

Dal Catalogo generale dei beni culturali al *knowledge graph* del patrimonio culturale italiano: il progetto ArCo

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00013

Chiara Veninata

ICCD

Le attività dell'ICCD sono da sempre indirizzate ad una maggiore condivisione e valorizzazione sia dei modelli di strutturazione della conoscenza sul patrimonio culturale sia dei dati prodotti nelle campagne di catalogazione. Negli ultimi anni l'ICCD ha concentrato le proprie attività sull'analisi e sull'applicazione delle potenzialità offerte dal semantic web e dai suoi strumenti. Uno dei risultati è il progetto ArCo, il grafo della conoscenza del patrimonio culturale italiano, costituito da una rete di ontologie e da oltre 169 milioni di triple riferite a oltre a 800 mila schede catalografiche. ArCo si basa sui dati del Catalogo generale dei beni culturali dell'Istituto centrale per il catalogo e la documentazione del MiBACT e sui dati dei suoi archivi fotografici. ArCo è distribuito congiuntamente con uno SPARQL endpoint, un software per convertire i record di catalogo in RDF e una ricca suite di materiale di documentazione (test, valutazione, istruzioni, esempi ecc.).

Premessa

L'Istituto centrale per il catalogo e la documentazione (ICCD) gestisce il Catalogo del patrimonio archeologico, architettonico, paesaggistico, storico artistico e demoetnoantropologico, realizzato ai sensi dell'art. 17 del Codice dei beni culturali. Il Catalogo nasce in Italia alla fine dell'Ottocento, quando il nascente Stato italiano dà avvio, nel 1875, alle prime rilevazioni sistematiche sul territorio con la finalità di censire il patrimonio storico artistico del nuovo stato italiano¹.

Il termine catalogazione fa generalmente riferimento a qualsiasi intervento volto a classificare, sulla base di caratteristiche comparabili, determinati oggetti. Esso

¹ Con la prima legge di tutela del 1902 si comincia a parlare di "Catalogo unico dei monumenti ed oggetti aventi pregio d'arte o di antichità". Ogni riferimento al Catalogo scompare nella seconda legge di tutela n. 364 del 20 giugno 1909, per ricomparire invece nel r.d. n. 1889 del 1923 in cui si richiama la necessità di compilare, per finalità sia di conoscenza che di conservazione, un catalogo dei monumenti e delle opere d'interesse storico, artistico e archeologico di proprietà statale. Anche se nulla viene normato sulla catalogazione nella legge n. 1089 del 1939, le soprintendenze continuano a catalogare il patrimonio: le schede così prodotte confluiscono nel 1969 nell'Archivio del nuovo Ufficio Centrale per il Catalogo, istituito quell'anno presso il Ministero della Pubblica Istruzione, trasformatosi in Istituto Centrale per il Catalogo e la Documentazione (ICCD) nel 1975, a seguito dell'istituzione del Ministero dei Beni Culturali.

comprende una serie di attività riconducibili alla registrazione, elencazione ed enumerazione organica di attributi di un oggetto e tuttavia non si esaurisce in essa. Il sistema della catalogazione del patrimonio culturale è infatti da sempre basato su due presupposti, l'uno rivolto alle specificità dell'oggetto di indagine, l'altro ad un recupero integrato delle informazioni pertinenti all'oggetto volto a garantirne la comunicazione nella dinamica spazio-tempo-contesto che lo caratterizza e che lo rende parte del patrimonio culturale.

Per garantire coerenza a tali presupposti, l'ICCD ha elaborato nel tempo un articolato "sistema di conoscenza": principi di metodo, strumenti e regole per acquisire le informazioni sui beni secondo procedure e criteri omogenei in tutto il territorio nazionale. La componente fondamentale di questo sistema è rappresentata dalle schede di catalogo, modelli descrittivi costituiti da una sequenza predefinita di voci, che raccolgono in modo formalizzato le notizie sui beni, seguendo un percorso conoscitivo che guida il catalogatore e al tempo stesso controlla e codifica i dati sulla base di precisi parametri e vocabolari. In relazione ai vari tipi di beni - mobili, immobili e immateriali - l'ICCD ha elaborato, ad oggi, trenta diverse schede di catalogo riferite a nove settori disciplinari (beni archeologici, beni architettonici e paesaggistici, beni demoetnoantropologici, beni fotografici, beni musicali, beni naturalistici, beni numismatici, beni scientifici e tecnologici, beni storici e artistici). Nel tracciato di ogni singola scheda si esplicitano informazioni identificative, descrittive, tecnico scientifiche, storico critiche, geografiche e di contesto, che consentono di porre in relazione il bene al territorio e agli altri beni secondo una prospettiva spazio temporale.

I dati del Catalogo generale del patrimonio culturale vengono governati attraverso il SIGECweb - Sistema informativo generale del catalogo - che gestisce circa tre milioni di schede di Catalogo che, pur nella loro numerosità, costituiscono una goccia nel mare del patrimonio culturale italiano e non consentono di realizzare, da sole, una attendibile ed esaustiva "mappa del patrimonio culturale"². Dal 2001 - con la riforma costituzionale che affida anche alle Regioni la valorizzazione del patrimonio culturale - le Regioni partecipano al processo di catalogazione del patrimonio culturale. È un sistema policentrico in cui si è dato vita a sistemi informativi regionali senza che, tuttavia, si sia mai avviato un organico e aggiornabile processo di comunicazione con il sistema informativo generale del catalogo gestito dall'ICCD³.

² Cfr. Laura Moro, *Il Catalogo del patrimonio culturale italiano. Nuova centralità e prospettive future*, «Economia della Cultura», 25 (2015), n. 3-4, p. 419.

³ Seguendo l'articolato sistema di relazione tra Stato e Regioni, a partire dagli anni Settanta, si arriva agli articoli 148 e 149 del D. Lgs. 112 del 1998 che definisce i rispettivi compiti nell'ottica delle distinte funzioni di tutela e valorizzazione dei beni culturali, distinzione poi confermata dalla riforma del Titolo V della costituzione (legge costituzionale 18 ottobre 2001, n. 3). L'Accordo nazionale Stato-Regioni del 2001 consolida un sistema policentrico in cui le Regioni avviano i loro sistemi informativi con riferimento alle necessità del territorio, in connessione con il sistema informativo generale del catalogo gestito, per il Ministero, dall'ICCD. Nel 2004, il Codice dei beni culturali all'art.

