica che proprio l'adesione agli standard mirava a garantire nei progetti di condivisione dei dati.

Dal 2010 il dibattito intorno alle potenzialità del nuovo *web* (*web* dei dati o *web* semantico), ha contribuito a rinverdire il confronto tra gli esperti che si sono occupati delle metodologie connesse alla descrizione del patrimonio culturale. I primi tentativi di esplorare i formalismi e le tecnologie connessi al *web* semantico – ed in particolar modo le ontologie ed i *linked open data* – sono in parte riconducibili a due opposte tendenze: la prima è rappresentata da tentativi di traduzione, a volte acritica, dei tracciati standard di metadati in chiave ontologica; la seconda è ravvisabile nella volontà di emanciparsi dai vincoli spesso presenti negli standard del settore per pubblicare i dati sul patrimonio culturale sulla base di modelli concettuali leggeri, anche non concepiti specificamente per le descrizioni dei beni culturali e delle loro collezioni (vedi, ad esempio, il caso di OCLC che pubblica l'intero catalogo bibliografico usando l'ontologia *schema.org*, pensata per una più efficiente indicizzazione dei siti *web* da parte dei motori di ricerca). In questa sede, quello che preme sottolineare è come, anche in questa nuova fase del dibattito teorico e scientifico, si assista, da una parte, al ricrearsi, nei settori librario, archivistico e museale, di atteggiamenti scarsamente propensi alla contaminazione in nome delle specificità delle varie discipline e metodologie descrittive; dall'altra parte, si nota un fiorire di iniziative scollegate le une dalle altre che si dedicano o a produrre nuovi modelli concettuali spesso estremamente complessi per finalità legate a specifici progetti di ricerca o ad adottare acriticamente modelli ontologici già disponibili per la rapida pubblicazione di dati aperti e facilmente relazionabili. Quello che è sembrato mancare fino ad ora è una seria riflessione suffragata da un numero adeguato di casi di studio sulle reali implicazioni che la scelta di un modello ontologico comporta per i *linked open data*, in termini di qualità dei collegamenti tra dati e quindi di accrescimento del potere conoscitivo offerto a chi consulta i dati attraverso i portali *web*.

Chi si è occupato di *linked open data* sa bene che la modellazione ontologica non è mai del tutto indifferente rispetto all'utilizzo che dei dati si intende fare e che spesso la scelta di una determinata ontologia è funzionale proprio ai molteplici utilizzi dei dati ritenuti plausibili dai modellisti. D'altra parte, neanche il ricorso ad ontologie ritenute i *golden standard* relativamente ad un determinato dominio, come ad esempio viene considerato CIDOC in ambito culturale, rende immuni da possibili scelte errate nella modellazione dei dati.¹⁴

¹⁴ È, ad esempio, il caso dei *linked open data* relativi al portale Cultura Italia la cui modellazione sulla base di CIDOC risulta infarcita di errori sia sintattici che semantici, dovuti ad una errata interpretazione dello *standard*. Cfr. Valdimir Alexiev, *How Not To Do Linked Data*, disponibile all'indirizzo <https://gist.github.com/VladimirAlexiev/090d5e54a525d57acb9b366121e77573>. Ed è anche il caso della modellazione dei dati del British Museum, che pur corretti da un punto di vista sintattico e semantico, vengono spesso citati tra gli addetti ai lavori come un esempio di modellazione eccessivamente complessa e astrattamente concettuale che rende difficoltosi l'*interlinking* con altri dati e l'effettuazione di *query* sull'*endpoint*: si fa riferimento alle criticità messe in evidenza nell'ambito di specifici tavoli di lavoro durante il quarto LOD-LAM Summit, tenutosi a Venezia presso la Fondazione Giorgio Cini, a giugno 2017 (cfr. <https://summit2017.lodlam.net/>), a cui la sottoscritta ha partecipato come delegata per l'Università La Sapienza di Roma. Un tavolo di lavoro, in particolare, era dedicato all'analisi dell'utilizzo di CIDOC come ontologia per la pubblicazione di LOD in ambito culturale (cfr. le note sulle sessioni di lavoro <http://bit.ly/2sPCj3e>, ultima consultazione 27/04/2019).

