



Un modello ontologico per l'integrazione delle informazioni del patrimonio culturale: CIDOC-CRM

Maria Teresa Biagetti

Introduzione

Uno degli obiettivi più significativi che le istituzioni della memoria perseguono, oltre ad assolvere alle basilari funzioni di custodia e di valorizzazione del patrimonio culturale, è rendere accessibili al pubblico le rappresentazioni digitali degli oggetti che conservano, per facilitare la conoscenza e la fruizione del patrimonio stesso, sostenere la ricerca scientifica e favorire l'educazione.

All'interno della cornice costituita da queste finalità, fornire l'accesso integrato alle informazioni e agli oggetti digitali gestiti da sistemi diversi, ciascuno dei quali presenta caratteristiche e modalità di organizzazione dei dati sue proprie, si è imposto negli ultimi anni come un obiettivo di primaria importanza, che può essere raggiunto grazie all'impiego di ontologie di alto livello.¹ Le

¹ La letteratura internazionale sulle ontologie formali è molto vasta e difficilmente sintetizzabile. Mi limito a citare qui un articolo fondamentale e un manuale tra i più completi: Guarino (1998) e Handbook on Ontologies (2009).
Doveroso citare il Laboratory for applied ontology





ontologie sono gli strumenti fondamentali per raggiungere l'interoperabilità semantica, poiché concettualizzano un dominio e agiscono come mediatori per la ricerca integrata degli oggetti digitali gestiti in diversi *repositories*. Nel Web semantico e nella esposizione dei dati in *dataset* costituiti da triple in RDF con le tecnologie *Linked open data*, le ontologie ricoprono un ruolo fondamentale nell'espressione delle relazioni semanticamente qualificate su cui si fonda il collegamento tra i dati.

Per permettere agli studiosi di accedere al patrimonio di dati e di informazioni presenti in fonti diverse, recentemente è stato sviluppato un approccio basato sul paradigma *Ontology-Based Data Access* (OBDA),² in base al quale i diversi *dataset* possono essere integrati virtualmente e interrogati attraverso un modello concettuale, un'ontologia, oggi prevalentemente scritta in OWL (OWL 2004; OWL 2009; Antoniou e Harmelen 2009), che modella concettualmente il dominio e definisce un vocabolario

(<http://www.istc.cnr.it/it/group/loa>) dell'Istituto di Scienze e Tecnologie della Cognizione del CNR di Trento (ISTC). Segnalo anche The International Association for Ontologies and its Applications (IAOA) <http://iaoa.org> e lo Special interest group su Semantic Web Applied Ontology (SWAO) sorto al suo interno. Sul sito di IAOA e sulla sua rivista pubblicata online, "Applied Ontologies" <http://iaoa.org/jao/jao.html>, si possono reperire le informazioni bibliografiche più recenti sull'argomento. In lingua italiana si può vedere Le ontologie (2010).

² Sono stati realizzati diversi sistemi *ontology-based*. Mi limito a citarne due: Ontop, sistema *open source* sviluppato dal gruppo di ricerca *Knowledge Representation meets Databases* (KRDB) della Libera Università di Bolzano <http://ontop.inf.unibz.it> e MASTRO, sviluppato presso il Dipartimento di Ingegneria informatica, automatica e sistemistica di Sapienza Università di Roma insieme alla Libera Università di Bolzano. Su MASTRO si può vedere, accessibile anche ai non specialisti, Civili, Console, De Giacomo (2013) e, più impegnativo per chi non è esperto, Calvanese, De Giacomo, Lembo (2011).



condiviso, funzionale alla ricerca nel settore specifico. Le tecnologie basate su strumenti semantici per la gestione dei dati permettono di ritrovare le informazioni distribuite in diversi *dataset* e di fare ricerche anche partendo da rappresentazioni diverse della conoscenza. In un sistema OBDA l'utente presenta la sua *query* utilizzando l'ontologia, e un sistema di *mapping* permette la connessione tra le classi e le proprietà dell'ontologia con i dati delle fonti documentarie. L'ontologia viene utilizzata per accedere ai dati gestiti nei *repositories*, senza che sia necessario conoscere la reale organizzazione dei dati stessi.