I sistemi, centrali e locali faticano ancora a scambiarsi i dati e la mancata interoperabilità tra di essi determina l'assenza di un Catalogo nazionale dei beni culturali, completo e aggiornato, ma solo la presenza delle varie e difformi "sue articolazioni"⁴.

Le possibilità offerte dalle tecnologie del web semantico

Con l'obiettivo di esplorare strade più innovative per lo scambio di dati tra sistemi informativi e per il loro arricchimento reciproco, l'ICCD ha di recente volto la sua attenzione alle possibilità offerte dal cosiddetto web semantico. In particolare la diffusione di modelli ontologici⁵ e di linked open data (LOD) sembrano offrire uno scenario in cui potrebbero trovare soluzione alcuni dei problemi suesposti. Si fa riferimento, in particolare, all'utilizzo di strumenti in grado di abilitare la descrizione, la pubblicazione e la condivisione dei dati attraverso database aperti e distribuiti (*triplestore*) in cui il singolo dato viene immediatamente reso disponibile sul web dall'istituzione responsabile, connotato semanticamente attraverso il ricorso a metadati espressi secondo formalismi (ontologie) comprensibili anche dai motori di ricerca più evoluti e collegabile facilmente ad altri dati attraverso relazioni significative, perché anch'esse riferite a modelli concettuali comprensibili dai software⁶. A livello internazionale, a partire dal 2010, si è manifestata una certa tendenza a promuovere i linked open data come una pratica standard di produzione e pubblicazione dei dati sul web, in linea con le raccomandazioni contenute nel *Library Linked data Incubator Group Final Report* rilasciato alla fine del 2011 dal *Library Linked data Incubator Group* del W3C⁷ e con gli indirizzi di Europeana⁸ e della

17 recepisce questo processo partecipativo delle regioni ma continua a fare perno sul Ministero, come garante dell'unitarietà della tutela, per assicurare la catalogazione del patrimonio culturale e definirne le metodologie con il concorso delle Regioni. Il quadro che ne emergeva era che Stato, Regioni ed enti pubblici territoriali catalogassero ciascuno i beni di loro proprietà e facessero affluire i dati prodotti al Catalogo nazionale, di fatto mai realizzato.

⁴ Cfr. art. 17 D. Lgs. 42/2004.

⁵ Le ontologie sono la formulazione di schemi concettuali esaustivi e rigorosi nell'ambito di un dominio dato. Tali schemi concettuali devono essere condivisi il più possibile nell'ambito della propria comunità di interesse e devono essere formalizzati secondo un linguaggio comprensibile anche alle macchine, che dovranno elaborare automaticamente quei dati consentendo di effettuare su di essi ragionamenti e query anche complesse.

⁶ Cfr. Commissione di coordinamento SPC, *Linee guida per l'interoperabilità semantica attraverso i linked open data*, 2012, disponibile all'indirizzo: <https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.

⁷ Disponibile all'indirizzo: <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>. Il *Final Report* forniva un'analisi dettagliata dei benefici derivanti dall'utilizzo delle tecnologie del *semantic web* nel settore dei beni culturali, sia per ciò che concerne gli utenti e i ricercatori sia per quanto riguarda gli stessi istituti culturali nei termini di una maggiore diffusione, comprensione e valorizzazione del loro immenso patrimonio informativo pubblicato sul "nuovo web".

⁸ Cfr. Harry Verwayen - Martijn Arnoldus - Peter B. Kaufman, *The Problem of the Yellow Milkmaid. A Business Model Perspective on Open Metadata*, Europea White Paper N. 2, novembre 2011. La

stessa direttiva 2013/37/UE sulla PSI (Public Sector Information), recepita anche in Italia nel 2015⁹. Negli ultimi dieci anni si stanno compiendo rapidi passi nella direzione della pubblicazione in formato linked open data dei dati relativi alle collezioni, soprattutto in ambito biblioteconomico e museale o per specifici progetti che riguardano il *cultural heritage* in maniera trasversale. Tali esperienze hanno determinato una crescente consapevolezza sulle potenzialità che i linked open data offrono alle istituzioni culturali per “aprire” i propri dati e pubblicarli secondo modalità innovative, che consentono l’arricchimento reciproco delle informazioni di partenza collegate ad altre fonti pubblicate secondo i medesimi paradigmi¹⁰. In questa direzione, oggi si inizia a parlare, anziché di “*cataloguing*”, di “*catalinking*”, secondo la felice espressione di Roy Tennant¹¹. Per l’ICCD, individuato l’obiettivo, si è trattato di comprendere come procedere alla pubblicazione dei linked open data del Catalogo senza sacrificare nulla della capacità, evoluta nel tempo, di esprimere complesse architetture della conoscenza sul patrimonio culturale.

ricerca focalizzava l’attenzione su costi/benefici dell’utilizzo dei linked data nel settore delle gallerie, biblioteche, archivi e musei (GLAMS). Venivano affrontati punti cruciali legati al rilascio come open data dei beni culturali, ovvero la licenza d’uso e la perdita del controllo delle risorse “liberate”: un nuovo modo di valorizzare le risorse culturali potrebbero essere costituito dallo sviluppo di servizi innovativi legati alle risorse culturali, applicazioni e servizi a favore dei ricercatori.

⁹ La direttiva 2013/37/UE interviene in materia di riutilizzo dell’informazione nel settore pubblico, attraverso la modifica della direttiva 2003/98/CE, la c.d. direttiva PSI (Public Sector Information), ed è finalizzata a favorire il riutilizzo dei dati delle pubbliche amministrazioni dell’Unione europea estendendo l’ambito di applicazione anche alle istituzioni culturali (biblioteche, musei e archivi) in precedenza escluse. In Italia è stata recepita con Decreto legislativo, 18/05/2015 n° 102 pubblicato in G.U. il 10/07/2015.