Il ruolo dei registri di ontologie e l'esperienza dei registri di metadati

I registri di ontologie, intesi come *repository* delle ontologie, si prefiggono lo scopo di aiutare il ricercatore ad individuare l'ontologia o le ontologie più adatte al proprio scopo e a valutarne la qualità e la riusabilità, sulla base di criteri di valutazione basati su parametri quanto più oggettivi possibili, ovvero su metriche misurabili.

Alcune metriche sono basate su statistiche e altre sono basate sul controllo di qualità. Per quanto concerne le prime, esse calcolano, ad esempio, il numero delle classi non anonime presenti nell'ontologia, numero delle proprietà (o *slot*), numero degli individui¹⁵, massima profondità delle relazioni gerarchiche tra le classi,¹⁶ numero medio e massimo delle classi sorelle ovvero sullo stesso livello nell'albero. Per quanto concerne invece le metriche basate sul controllo di qualità, esse si basano su parametri quali, ad esempio, classi con una sola sottoclasse (situazione che spesso indica una gerarchia poco specificata o una poco appropriata distinzione tra la classe e la sottoclasse), classi con più di 25 sottoclassi (spesso una classe così articolata è eligibile per distinzioni e categorizzazioni ulteriori), classi senza definizioni.¹⁷

Tuttavia, la sola disponibilità di elementi quantitativi o comunque oggettivi non appare sufficiente a decretare la bontà di un modello rispetto ad un altro. La valutazione non prescinde mai da elementi soggettivi e occorre capire se esistano strumenti in grado di agevolare anche le interpretazioni da parte dei modellisti e se i registri di ontologie possono rivelarsi utili in tal senso.

I registri di ontologie possono consentire di valutare come un concetto identificato da un utente sia stato espresso in varie ontologie, rendendo esplicite le equivalenze e gli allineamenti tra i vari modelli concettuali registrati, con strumenti in grado non solo di esplicitare gli assiomi del linguaggio RDFS e OWL appositamente previsti per la gestione degli allineamenti concettuali (relazioni di sottoclasse/proprietà o di equivalenza tra classi o proprietà di due ontologie differenti) ma in grado anche di prevedere costrutti sintattici per l'esplicitazione di relazioni più complesse (ad es. un concetto di una ontologia che corrisponde a più concetti o ad una concatenazione di classi e proprietà in un'altra ontologia).

I registri di ontologie sono fortemente correlati all'esperienza dei registri di metadati.

Uno degli studi più completi riguardo l'utilità dei registri di metadati nasce nell'ambito delle *digital humanities*. Si tratta del "*Principles of Metadata Registries. A White Paper of the DELOS Working Group on Registries*",¹⁸ sponsorizzato dalla *DELOS Network of Excellence on digital libraries*, una iniziativa della Commissione Europea nata per promuovere la ricerca e la cooperazione internazionale nel campo delle *digital libraries*.

¹⁵ Questa metrica è significativa solo per specifici linguaggi di rappresentazione delle ontologie dato che per esempio nelle ontologie di formato OBO non sono presenti individui.

¹⁶ Per le ontologie espresse in OWL e RDFS, considera come relazione gerarchica la relazione "is-a". Per le ontologie nel formato OBO, sono considerate relazioni gerarchiche le relazioni "is-a", "has-part", inversa di "develops-from".

¹⁷ Alcune di queste proprietà sono definite nella OMV Ontology ed hanno omv:prefix; il resto sono definite localmente nella BioPortal Metadata Ontology (metrics:prefix).

¹⁸ Baker, Thomas *et al.* "Principles of Metadata Registries A White Paper of the DELOS Working Group on Registries." (2003), disponibile all'indirizzo <https://pdfs.semanticscholar.org/01ea/e200c915fbb38faf2584e87230bb15d2d683.pdf> (consultato il 26/12/2018).

Rispetto ai registri di metadati, i registri di ontologie possono essere utilizzati, sfruttando le caratteristiche dei linguaggi RDF e OWL, per rendere comprensibili i modelli concettuali formalizzati in ontologie oltre che agli esseri umani anche alle macchine, per favorire l'esplicitazione delle relazioni semantiche tra classi e proprietà delle ontologie al fine di effettuare operazioni di mappatura e *query* complesse su dati espressi sulla base di modelli dati differenti.