Questo approccio risulta particolarmente efficace nei casi in cui i *dataset* abbiano subito nel tempo trasformazioni nella loro organizzazione, e le fonti informative si presentino al momento distribuite o replicate. I diversi sistemi sviluppati secondo il paradigma OBDA assicurano l'uso di una ontologia elaborata in accordo con la concettualizzazione del dominio di interesse per il quale il sistema è stato sviluppato, e indipendente da esso.

Sia la ricerca storica che la ricerca prodotta nell'ambito degli studi relativi al patrimonio culturale, inteso in senso lato, si avvalgono oggi di sistemi che sfruttano questo paradigma. Un esempio recentissimo è costituito dal progetto EPN_{et}, avviato nel 2014, relativo allo studio del sistema economico e commerciale dell'Impero romano sulla base di ampi *dataset* costituiti dai dati epigrafici reperiti sulle anfore (Calvanese et al. 2015), nel quale viene utilizzato un sistema *ontology-based* per il collegamento virtuale e l'accesso ai dati strutturati in diversi database.

Tuttavia, anche i sistemi informativi più recenti basati sul paradigma OBDA hanno la necessità di assicurarsi l'interoperabilità con altri sistemi e con altre iniziative scientifiche,



intraprese nello stesso settore di ricerca. Per questo motivo, tutti i sistemi devono tenere presente CIDOC-CRM di ICOM,³ la più importante e vasta ontologia del settore del patrimonio culturale. Infatti, la porzione maggiore del Conceptual Reference Model per il sistema EPNet è stata modellata tenendo presente CIDOC-CRM, di cui costituisce una estensione (Calvanese et al. 2015, 293).⁴

CIDOC-CRM è una ontologia formale il cui scopo è rendere possibile lo scambio e l'integrazione delle descrizioni, delle informazioni e della documentazione per la ricerca scientifica tra fonti eterogenee del patrimonio culturale: collezioni museali, siti archeologici, monumenti, e documentazione scientifica conservata negli archivi e nelle biblioteche (Doerr 2003; Doerr 2009, 463–86).

È necessaria, quindi, un'analisi approfondita di questa ontologia, impostasi a livello internazionale e ampiamente utilizzata nei settori artistico, storico e archeologico,⁵ per contribuire a

³ ICOM – International council of museums <http://icom.museum>, istituita nel 1946, è una organizzazione internazionale di istituzioni e di professionisti nel campo museale, dedicata a definire standard e buone pratiche per la gestione delle collezioni museali. Attualmente fanno parte di ICOM 20.000 musei e 35.000 esperti del settore, e sono riconosciuti 119 comitati nazionali e 30 comitati internazionali, tra i quali International Committee for Documentation, CIDOC.

⁴ Più approfondito e ricco di dettagli tecnici Calvanese et al. (2016).

⁵ La letteratura sull'utilizzo di CIDOC-CRM in ambito archeologico è particolarmente ricca. L'ontologia viene impiegata per indicizzare, scambiare e riutilizzare le informazioni durante e dopo gli scavi archeologici. Rinvio ai lavori di Douglas Tudhope e Ceri Binding, della Glamorgan University, per i ragguagli sull'argomento, e per quanto riguarda l'Italia, alle ricerche di Franco Niccolucci e di Achille Felicetti del VASTLAB - PIN ScRL dell'Università degli Studi di Firenze.



metterne in rilievo i fondamenti, tutte le particolarità e i meccanismi di funzionamento.

Il presente articolo è strutturato in due parti: la prima è dedicata ad illustrare le peculiarità delle ontologie e della *Description Logics*, e costituisce lo sfondo e il supporto teorico per la parte successiva, nella quale sono presentate le caratteristiche dell'ontologia CIDOC, dando spazio all'analisi della sua modellazione logica e alle scelte teoretiche di fondo operate dai suoi realizzatori. Obiettivo di questo articolo non è tanto presentare una rassegna dei progetti nei quali CIDOC-CRM viene utilizzato, oggi anche nella realizzazione di *endpoint* SPARQL per la consultazione di dati collegati, ma di fare emergere i fondamenti logici sui quali si basa e portare alla luce l'impostazione teoretica sottostante.

Ontologie, Logica e rappresentazione della conoscenza

Prima di presentare nei dettagli e nelle sue articolazioni l'ontologia CIDOC-CRM, è necessario mettere in evidenza i fondamenti logici su cui si basano le ontologie, in modo da poter fornire una introduzione teoretica al successivo esame dell'ontologia oggetto di questo studio, della quale, infatti, oltre ai meccanismi di funzionamento e alle caratteristiche strutturali, si intende mettere in evidenza proprio l'impostazione teoretica di base, e le sue peculiarità specifiche per l'organizzazione della conoscenza storica.