¹⁰ Una prima analisi dei progetti c.d. LOD-LAM (Linked Open Data for Libraries, Archives and Museums) è stata condotta con la *Review on linked open data sources*, effettuata nell’ambito del Progetto Athena Plus dell’ottobre 2013; essa è stata successivamente integrata e aggiornata da una ricerca di OCLC condotta tra luglio e agosto 2014 e poi ripubblicata con alcune correzioni nel 2015. I risultati della prima indagine del 2014, portata a termine nel 2017, sono stati successivamente confrontati con la revisione 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. Le istituzioni italiane coinvolte nell’indagine OCLC del 2018 erano 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. Le attività di ricerca condotte dalla scrivente nel 2018 arricchiscono notevolmente l’indagine sul fronte italiano. Cfr. Chiara Veninata, *Linked open data e ontologie per la descrizione del patrimonio culturale: criteri per la progettazione di un registro ragionato*, disponibile all’indirizzo <<https://iris.uniroma1.it/handle/11573/1325813#.XmuFM6hKg2w>>.

¹¹ Cfr. su Twitter lo storify dell’intervento *From linked data to shared knowledge* di Roy Tennant, Jon Voss e Ingrid Mason, alla conferenza di Information Online 2013 dal titolo “*Be different. Do differently*”. Information Online 2013 disponibile all’indirizzo <<https://storify.com/ALIAonline/keynote-roy-tennant-jon-voss-and-ingrid-mason>>.

La scelta del modello ontologico per il knowledge graph del patrimonio culturale

L’Agenzia per l’Italia Digitale (AGID) mira ad individuare alcuni ecosistemi digitali¹², in ciascuno dei quali una rete di ontologie per la pubblica amministrazione definita OntoPiA garantirebbe integrazione tra i dati prodotti dai soggetti che contribuiscono all’ecosistema. In questa prospettiva, l’ICCD ha inteso aprire i dati del Catalogo generale pensando ad essi come un tassello fondamentale dell’ecosistema digitale “Beni culturali e turismo”. Una volta scelto il paradigma dei linked open data, il Catalogo può assumere un nuovo ruolo come fondamento del *knowledge graph* sul patrimonio culturale, cui gli altri detentori di informazioni (catalografiche e non) sul patrimonio sono liberi di contribuire, in una architettura della conoscenza aperta e condivisa.

Al fine di garantire una ottimale pubblicazione delle informazioni contenute nelle varie tipologie di schede catalografiche, a novembre 2017 è stato avviato il progetto ArCo, con cui l’ICCD e il Semantic Technology Laboratory (STLab) dell’Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR¹³ mirano a valorizzare il patrimonio artistico e culturale italiano attraverso l’individuazione (o la creazione, come vedremo) di una rete di ontologie che fornisca un’architettura della conoscenza dei beni culturali e consenta la pubblicazione di dati secondo il paradigma linked open data e che sia il più possibile integrata con OntoPiA.

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. Ciò agevola l’interoperabilità semantica tra diverse basi di conoscenza.

Nell’individuare il modello ontologico più adeguato a render conto della complessità della metodologia catalografica e della ricchezza dei dati presenti nelle schede, si sono analizzate alcune ontologie del dominio¹⁴ dei beni culturali ed in particolare CIDOC-CRM, EDM, Cultural-ON, Dublin Core, DBPedia e Schema.org.

¹² Cfr.: <https://docs.italia.it/italia/piano-triennale-ict/pianotriennale-ict-doc/it/2019-2021/07_ecosistemi.html>.

¹³ Il team del CNR che lavora sul progetto ArCo è composto da Aldo Gangemi, Valentina Presutti, Luigi Asprino, Valentina Corriero, Andrea Nuzzolese, Margherita Porena, Alessandro Russo e Ludovica Marinucci. Per ICCD, responsabili scientifici del progetto sono Chiara Veninata e Maria Letizia Mancinelli.

¹⁴ La locuzione “di dominio” fa riferimento alla seguente distinzione a seconda delle finalità e dei confini tematici delle ontologie:

1) Upper-level ontologies (dette anche ontologie pesanti o ontologie fondazionali). Un’ontologia di alto livello rappresenta il tentativo di modellare l’intera realtà in un’unica tassonomia, da cui le specifiche ontologie ereditano al fine di classificare i diversi ambiti del sapere. Hanno un livello di astrazione molto elevato e sono spesso frutto di competenze multidisciplinari che coniugano le tecnologie dell’informazione con la filosofia analitica e l’epistemologia. Modellano i concetti fondazionali comuni a tutti i domini (spazio, tempo ecc.) e perciò sono adatte a scambiare informazioni attraverso domini diversi, dove i problemi di ambiguità terminologica acquistano fondamentale rilevanza. Le ontologie fondazionali sono sviluppate proprio allo scopo di facilitare la negoziazione dei significa-

Nell'ambito dei beni culturali, CIDOC-CRM è un'ontologia creata per la modellazione, l'integrazione e lo scambio di informazioni eterogenee riguardanti i beni culturali ed è stata ottenuta da un lavoro di oltre dieci anni da parte del CIDOC Documentation Standards Working Group¹⁵. CIDOC-CRM si pone in realtà a cavallo tra le ontologie c.d. fondazionali e le ontologie di dominio. L'ontologia è incentrata sulle relazioni tra l'oggetto culturale fisico e le relazioni tra esso e gli eventi che lo hanno visto coinvolto, nonché con le entità che hanno interagito con esso (persone, organizzazioni). CIDOC-CRM ha un approccio fortemente evento-centrico, focalizzandosi soprattutto sull'aspetto temporale delle attività e degli eventi, messi in relazione con la partecipazione di attori, luoghi, date e oggetti in generale. Ogni evento può essere concepito come un incontro tra risorse, in cui specifiche proprietà definiscono il tipo di partecipazione ad esso.