Un registro di ontologie presenta infatti i seguenti possibili scenari di utilizzo, molti dei quali sono comuni ai registri di metadati, anche se i formalismi connessi alla rappresentazione ontologica ne potenziano gli effetti:

- individuare il modo migliore per descrivere e modellare una risorsa (una *query* al registro dovrà restituire una lista di classi e proprietà con indicazioni circa il loro utilizzo ed esempi concreti);
- armonizzare le ontologie usate in un determinato dominio da varie istituzioni o nell'ambito di vari progetti (il registro presenterà la descrizione di come le singole classi e proprietà sono state utilizzate per favorire operazioni di *crosswalk* tra ontologie);
- pubblicare in formato *linked open data* certi dati facendo massimo riuso di ontologie già esistenti (occorrerà effettuare *query* sul registro e visualizzare sul registro tutte le classi e le proprietà afferenti a un determinato concetto ricercato, contestualizzate in relazione alle caratteristiche del modello concettuale di riferimento e a casi d'uso);
- rendere comprensibile a lungo termine una ontologia sulla cui base sono stati modellati certi dati (il registro potrà conservare una copia in locale delle varie versioni delle ontologie, fungendo anche da repository e tenere conto delle variazioni intercorse nel tempo tra una versione e l'altra della medesima ontologia, definendo anche la compatibilità tra le versioni);
- individuare, per ogni ontologia esaminata, i riferimenti a chi la mantiene, a chi la aggiorna, a chi la usa e in quali progetti, ed esplicitare se i suoi autori hanno sviluppato ontologie in domini contermini (il registro terrà conto della frequenza di aggiornamento dell'ontologia, dei suoi autori, dei *dataset* che sono su di essa basati, degli autori che l'hanno modellata);
- verificare se una ontologia è stata usata in un determinato dominio, quale classe dell'ontologia è stata usata; come è definita una determinata classe in una determinata ontologia, da quale istituzione è stata utilizzata ed in che modo, a quali altre classi è eventualmente dichiarata equivalente e che relazioni ci sono tra una classe ed un'altra nell'ambito della medesima ontologia e/o con classi di altre ontologie, attraverso quale proprietà sono espresse.

Potenzialmente, il registro è in grado di garantire la semplificazione delle operazioni di mappatura e integrazione fra set di metadati, resa possibile da un'architettura in grado di estrapolare le entità comuni dai singoli vocabolari, esplicitata attraverso il ricorso a meccanismi di mappatura verso ontologie top level e l'effettuazione di operazioni di mappatura tra classi e proprietà e/o costrutti complessi di classi e proprietà da parte di utenti registrati e accreditati.

Fondamentali inoltre sono le funzionalità che prevedono la raccolta automatica dei requisiti funzionali alle politiche adottate per garantire non solo lo *storage* ma anche la disponibilità a lungo termine dell'ontologia pubblicata, sia riguardo al formato informatico dei file registrati, sia relativamente ai metadati necessari per la conservazione a lungo termine delle ontologie.

Registri di ontologie

Una prima disamina di registri di ontologie viene effettuata nel 2001 da Ding e Fensel che censiscono 8 registri di ontologie, quasi nessuno dei quali tuttavia risulta ad oggi mantenuto.¹⁹

D'Aquin e Noy nel 2012²⁰ forniscono un'analisi più aggiornata dei registri di ontologie, censendone ben 11, di cui solo 6 risultano ad oggi mantenuti: BioPortal, sviluppato per il settore biomedico dal National Center for Biomedical Ontology; Cupboard, creato nell'ambito del NeOn Project, si basa sul concetto di spazio in cui ciascun utente può selezionare e lavorare (annotare, allineare semanticamente e commentare) un *subset* specifico di ontologie; OBO Foundry initiative, che mira a creare degli insiemi di ontologie di ambito biomedico ben documentate e ben definite in grado di interrelarsi le une con le altre.; oeGov - Ontologies for e-governement, un'iniziativa per la creazione e lo storage di ontologie dedicate all'amministrazione digitale, mantenuta da TopQuadrant; OLS - Ontology Lookup Service, sviluppato dall'European Bioinformatics Institute, un registro usato in vari progetti di ambito biomedico; OntologyDesignPatterns.org (ODP) catalogo di *ontology design patterns*, che costituiscono di per sé delle micro-ontologie in alcuni casi o, in altri, dei componenti ontologici che possono essere facilmente integrabili in ontologie di dominio; OntoSelect, provvisto di un avanzato meccanismo di ricerca sulle ontologie che organizza i risultati sulla base di un meccanismo di ranking basato su parametri quali, ad esempio, il numero di importazioni o il numero di linguaggi; OntoSearch2, che intende fornire un efficiente meccanismo di query sulle ontologie, potenziando il linguaggio SPARQL; ONKI ontology server, a supporto di vari servizi informatici del governo finlandese; TONES repository, un registro di ontologie tutte rigorosamente espresse in OWL; Schema-Cache, Sviluppato sulla Talis Platform.