Le ontologie di dominio sono costituite dalla specificazione formale di una concettualizzazione di un dominio scientifico o applicativo; le *upper ontologies* (o *top-level ontologies*), dette anche in



letteratura ontologie fondazionali, sono invece ontologie che rappresentano concetti e proprietà molto generali, ai quali si possono collegare i termini più specifici presenti nelle ontologie di dominio e leggere, e sono applicabili ad un vasto insieme di ambiti cognitivi. Tra le ontologie fondazionali si ricordano SUMO e DOLCE. *Suggested Upper Merged Ontology* (SUMO),⁶ è stata sviluppata a partire dal 2000 e attualmente è gestita da Institute of Electrical and Electronics Engineers (IEEE). Si è imposta come standard per questo tipo di ontologie e agisce come ontologia fondazionale per le ontologie di dominio. La sua sezione *upper-level*, visionabile separatamente, contiene infatti le entità generali che non appartengono ad alcun dominio scientifico, e che possono servire ad un ampio numero di campi scientifici: le funzioni numeriche, le unità di misura, i concetti temporali, le relazioni spaziali, i processi e le proprietà generali. L'ontologia fondazionale DOLCE,⁷ sviluppata da Nicola Guarino e dall'Istituto di Scienze e Tecnologie della Cognizione del CNR di Trento nell'ambito del progetto WonderWeb, rispecchia le strutture del linguaggio naturale e dell'attività cognitiva umana e di senso comune.

Le ontologie sono realizzate adoperando un linguaggio formale, che deve prevedere sintassi e semantica definite, requisiti necessari per il trattamento delle informazioni da parte delle macchine. La semantica formale assicura la descrizione di un significato univoco, ed impedisce l'interpretazione soggettiva da

⁶ www.ontologyportal.org. Sul sito sono visionabili diverse ontologie di dominio, dedicate alla Comunicazione, all'Economia, alla Finanza, alla Geografia, all'Ingegneria ecc., mantenute collegate a SUMO.

⁷ Una versione abbreviata è visionabile all'URL: <http://www.loa.istc.cnr.it/ontologies/DUL.owl>. Si può leggere una presentazione in lingua italiana in Gaio et al. (2010).



parte di agenti differenti, che siano esseri umani o macchine, e in particolare permette di descrivere gli oggetti e le relazioni dichiarate in una ontologia in una modalità accessibile alle macchine.

Le ontologie per il Semantic Web oggi vengono modellate utilizzando il linguaggio OWL. OWL è basato sulla *Description Logics*, un sottoinsieme della Logica del primo ordine (*First-order Logic*), in base alla quale si possono esprimere le proposizioni sugli oggetti, sulle relazioni tra gli oggetti e sulle proprietà che gli oggetti presentano.

Description Logics è un linguaggio formale per la rappresentazione della conoscenza e permette il ragionamento su di essa attraverso le inferenze. E' dotata di una semantica formale, che costituisce la base del ragionamento sugli oggetti e sulle conoscenze tramite le relazioni di appartenenza di un individuo ad una classe (*class membership*), l'equivalenza delle classi e la coerenza dei rapporti tra classi e sottoclassi; consente inoltre il processo di classificazione per mezzo del riconoscimento delle proprietà. *Description Logics* utilizza formalismi descrittivi che comprendono gli operatori di congiunzione, di negazione, di intersezione di concetti, di quantificazione esistenziale, di restrizione dei valori e di restrizione numerica (cardinalità) (Nardi e Brachman 2003, 1–40).⁸

I linguaggi basati sulle Logiche descrittive forniscono agli utenti - agenti umani o automatici - la possibilità di compiere inferenze, cioè dedurre conoscenze dalla rete di conoscenze esplicitate. Le inferenze vengono realizzate applicando gli algoritmi di

⁸ Una trattazione più approfondita e tecnica, e quindi meno facilmente fruibile dai non specialisti, delle caratteristiche della DL, in Baader e Nutt (2003, 47–100).



subsumption, che determina le relazioni gerarchiche tra i concetti, di *instance*, che specifica l'appartenenza di un individuo ad una classe, e di *consistency*, che rende possibile l'analisi della coerenza logica e dell'assenza di contraddizione tra i concetti (Baader, Horrocks, e Sattler 2009, 21–43). Le relazioni gerarchiche tra i concetti, a partire dalla relazione fondamentale *IsA* (es.: Mother *IsA* Parent), costituiscono le basi per l'ereditarietà delle proprietà, cioè il fatto che un concetto più specifico eredita le proprietà del concetto più generale, e danno luogo alle reti di ereditarietà.