CIDOC-CRM non dà nessuna indicazione circa l'obbligatorietà di certe classi o proprietà, lasciando massima libertà nella modellazione dei dati, in modo da potersi adattare a qualunque requisito espressivo richiesto nei vari ambiti culturali per differenti oggetti culturali. Un'altra caratteristica di CIDOC-CRM è la struttura poli-gerarchica delle classi (e delle proprietà) che consente una grande libertà ed espressività nella modellazione dei dati. Il fatto che la questione terminologica venga demandata alla possibilità di utilizzare propri vocabolari, integrandoli con il modello ontologico, consente un ulteriore arricchimento semantico dei dati, senza sacrificare l'unicità del modello ontologico. I vocabolari stessi possono essere inoltre collegati tra loro per esplicitare dipendenze, corrispondenze e relazioni tra i termini in generale. Ad oggi CIDOC-CRM risulta una delle ontologie più utilizzate in ambito museale¹⁶.

ti dei termini di un certo dominio attraverso diverse comunità scientifiche o linguistiche.

2) Ontologie di dominio (*domain ontologies*). Si tratta di schemi concettuali che rappresentano la conoscenza di un dominio applicativo più o meno specifico, come possono essere gli archivi, le biblioteche, le opere d'arte, i beni archeologici ecc. Le ontologie utilizzate nei sistemi informativi sono di solito ontologie di dominio, ma di fatto i sistemi informativi basati su queste ontologie funzionano bene fintanto che le persone e le macchine condividono lo stesso vocabolario e il significato dei termini, come accade in comunità chiuse. Questo tipo di ontologie risulta essere debole relativamente a possibili fraintendimenti semantici quando gli utilizzatori del sistema informativo non condividono le stesse assunzioni sul significato dei termini del vocabolario. Per essere in grado di facilitare la negoziazione di significati, è estremamente utile che le ontologie di dominio contengano riferimenti semantici alle ontologie fondazionali, che possano chiarire i termini più specifici usati nelle ontologie di dominio.

¹⁵ A partire dal dicembre 2006 il modello concettuale è diventato uno standard ISO, e intende diventare il riferimento per la rappresentazione semantica di informazioni relative al patrimonio culturale. Nel 2014 l'ontologia è stata tecnicamente rivisitata a cura della commissione competente ISO / TC 46, "Informazione e documentazione", sottocomitato SC 4, "interoperabilità tecnica", in collaborazione con il Consiglio Internazionale dei Musei - Comitato per la Documentazione (ICOM CIDOC).

¹⁶ Nel tempo sono stati creati ulteriori CRM; nati come estensioni di CIDOC: FRBRoo, Bibliographic References, PRESSoo: Periodicals, CRMInf: Argumentation Model, CRMsci: Scientific Observation Model, CRMdig: Digital Provenance Model, CRMgeo: Spatiotemporal Refinement. È stato da poco rilasciato anche il profilo CRMarcheo a cura del team del Progetto Ariadne.

L'Europeana Data Model (EDM) è l'evoluzione – in chiave ontologica – dell'Europeana Semantic Elements (ESE), il modello di dati con cui nel portale Europeana viene gestita l'importazione dei dati relativi alle descrizioni di oggetti del patrimonio culturale. Tale ontologia fa riuso diretto delle ontologie Dublin Core, DC Terms, OAI-ORE e stabilisce relazioni semantiche con CIDOC, Dolce-Lite, Abc, DCTems.

Cultural-ON è l'ontologia sviluppata nel 2015 dal MIBACT, insieme all'Istituto di Scienze e Tecnologie della Cognizione del Consiglio Nazionale delle Ricerche (ISTC-CNR) per la rappresentazione secondo il paradigma dei linked open data delle informazioni relative agli Istituti e luoghi della cultura e agli eventi culturali. Si articola in 56 classi e 90 proprietà e stabilisce relazioni semantiche con le ontologie Dolce, CIDOC, ORG, Event, Ticket ontology, LOD, FOAF, Fabio, Cito, Geonames, GoodRelation, PRO, SAN e con un set di ontologie di ISTAT. Di recente è stata rifattorizzata¹⁷ per integrarla nella rete di OntoPiA.

Dublin Core terms e refinements (DC e DCTerm) costituiscono dei data model leggeri che hanno lo scopo di descrivere attraverso un set minimo di metadati, una grande varietà di risorse in formati diversi. Dublin Core è estremamente diffuso a livello internazionale e il suo utilizzo è considerato un ottimo punto di partenza per l'interoperabilità tra sistemi diversi. Il nucleo (DC) è costituito da 15 proprietà cui è possibile affiancare dei "raffinamenti" (DCT) che consentono una descrizione delle risorse ancora più granulare¹⁸.

DBPedia ontology è una ontologia leggera e cross-domain, creata sulla base degli infobox più comunemente utilizzati all'interno di Wikipedia. L'ontologia mira a rappresentare i dati di DBPedia, che nasce per estrarre contenuti strutturati dalle informazioni create in vari progetti Wikipedia. Questa informazione strutturata assomiglia a un *open knowledge graph* (OKG), ovvero un tipo di database aperto che memorizza le conoscenze in una forma leggibile dalla macchina e fornisce un mezzo per raccogliere, organizzare, condividere, cercare e utilizzare dati e informazioni.

Schema.org è un vocabolario nato nel 2011 come tentativo di descrivere sul web qualsiasi tipologia di risorsa adoperando un linguaggio condiviso, per facilitare le ricerche. Fondatori del progetto Schema.org sono stati Google, Microsoft, Yahoo e Yandex. Nonostante la sua diffusione sia prevalente sui siti di e-commerce, da diver-

¹⁷ In ingegneria del software, il *refactoring* è una tecnica per modificare la struttura interna di porzioni di codice senza modificarne il comportamento esterno, applicata per migliorare alcune caratteristiche non funzionali del software quali la leggibilità, la manutenibilità, la riusabilità, l'estensibilità del codice nonché la riduzione della sua complessità. Cfr Martin Fowler, *Refactoring*, <https://refactoring.com/>