Nel 2016 Debashis Naskar e Biswanath Dutta identificano i seguenti 10 registri classificandoli a seconda delle loro funzionalità in *repository*, *directory* e *registry*,²¹ laddove attribuiscono agli *ontology registry* funzionalità di ricerca, navigazione, mapping e metadattazione; *le ontology directory* si caratterizzerebbero per essere relative ad un particolare dominio della conoscenza e per svolgere funzioni di reference ovvero di assistenza per gli utenti; non avrebbero tuttavia funzionalità di mapping o di metadattazione; gli *ontology repository* aggiungerebbero alle funzionalità presenti negli *ontology registry* anche funzionalità di *storage*, conservazione a lungo termine e di standardizzazione:

- 1) Bioportal, repository di dominio;
- 2) Agroportal, repository di dominio;
- 3) COLORE (COmmon Logic Ontology repository), repository generale;
- 4) Romulus (repository of Ontologies for MULTiple Uses), repository misto;
- 5) OeGov, directory di dominio;
- 6) ODP - OntologyDesignPatterns, directory generale;
- 7) ONKI, directory mista;

¹⁹ Ding, Y., Fensel, D. (2001). *Ontology library systems. The key to successful ontology reuse*. In *First Semantic Web Working Symposium*, 93–112.

²⁰ Noy, Natasha F., D'Aquin, Mathieu. *Where to Publish and Find Ontologies? A Survey of Ontology Libraries*, [op. cit.].

²¹ Cfr. Naskar, Debashis and Dutta, Biswanath, *Ontology Libraries: A Study from an Ontofier and an Ontologist Perspectives*, 2016, disponibile all'indirizzo https://www.researchgate.net/publication/305368094_Ontology_Libraries_A_Study_from_an_Ontofier_and_an_Ontologist_Perspectives.

- 8) DAML, directory mista;
- 9) MMI-ORR, registro di dominio;
- 10) Protégé ontology library, registro misto.

Tra i censimenti più recenti, va infine menzionata l'analisi di Jonquet, Toulet, Arnaud, et al. (2018)²² che riguarda tre registri del dominio "agricoltura":

- 1) FAIRSharing,²³ un database che raccoglie oltre 1200 standard sviluppati da comunità scientifiche e di sviluppatori organizzati a seconda del tipo di standard (Terminology Artifact, Model/Format, Reporting Guideline, Identifier Schema), del dominio coperto, del soggetto, delle tassonomie, del Paese di creazione e dell'organizzazione responsabile.
- 2) FAO's VEST Registry, non più disponibile perché confluito nel Gogan Vest / Agroportal map of standards.²⁴
- 3) agINFRA linked data vocabularies, punto di ingresso per i *value vocabularies* e *metadata vocabularies* utilizzati nel progetto agINFRA.²⁵

Allo stato attuale, le realizzazioni più interessanti sono rappresentate dai registri LOV – Linked Open Vocabularies, AgroPortal, Bioportal e Finto. Essi sono correntemente mantenuti e presentano le più avanzate funzionalità di gestione e presentazione delle ontologie (in termini di *storage*, conservazione, metadattazione, ricerca, annotazione etc.), con interfacce utente particolarmente amichevoli.