La logica convenzionale è monotonica, e questo significa che se una inferenza è valida, le sue conclusioni non potranno mai essere eliminate, neanche quando una nuova premessa che contraddice le premesse esistenti venga aggiunta. Una inferenza è valida quando le conclusioni sono vere in tutti i casi in cui le premesse sono vere, e in base alla logica monotonica essa rimane valida anche quando vengano aggiunti assiomi, o proposizioni considerate vere in contrasto con quelle già esistenti. Nella realizzazione di conseguenze logiche, la monotonicità assicura che le premesse di ogni conclusione derivata possano essere estese, inserendo nuove premesse, senza per questo invalidare le conclusioni originarie. Non si produce, quindi, alcuna riduzione nell'insieme delle conseguenze logiche originario. Semplificando, si può affermare che un nuovo elemento conoscitivo non può indebolire, o ridurre, ciò che già era conosciuto.

Molti studiosi di Logica, di Filosofia e di Psicologia cognitiva hanno rilevato da tempo quanto questa impostazione sia in contrasto con l'attività cognitiva umana, naturalmente portata a riconsiderare le proprie conoscenze a seguito di nuovi elementi o fatti accaduti. Gli scienziati cognitivisti hanno sviluppato un nuovo paradigma, la logica probabilistica (*p-Logic*), basata sull'accettazione delle condizioni soggettive, nell'ambito della



quale il concetto di validità probabilistica sostituisce quello della validità logica.⁹

La logica probabilistica è non-monotonica, così come il ragionamento umano comune, il ragionamento scientifico e quello abduttivo. Al contrario della classica *First-order Logics*, che è monotona, le logiche non-monotone (Strasser e Antonelli 2015) sviluppate nell'ambito dell'Intelligenza artificiale, ammettono la fallibilità e permettono la correzione delle conclusioni in presenza di fatti che provino il contrario di quanto asserito dalle premesse; si tratta di logiche ampliative, che ammettono che le conclusioni vadano oltre i significati coinvolti nelle premesse. Nel ragionamento invalidabile, *defeasible reasoning*,¹⁰ a premesse vere possono seguire anche conclusioni false, in quanto la relazione tra le premesse e le conclusioni può essere invalidata da informazioni aggiuntive.¹¹ La logica non-monotona prevede che alcune conclusioni possano essere invalidate nel momento in cui viene aggiunta nuova conoscenza in una base di conoscenza.

⁹ Per una presentazione dei fondamenti della Logica convenzionale, dei nuovi paradigmi della *p-Logic*, cioè della Logica probabilistica, e della teoria dei modelli mentali, si può vedere Johnson-Laird, Khemlani, Goodwin (2015).

¹⁰ Si veda la trattazione dal punto di vista filosofico di Pollock (1987).

¹¹ "Reasoning is *defeasible* when the corresponding argument is rationally compelling but not deductively valid. The truth of the premises of a good defeasible argument provide support for the conclusion, even though it is possible for the premises to be true and the conclusion false. In other words, the relationship of support between premises and conclusion is a tentative one, potentially defeated by additional information." (Koons 2014).



Un modello concettuale di riferimento per il patrimonio culturale: CIDOC-CRM

Per le istituzioni del patrimonio culturale risulta fondamentale poter integrare le informazioni che riguardano i diversi contesti storici di cui i singoli oggetti conservati fanno parte, e dai quali traggono la propria forza documentaria. L'integrazione delle informazioni non può basarsi sugli schemi di metadati, ma necessita di un modello più ricco e articolato, capace di cogliere tutti gli elementi dei contesti storici di riferimento, e di rappresentare la semantica soggiacente a decine di schemi particolari.

Il modello concettuale predisposto da CIDOC è stato sviluppato tenendo presenti in particolare le esigenze dei musei. Il museo acquisisce, conserva e rende disponibili al pubblico in modo permanente il patrimonio tangibile e intangibile dell'umanità e il suo ambiente, per fini di studio e di libera fruizione.