¹⁸ Il documento relativo alla trasposizione in RDF dello standard è una specifica aggiornata e autorevole di tutti i metadati gestiti dalla Dublin Core Metadata Initiative (DCMI). Sono inclusi i quindici termini del Dublin Core Metadata Element Set, che sono stati anche pubblicati come IETF RFC 5013 [RFC5013], ANSI / NISO Standard Z39.85-2007 [NISOZ3985] e ISO Standard 15836: 2009 [ISO15836].

so tempo il suo utilizzo si sta affermando in ambito internazionale anche per la descrizione del patrimonio culturale per aumentarne la visibilità sul web. In seno al W3C è stato creato lo Schema Bib Extend Community Group, col compito di preparare proposte per estendere Schema.org per una migliore rappresentazione e condivisione delle informazioni bibliografiche. L'iniziativa è complementare al lavoro che la Library of Congress conduce su BIBFRAME, ed è finalizzata non a rappresentare la ricchezza presente nei dati MARC ma a fare in modo che i dati bibliografici possano essere marcati in un modo che i motori di ricerca possano comprenderli e quindi utilizzarli appropriatamente. Sebbene infatti Schema.org non sia stato progettato per sostituire gli standard delle biblioteche, l'ontologia definisce un *reasonably coherent commonsense model* con classi e proprietà rilevanti per la descrizione delle risorse bibliografiche tra cui lavoro creativo, persona, autore, regista, luogo, organizzazione, editore, data di copyright, libro, ISBN e così via. Schema.org è proposto come un'impalcatura concettuale di massima a cui possono essere collegate ontologie più dettagliate progettate dalle varie comunità degli specifici domini.

A lungo il riuso di una ontologia, raccomandato tra le buone pratiche nell'implementazione dei linked data, è stato inteso come riuso diretto, ovvero adozione di un'ontologia già sviluppata per la modellazione dei propri dati. Nella pratica si tratta di fare riferimento in una ontologia A ad un'ontologia B tramite richiamo della sua URI. L'ontologia B contiene definizioni il cui significato diviene così parte integrante del significato dell'ontologia A. Poiché la proprietà è transitiva, se l'ontologia B ha importato a sua volta una terza ontologia C, anche tale ontologia C viene ad essere parte integrante dell'ontologia A. Ciò implica che, per un riuso diretto consapevole, è bene conoscere tutte le catene di importazione cui l'ontologia importata è eventualmente vincolata.

Esiste anche una seconda via per fare riferimento a ontologie preesistenti e ritenute pertinenti al proprio dominio, ed è quella che prevede il riuso indiretto delle ontologie connettendo esplicitamente la propria ontologia ad altre ontologie significative attraverso meccanismi di allineamento semantico (*ontology mapping*). Questa modalità presenta un duplice vantaggio, ovvero quello di mantenere autonomo il proprio schema concettuale da eccessive dipendenze da schemi concettuali mantenuti da agenti terzi su cui non si ha alcun potere di controllo e rendere la propria ontologia totalmente autoconsistente nel modellare un determinato dominio, con indubbi vantaggi anche per gli eventuali riutilizzatori dell'ontologia creata.

Con questa consapevolezza e poiché nessuna delle ontologie esaminate riusciva ad esprimere la complessa semantica sottesa alle trenta normative catalografiche rilasciate da ICCD, il team di ricerca ICCD-CNR ha stabilito di procedere al disegno di una nuova ontologia, pubblicata con il nome di "ArCo - Architettura della Conoscenza". ArCo si presenta come una rete di ontologie rilasciate dall'ICCD insieme all'Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR tra il 2018 e il 2020 per la strutturazione della conoscenza per i beni culturali.

La rete di ontologie di ArCo – Architettura della Conoscenza

ArCo intende rispecchiare l'analiticità strutturale delle schede catalografiche elaborate dall'ICCD nonché valorizzare il contesto dei beni culturali e restituire la complessità dell'oggetto "bene culturale", rafforzando la ricchezza semantica dei dati tipicamente catalografici.

Le ontologie di ArCo si compongono di più moduli. Il modulo ArCo importa sia Cultural-ON che altre ontologie di OntoPiA. Gli altri moduli finora rilasciati riguardano la "context-description", il "cultural-event", la "denotative-description", il "catalogue record" e la "location".

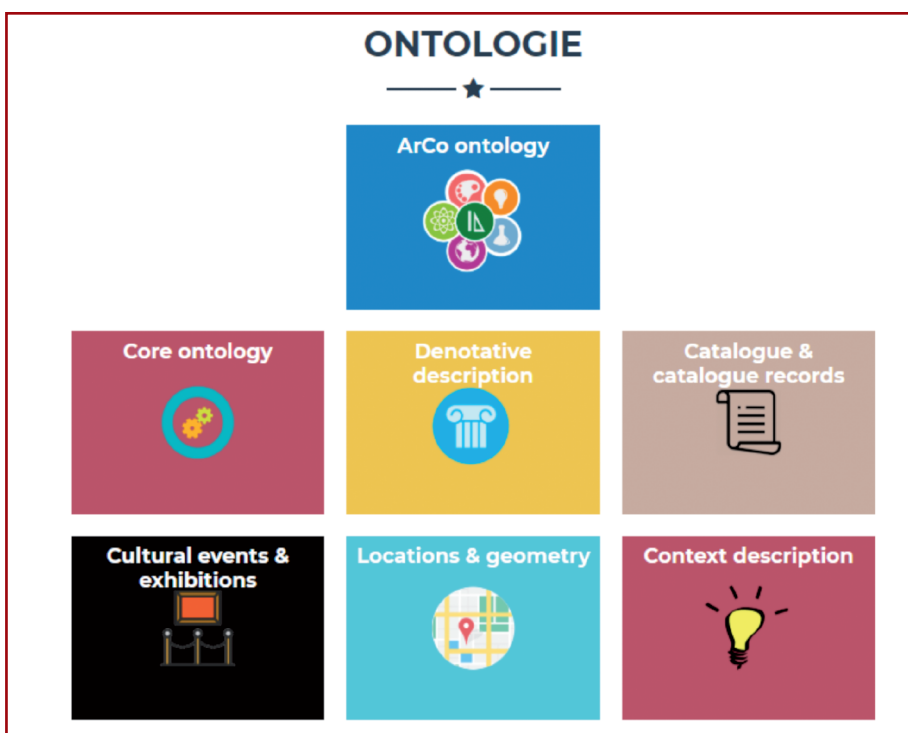


Figura 1. Le ontologie del progetto ArCo

ArCo viene rilasciato con licenza CC-BY-SA 4.0 sia su GitHub che sul sito ufficiale dei dati aperti del MiBACT. È accessibile online e può essere anche scaricato e installato localmente con un *docker*¹⁹. ArCo include:

- una rete di ontologie che modellano il dominio dei beni culturali con un alto livello di dettaglio;

¹⁹ *Docker* è un progetto *open-source* che consente di rilasciare una applicazione all'interno di un contenitore software fornendo anche una virtualizzazione del sistema operativo di Linux, evitando l'installazione e la manutenzione di una macchina virtuale. Cfr. <https://www.docker.com/>.