I quattro registri esaminati, in particolare, sono conformi al modello FAIR nel senso che consentono di rendere al loro interno le risorse (le ontologie) *findable*, *accessible*, *interoperable* e *reusable* conformemente ai 15 principi FAIR.²⁶ Inoltre essi si caratterizzano per la volontà di gestire in maniera uniforme i metadati delle ontologie registrate, secondo un preciso modello concettuale che è posto a loro fondamento, a differenza di altre realizzazioni che si limitano a trattare le ontologie sulla base dei soli metadati in esse esplicitamente previsti e sulla base di metriche automaticamente. In particolare, LOV – che è il portale di gran lunga più consultato da chi si occupa di ontologie – ha un orizzonte poco specialistico e mira a ricomprendere qualunque ontologia realizzata nei vari domini della conoscenza, mentre Bioportal e Agroportal si specializzano rispettivamente nel settore biomedico e agronomico. Dall'analisi delle loro funzionalità e dai test sul loro utilizzo è emerso chiaramente che strumenti generalisti come LOV risultano efficaci solo nel caso in cui un utente miri ad ottenere una prima panoramica sulle ontologie esistenti e ad identificare le ontologie afferenti al dominio di proprio interesse.

Nella restituzione dei risultati di ricerca l'analisi ha evidenziato come la metodologia di calcolo del *ranking* attribuito a ciascuna classe/proprietà dell'ontologia presti – soprattutto in LOV – una attenzione nettamente maggiore al concetto di "diffusione" dell'utilizzo di determinate ontologie (diffusione calcolata sull'uso delle classi e delle proprietà nella metadattazione dei dati censiti nella *lod*-

²² Jonquet, C., Toulet, A., Arnaud, E., Aubin, S., Dzale Yeumo, W. E., Emonet, V., Graybeal, J., Laporte, M.-A., Musen, M. A., Pesce, V., Larmande, P. (2018), "AgroPortal: A vocabulary and ontology repository for agronomy", *Computers and Electronics in Agriculture*, 144, 126–143, DOI: [10.1016/j.compag.2017.10.012](https://doi.org/10.1016/j.compag.2017.10.012), disponibile all'indirizzo <https://prodinra.inra.fr/record/427822>.

²³ Cfr. <https://fairsharing.org/>.

²⁴ Cfr. <https://www.godan.info/datasets/vest-agroportal-map-standards>.

²⁵ Cfr. <https://aginfra.eu/>.

²⁶ Cfr. <https://www.nature.com/articles/sdata201618>.

cloud) a discapito del concetto di “provenienza”, ovvero di collocazione in un determinato contesto – scientifico o meno – dell’ontologia medesima.

In riferimento alle caratteristiche dei registri, si riportano di seguito (fig. 1) in forma schematica alcuni risultati del processo di valutazione dei 4 registri analizzati in dettaglio.

	Dominio	Metadati ontologie	Formato dei file	Id. ontologie	Inserimento	Navigazione e ricerca di ontologie	Navigazione e ricerca di classi e proprietà	Mapping e relazioni	Accesso web	Uso	Versioning ontologie	Editing ontologie	Valutaz. ontologie
LOV	Non specialistico	VOAF	OWL, RDF	URI	Amministratori su segnalazione	Navigazione; Navigazione Gerarchica; Ricerca avanzata	Ricerca semplice e avanzata	Mapping tra classi, solo se presente nelle ontologie. Scarsamente implementato	REST	View, Download	Gestito	Non gestito	Non gestito
Bioportal	Biomedico	OMV	OWL, RDF, OBO	URI	Utenti registrati	Navigazione; Navigazione Gerarchica; Ricerca avanzata	Ricerca semplice e avanzata	Mapping tra classi	REST	Upload, Edit, View, Download, User evaluation	Gestito	Non gestito	Non gestito
Agroportal	Agronomico	MOD	OWL, RDF, OBO, UMLS	URI	Utenti registrati	Navigazione; Navigazione Gerarchica; Ricerca avanzata	Ricerca semplice e avanzata	Mapping tra ontologie	REST	Upload, Edit, View, Download, User evaluation	Gestito	Non gestito	Utenti registrati
Finto	E-government, Non specialistico	DC	RDF, SKOS	URI	Amministratori	Navigazione	Ricerca semplice	Non supportato	REST	Upload, View, Download	Gestito	Non gestito	Non gestito

Fig. 1. Risultati del processo di valutazione di 4 registri di ontologie

Alcune annotazioni finali e sviluppi possibili

Come ogni risorsa, anche le ontologie, i *thesauri* e i vocabolari controllati devono essere descritti con metadati pertinenti e adatti a facilitarne l’identificazione, la selezione, il riutilizzo e la conservazione a lungo termine. I registri di ontologie sono strumenti che descrivono formalmente i modelli ontologici disponibili sul *web*, a fronte di modelli di metadattazione il più possibile uniformi, e ne agevolano il reperimento e la valutazione, incentivandone il riuso e facilitando i processi di allineamento semantico e di interoperabilità. Un registro di ontologie per i beni culturali dovrebbe essere ideato per travalicare i limiti di un mero catalogo di risorse formali da prendere a riferimento per operare le scelte più opportune nella valorizzazione semantica dei propri dati. Esso dovrebbe anzitutto costituire un primo nucleo di quello che in prospettiva potrebbe diventare un repertorio ufficiale mantenuto, ad esempio, o dal MiBACT o da AgID, nell’ambito delle proprie funzioni istituzionali di promozione dell’omogeneità dei linguaggi, delle procedure e degli standard, connesse in particolare alle politiche di valorizzazione del patrimonio informativo pubblico nazionale, ivi compresa la definizione della strategia in materia di dati aperti nonché lo sviluppo e la gestione del portale nazionale dei dati aperti. Analogamente a quanto già avviene nell’ambito della conservazione digitale infatti, AgID inizia a porsi

come garante di strumenti che raccolgono tecnologie e teorie formali di tipo semantico certificate o accreditate mediante un sistema di valutazione e certificazione condiviso.²⁷

Nell'ambito dei beni culturali, un possibile ulteriore obiettivo è che un registro di dominio possa costituire il fondamento semantico di una *digital library* trasversale ai vari domini della cultura (dominio biblioteconomico, archivistico e museale), divenendone una componente di servizi interni, al fine di consentire la mappatura delle varie ontologie utilizzate nei sistemi di origine, che interagiscono con la *digital library*, e per ottimizzare le prestazioni delle interrogazioni. In tale veste, il registro garantirebbe una architettura in grado di estrapolare le entità comuni dai singoli modelli ontologici e rendere più semplici le mappature e le integrazioni fra *set* di metadati. Questa finalità riguarda la possibilità di usare il registro per facilitare la costruzione di indici comuni a risorse metadate sulla base di ontologie differenti per la ricerca delle risorse culturali e l'accesso ad esse. La disponibilità di un simile strumento consentirebbe di superare una delle maggiori difficoltà ravvisate nei tentativi anche recenti di costruzione di *digital library cross-domain*. Infatti le risorse relative a ciascun dominio potrebbero continuare ad essere pubblicate sulla base dei modelli concettuali e dei sistemi di metadatezione propri di ciascun ambito disciplinare (archivistico, librario, archeologico, demotnoantropologico, storico-artistico etc.) mentre la *reductio ad unum* sarebbe demandata solo alla fase di ricerca nella *digital library*. Si supererebbe in questo modo uno dei principali limiti degli attuali portali aggregatori di risorse come CulturaItalia o Europeana: per inviare i dati all'aggregatore è necessario effettuare operazioni di mappatura semantica a partire dai sistemi originari sulla base del tracciato dati individuato per l'aggregatore. Tale tracciato spesso coincide con un tracciato di minima scelto come modello dati comune alle varie tipologie di risorse gestite. Le operazioni di mappatura hanno pertanto comportato operazioni di schiacciamento semantico dei dati che in origine si presentavano molto ricchi (si pensi ai dati espressi secondo i tracciati descrittivi all'interno del Catalogo generale dei beni culturali) o di forzature semantiche di dati espressi secondo modalità consolidate di metadatezione peculiari di un determinato dominio (si pensi ai dati archivistici). Una *digital library* basata sulle tecnologie del web semantico dovrebbe invece offrire la possibilità di effettuare ricerche sul complesso degli elementi descrittivi di ciascuna risorsa, indicizzati ciascuno secondo il proprio modello descrittivo; inoltre, grazie all'utilizzo di un registro di ontologie in cui sia reso esplicito (e computabile dai software) lo strato di mappatura concettuale tra i modelli descrittivi dei vari domini, sarebbe possibile anche ricercare su chiavi comuni, nel senso di entità espresse in maniera definita con un procedimento che raccoglie e aggrega aspetti particolari che una molteplicità di oggetti informativi hanno in comune.