- un dataset di dati LOD che conta 169 milioni di triple, che descrivono oltre 800.000 beni culturali;
- un software per convertire automaticamente i record di catalogo conformi alle normative ICCD in LOD conformi ad ArCo, che consente aggiornamenti automatici e frequenti e facilita il riutilizzo delle ontologie;
- una documentazione dettagliata che riporta: (i) i requisiti ontologici espressi sotto forma di user story, le cosiddette domande di competenza, (ii) i modelli ontologici con diagrammi ed esempi di utilizzo;
- una serie di esempi che i potenziali utenti possono utilizzare come materiale di formazione: vi sono domande poste in linguaggio naturale e tradotte in corrispondenti query SPARQL, che possono essere testate direttamente con lo SPARQL endpoint del sito dati.beniculturali.it;
- una suite di test, implementata come file OWL e query SPARQL, utilizzata per convalidare il grafo della conoscenza di ArCo;
- un file di allineamento esterno tra le classi e le proprietà di ArCo e le classi e le proprietà delle ontologie di CIDOC, EDM, Schema, OntoPiA per dimostrare l'interoperabilità semantica tra i modelli ontologici.

La rete di ontologie ArCo è il risultato di un progetto di eXtreme Design (XD), una metodologia consolidata basata sul riutilizzo di modelli di progettazione ontologica (*ontology design patterns, ODP*).

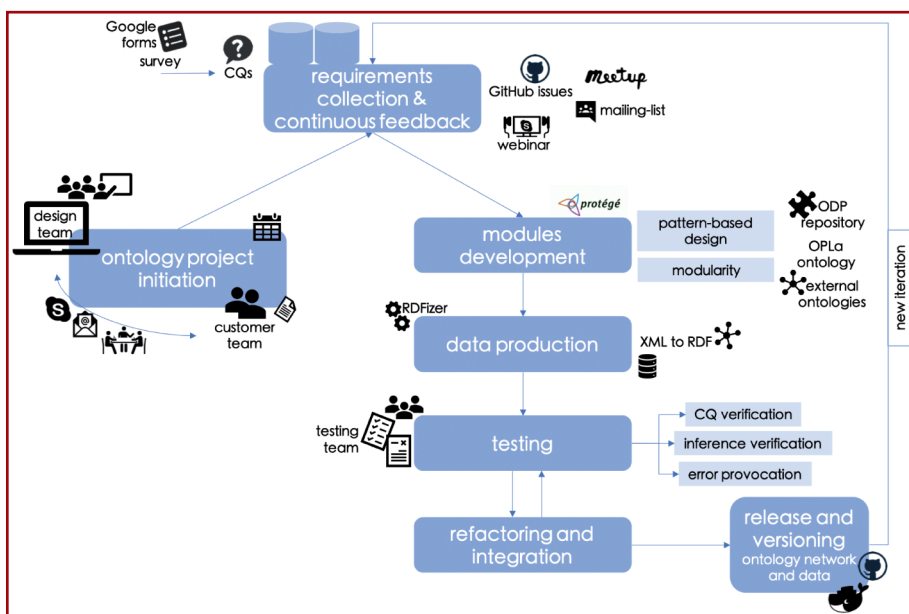


Figura 2. La metodologia eXtreme Design implementata per il knowledge graph di ArCo

Tale metodologia prevede un ciclo iterativo di test e feedback che coinvolge diversi attori:

- (i) un gruppo addetto alla progettazione dell'ontologia;
- (ii) un team di test, disgiunto dal team di progettazione, che si occupa di sperimentare l'ontologia;
- (iii) un team di attori interni ed esterni a ICCD, che esplicita i requisiti che vengono tradotti dal team di progettazione e dal team di test in impegni ontologici che guidano lo sviluppo dell'ontologia. Per rilevare eventuali incoerenze nell'ontologia, è stato eseguito regolarmente un ragionatore, Hermit, durante la fase di modellazione.

Tutte le proposte di miglioramento e le segnalazioni di malfunzionamento sono state raccolte su GitHub, su cui ArCo viene pubblicato in modo incrementale. Le comunità di sviluppatori e sperimentatori hanno interagito attraverso apposite mailing list, nonché partecipando a incontri in presenza e virtuali. Parallelamente allo sviluppo di moduli ontologici, sono sempre stati prodotti anche i dati in formato RDF²⁰ e a ciascuna risorsa RDF è stato assegnato un URI²¹. I dati sono stati infine collegati automaticamente a dati esterni usando il software LIMES, sulla base di euristiche stringenti che hanno garantito solo allineamenti verificabili.

Nella modellazione di ArCo, l'ICCD – insieme al CNR – ha coinvolto fin dall'inizio numerosi potenziali sviluppatori di applicazioni che usano le ontologie e i LOD prodotti nel settore dei beni culturali, includendo un programma c.d. di *early adoption*. Attraverso incontri periodici svolti con lo strumento dei *meetup*²² e il rilascio regolare e progressivo delle ontologie e dei dati – anche se in formato instabile – su Github, ICCD e STLab hanno voluto inaugurare una nuova stagione di partecipazione con gli *early adopter*, che hanno contribuito alla raccolta dei requisiti delle ontologie e allo sviluppo e miglioramento del progetto fornendo continui feedback.