²⁷ Cfr. OntoPiA, <https://github.com/italia/daf-ontologie-vocabolari-controllati>.

Riferimenti bibliografici

Allocca, Carlo, Mathieu D'Aquin, Enrico Motta. 2009. *DOOR-towards a formalization of ontology relations*. In *International conference on knowledge engineering and ontology development*, keod'09, 13–20. Madera, Portugal.

Allocca, Carlo, Mathieu D'Aquin, Enrico Motta. 2012. "Impact of Using Relationships between Ontologies to Enhance the Ontology Search Results." In *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 164–176. Springer Berlin Heidelberg.

D'Aquin, Mathieu, Holger Lewen. 2009. "Cupboard - a place to expose your ontologies to applications and the community." In *Lecture Notes in Computer Science*, volume 5554, 913–918. Berlin: Springer.

Debashis, Naskar. 2014. *Ontology and ontology libraries: a critical study*, in *Master's Dissertation* (carried under the supervision of Biswanath Dutta), Bangalore, India: DRTC, Indian Statistical Institute, 10–49.

Ding, Ying, Fensel, Dieter. 2001. "Ontology Library Systems: The key for Successful Ontology Reuse." In Isabel F. Cruz, Stefan Decker, Jerome Euzenat, Deborah L. McGuinness, *Proceedings of SWWS'01, The First Semantic Web Working Symposium*, 93–112, <http://sw-portal.deri.org/papers/publications/ding+01.pdf> ultima consultazione 30/12/2019.

Hartmann, Jens, Raúl Palma, Asunción Gómez-Pérez. 2009. "Ontology Repositories", *Handbook on Ontologies*, 906–915. Berlino, Heidelberg: Springer-Verlag.

Jonquet, Clement, *et al.* 2018. "AgroPortal: A vocabulary and ontology repository for agronomy." *Computers and Electronics in Agriculture*, 144:126–143. <https://prodinra.inra.fr/record/427822>.

Jonquet, Clement, Anne Toulet, Biswanath Dutta, Vincent Emonet. 2018. "Harnessing the Power of Unified Metadata in an Ontology Repository: The Case of AgroPortal." *Journal on Data Semantics* 7:191–221, <https://doi.org/10.1007/s13740-018-0091-5>.

Magkanaraki, Aimilia, Sofia Alexaki, Vassilis Christophides, Dimitris Plexousakis. 2002. "Benchmarking RDF Schemas for the Semantic Web", *Lecture Notes in Computer Science*, 2342, Berlin, Heidelberg: Springer.

Mccarthy, John W., *et al.* 2009. *Data Modeling and Harmonization with OWL: Opportunities and Lessons Learned*. http://ceur-ws.org/Vol-524/swese2009_7.pdf, ultima consultazione 30/12/2019.

Noy, Natalya F., *et al.* 2009. "Bioportal: ontologies and integrated data resources at the click of a mouse." *Nucleic Acids Research*, vol. 37.

Noy, Natalya F., Mathieu D'Aquin. 2012. "Where to Publish and Find Ontologies? A Survey of Ontology Libraries, in Web Semantics: Science, Services and Agents on the World Wide Web." *Journal of Web Semantics*, vol. 11, January, <http://www.websemanticsjournal.org/index.php/ps/article/view/217>, ultima consultazione 30/12/2019.

Stadtmuller, Steffen, Andreas Harth, Marko Grobelnik. 2013. "Accessing Information About Linked Data Vocabularies with vocab.cc." In Li J., Qi G., Zhao D., Nejd W., Zheng HT. (eds). *Semantic Web and Web Science. Springer Proceedings in Complexity*, New York: Springer.

Thomas, Edward, Jeff Z. Pan, Jeff, Derek Sleeman. 2007. "ONTOSEARCH2: Searching ontologies semantically." In *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions, Innsbruck, Austria, June 6-7, 2007*, ceur-ws.org.

Xiang, Zuoshuang, Mélanie Courtot, Ryan R. Brinkman, Alan Ruttenberg, Yongqun He. 2010. "OntoFox: web-based support for ontology reuse." *BMC Research Notes*, vol. 3.