²⁰ Ciò è avvenuto tramite lo sviluppo del software RDFizer: il suo componente principale è il convertitore XML2RDF, che prende in entrata un file XML conforme agli standard di catalogazione ICCD e lo trasforma in file in formato RDF attraverso fogli di stile XSLT.

²¹ L'URI per gli individui è <https://w3id.org/arco/resource/> con il prefisso data:. L'ID di ogni individuo è preceduto dal nome del suo tipo (ad esempio Agent, CulturalProperty ecc.). Per ogni tipo, sono stati attribuiti manualmente un insieme di elementi che costituiscono una possibile chiave (ad esempio AUTN, il nome dell'autore). Poiché i valori di questi elementi sono spesso stringhe, è stata rimossa la punteggiatura dai valori, sono stati convertiti in lettere minuscole, concatenati e ordinati in ordine alfabetico (ad esempio michelangelo-buonarroti). Infine è stato calcolato un MD5 checksum sulla stringa risultante, utilizzata come ID dell'URI, ad esempio dati: Agent/dcd4ca7b54dd3d7dac083dd4c54a9eef.

²² I meetup sono riunioni informali in cui si incontrano persone che hanno in comune un interesse culturale, tecnologico, sociale, ricreativo ecc. (dall'inglese "to meet up" che significa "incontrarsi per caso"). Si veda: <<https://www.meetup.com/it-IT/Meetup-Web-Progetto-ArCo-Architettura-della-conoscenza>>.

Conclusioni e prospettive

La rete di ontologie ArCo è stata dunque utilizzata per la modellazione dei dati del Catalogo generale dei beni culturali, gestita da ICCD e per la modellazione dei dati degli archivi fotografici di ICCD. Essa è stata utilizzata anche per la pubblicazione di LOD del Museo del Paesaggio sonoro²³ di Riva presso Chieri (TO) e per la pubblicazione dei LOD delle opere d'arte da parte dell'Istituto per i beni artistici culturali e naturali (IBACN) della Regione Emilia-Romagna. Anche la Regione Campania ha dichiarato di volere usare ArCo per i propri linked data nell'ambito del progetto Move to Cloud²⁴.

L'attività con StLab del CNR si è collocata in una fase in cui il Catalogo generale dei beni culturali online mostrava chiaramente alcune debolezze, riconducibili in buona parte ad una interfaccia in cui poca attenzione era data all'esperienza utente. La maggior parte degli utenti del Catalogo era infatti riconducibile a funzionari della pubblica amministrazione. Per ampliare l'utenza del sito e sperimentare i benefici del nuovo paradigma di pubblicazione dei LOD, ICCD ha deciso di ripensare il proprio sito di pubblicazione dei dati catalografici e ha promosso una *call to co-creation*, un contest a premi con cui sono stati coinvolti gli *early adopters* nella progettazione di modalità di fruizione innovative del sistema del Catalogo generale dei beni culturali. Alla gara hanno partecipato cinque aziende, tra cui una giuria di esperti ha selezionato la proposta vincitrice. Dalle ricerche fatte all'interno della piattaforma scelta e grazie ai LOD, l'utente potrà ottenere non solo la consultazione della risorsa digitale e dei dati descrittivi sul singolo elemento o gruppo di beni ricercato, ma anche la ricostruzione del contesto nel quale essi si collocano, evidenziando le relazioni esistenti fra gli elementi del patrimonio, i soggetti che li riguardano, i luoghi che ne sono lo scenario, le persone a cui sono legati, arricchendo le fonti di partenza con informazioni di qualità utili a fini conoscitivi, educativi, di ricerca, oltre che di valorizzazione. I dati del Catalogo mostreranno tutti i collegamenti verso altre basi di conoscenza (Regioni che pubblicano in LOD i dati sul proprio patrimonio, i LOD della Fototeca Zerì, Wikidata, Geonames, Europeana ecc.), ottenuti dinamicamente tramite procedure di riconciliazione semantica²⁵. I medesimi meccanismi sono stati utilizzati anche per collegare automaticamente i dati delle schede di catalogo con i dati presenti in quei *repository* a loro volta collegati a Wikidata (Enciclopedia Treccani, British Museum, VIAF ecc.). Infine, la riconciliazione è stata effettuata anche verso dati "interni", ovvero presenti nelle descrizioni inventariali degli archivi fotografici dell'ICCD, con l'obiettivo di dare or-

²³ <http://museopaesaggiosonoro.org/sound-archives-musical-instruments-collection-samic/>.

²⁴ Cfr. <<https://cultura.regione.campania.it/movetocloud>>.

²⁵ Con tale termine si intende fare riferimento ai meccanismi di allineamento basati sul concetto di identità (o di somiglianza) tra risorse nell'ambito del *semantic web*. Due risorse identiche a e b, condividono gli stessi attributi e le stesse relazioni. Se la risorsa a viene aggiornata (nei suoi attributi e nelle sue relazioni), di fatto si aggiorna anche la risorsa b. Con lo stesso significato, si parla anche di data linking, interlinking, data matching.

ganicità al complesso di dati catalografici e immagini fotografiche sul patrimonio che finora sono stati gestiti da ICCD su due banche dati separate.

Il passo successivo sarà l'integrazione dei LOD del Catalogo con ulteriori LOD sul patrimonio culturale, ad esempio della Regione Veneto²⁶, della Regione Emilia-Romagna²⁷, della Regione Campania²⁸ e gli open data della Regione Lombardia²⁹ e della Regione Sardegna³⁰: ciò consentirà, da una parte, di sperimentare l'interoperabilità tra i dati, superando i limiti dell'interoperabilità tra i sistemi, prefigurate già nel 2012 dalle *Linee guida per l'interoperabilità semantica*; dall'altra, contribuirà ad accrescere il *knowledge graph* sul patrimonio culturale e a fare un passo notevole verso la realizzazione del Catalogo nazionale del patrimonio culturale italiano, almeno in termini quantitativi.

In termini qualitativi, ICCD ha avviato, insieme all'Università di Bologna, il progetto ArCo4Science che, insieme alle collaborazioni con le università di Ferrara, La Sapienza, Roma Tre, mira all'implementazione del *knowledge graph* del patrimonio culturale italiano, integrando i dati del catalogo con i risultati della ricerca sul patrimonio culturale svolta in ambito universitario. Ciò avverrà, in particolare, attraverso la definizione ed esecuzione del processo di acquisizione e produzione di *knowledge graph* collegati al Catalogo generale dei beni culturali anche attraverso estrazione automatica di conoscenza da testo e attraverso la produzione e acquisizione di immagini e ricostruzioni 3D per la costruzione di ambienti virtuali semantici basati sul *knowledge graph* ArCo.

ICCD always aimed at sharing the models for structuring knowledge on cultural heritage and the data produced in cataloging campaigns. In recent years, the ICCD has focused its activities on the semantic web and its tools. One of the results is the ArCo project, the graph of knowledge of the Italian cultural heritage, consisting of a network of ontologies and more than 169 million triples referring to over 800 thousand catalog cards. ArCo is based on the data of the General Catalog of Cultural Heritage of the Central Institute for the catalog and documentation of MiBACT and on the data of its photographic archives. ArCo is jointly distributed with a SPARQL endpoint, a software to convert catalog records into RDF and a rich suite of documentation material (tests, evaluation, instructions, examples, etc.).

²⁶ <https://www.culturaveneto.it/it/footer/ontologia>.

²⁷ <https://ibc.regione.emilia-romagna.it/servizi-online/lod>.

²⁸ <https://cultura.regione.campania.it/ecosistema>.

²⁹ <https://www.dati.lombardia.it/Cultura/Opere-d-Arte-conservate-nei-Musei/dsrv-9ish>.

³⁰ <http://dati.regione.sardegna.it/dataset>.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

RIFERIMENTI BIBLIOGRAFICI

Valentina Anita Carriero – Aldo Gangemi – Maria Letizia Mancinelli – Andrea Giovanni Nuzzolese – Valentina Presutti – Chiara Veninata. *Pattern-based design applied to cultural heritage knowledge graphs*. (2019). ArXiv, <<https://arxiv.org/abs/1911.07585>>.

Valentina Anita Carriero – Aldo Gangemi – Maria Letizia Mancinelli – Ludovica Marinucci – Andrea Giovanni Nuzzolese – Valentina Presutti – Chiara Veninata. *ArCo: the Italian Cultural Heritage Knowledge graph*. In: *The Semantic Web – ISWC 2019. ISWC 2019. Lecture Notes in Computer Science*. Cham: Springer, 2019, p. 36–52. <<https://arxiv.org/abs/1905.02840>>.

Valentina Anita Carriero – Aldo Gangemi – Maria Letizia Mancinelli – Ludovica Marinucci – Andrea Giovanni Nuzzolese – Valentina Presutti – Chiara Veninata. *ArCo Ontology Network and LOD on Italian Cultural Heritage*. In: *Proceedings of the First International Workshop on Open Data and Ontologies for Cultural Heritage co-located with the 31st International Conference on Advanced Information Systems Engineering (CAiSE 2019), Rome, Italy, June 3, 2019*. <<http://ceur-ws.org/Vol-2375/>>.

Martin Doerr. *The CIDOC Conceptual Reference Module An Ontological Approach to Semantic Interoperability of Metadata*. «AI Magazine», 24 (2003), n. 3, p. 75-92.

Gordon Dunsire – Corey Harper – Diane Hillmann – Jon Phipps. *Linked data Vocabulary Management: Infrastructure Support, Data Integration, and Interoperability*. «Information Standards Quarterly (ISQ)», 24 (Spring/Summer 2012), n. 2/3.

Claudio Gnoli – Carlo Scognamiglio. *Ontologia e organizzazione della conoscenza: introduzione ai fondamenti teorici dell'indicizzazione semantica*. Lecce: Pensa Multimedia, 2008.

Francesca Tomasi – Marilena Daquino. *Linked Cultural Objects: dagli standard di catalogazione ai modelli per il web of data. Spunti di riflessione dalla Fototeca Zerì*. «Umanistica digitale», (2017), 1, p. 29-43.

Giorgia Lodi – Luigi Asprino – Andrea Giovanni Nuzzolese – Valentina Presutti – Aldo Gangemi – Diego Reforgiato Recupero – Chiara Veninata – Annarita Orsini. *Semantic Web for cultural heritage valorization*. In: *Data Analytics in Digital Humanities*, Cham: Springer, 2017, p. 3-37. <https://link.springer.com/chapter/10.1007/978-3-319-54499-1_1>.

Maria Letizia Mancinelli – Chiara Veninata. *Architettura della conoscenza: il sistema ICCD come modello per la descrizione dei beni culturali*. In: *Conferenza GARR 2018: selected papers. Data (R)evolution. Cagliari, 3-5 ottobre 2018*. Roma: Associazione consortium GARR, 2019, p. 130-134. DOI: 10.26314/GARR-Conf18-proceedings.

Laura Moro. *Il Catalogo del patrimonio culturale italiano. Nuova centralità e prospettive future*. «Economia della Cultura», 25 (2015), n. 3-4, p. 419-432.

Laura Moro. *Dalla somiglianza alla sinergia. La descrizione del patrimonio culturale dalla specificità all'integrazione digitale*. Intervento presentato in occasione dell'Incontro di studio tenutosi presso la Sala convegni dell'Archivio storico della Presidenza della Repubblica: *Dalla somiglianza alla sinergia. La descrizione del patrimonio culturale dalle specificità all'integrazione digitale*. Roma, 17 ottobre 2017. <<http://www.iccd.beniculturali.it/getFile.php?id=6083>>.

Sandra Vasco Rocca. *Beni culturali e catalogazione. Principi teorici e percorsi di analisi*. Roma: Gangemi, 2002.

Chiara Veninata. *I dati aperti di ICCD: verso l'interoperabilità semantica attraverso le ontologie e i linked open data*. In: *Atti del Convegno ArcheFoss. Cagliari 7-9 ottobre 2016*. In corso di pubblicazione.