La definizione di museo fornita da International Council of Museums (ICOM) abbraccia non solo le collezioni museali, ma anche i monumenti, i siti archeologici, e gli ambienti culturali in cui essi sono inseriti e, facendo riferimento al patrimonio intangibile, accoglie anche le tradizioni culturali delle diverse comunità. Ai musei viene affidato il compito di conservare e rendere fruibile il patrimonio culturale tramandato oralmente dai gruppi culturali, registrato e digitalizzato, insieme alla musica, alla danza e al teatro tradizionali delle diverse culture.¹² Per quanto

¹² "Section 1. Museum. A museum is a non-profit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study



riguarda il patrimonio culturale intangibile, la definizione presentata da ICOM è in linea con la *Convention for the Safeguarding of the Intangible Cultural Heritage* approvata da UNESCO nel 2003¹³ e che faceva riferimento appunto a cinque ampi settori di manifestazione del patrimonio culturale intangibile: le tradizioni orali e le espressioni linguistiche, la musica, la danza, le rappresentazioni sceniche (*performing arts*), le pratiche sociali e rituali, le conoscenze e le pratiche che riguardano il patrimonio naturale, e le abilità artigiane. La Convenzione si era proposta come obiettivo quello di assicurare la protezione, lo sviluppo, la documentazione e la preservazione del patrimonio intangibile.

CIDOC-CRM si propone come un modello ampio ed elastico, in grado di rendere evidente la logica di ciò che le diverse strutture informative documentano. Propone un modello di struttura intellettuale, in termini logici, della documentazione culturale, e quindi rende possibile l'interoperabilità semantica, cioè la capacità dei diversi sistemi informativi di comunicare informazioni coerenti con un significato "inteso", ma non definisce la terminologia usata dalle diverse strutture.

and enjoyment." <http://icom.museum/the-organisation/icom-statutes/3-definition-of-terms/#sommairecontent>.

"Intangible Heritage. Cultural heritage does not only embrace the tangible expressions like monuments and objects throughout the years. It also includes the living expressions like the traditions that many groups and communities worldwide have been passed down by their ancestors and will continue to pass on to their descendants, mostly by word of mouth. Although it is the motor of cultural diversity, this heritage is fragile."

<http://icom.museum/programmes/intangible-heritage>.

¹³ <http://www.unesco.org/culture/ich/en/1com>.



Concretamente, si offre come supporto alla modellazione concettuale per gli sviluppatori dei sistemi informativi; come linguaggio formale per identificare i contenuti informativi comuni nei diversi formati di dati; come modello per la integrazione e lo scambio dei dati tra i singoli sistemi informativi e un contenitore globale, senza correre il rischio di perdere i significati; infine, come modello per formulare le *queries* da rivolgere a sistemi di risorse integrate.

La possibilità di scambiare e integrare la documentazione scientifica eterogenea delle collezioni museali si applica a tutti i materiali coperti dalla definizione di ICOM: collezioni, siti e monumenti che riguardano l'etnografia, l'archeologia, l'arte, la storia naturale, la storia della scienza e della tecnologia e il patrimonio immateriale registrato. La documentazione delle collezioni include la descrizione dettagliata degli item individuali contenuti nelle collezioni, e in particolare le informazioni relative ai contesti storici e geografici. Un secondo obiettivo è lo scambio di informazioni con biblioteche e archivi. Non rientra invece nelle finalità di CRM il trattamento dei documenti amministrativi, relativi alla gestione delle istituzioni culturali.

CIDOC-CRM è stata sviluppata a partire dal 1996 da un gruppo di lavoro composto da informatici, archeologi, storici dell'arte, filosofi e studiosi di Biblioteconomia, con la partecipazione di ICOM e di CIDOC.¹⁴ La realizzazione ha seguito un procedimento *bottom up*, durante il quale sono state selezionate

¹⁴ International Committee for Documentation (CIDOC) è una delle commissioni internazionali di ICOM. Si occupa della preservazione degli oggetti digitali, degli standard e dello sviluppo di modelli di dati per i musei, dello sviluppo di strumenti per la ricerca semantica applicata ai musei, e dal 1992 della facilitazione dello scambio della documentazione durante gli scavi archeologici. <http://network.icom.museum/cidoc>.



dalle basi di dati strutturati, rappresentative del settore, le entità utili a rappresentare la conoscenza relativa al patrimonio culturale conservato in particolar modo nei musei, e sono stati integrati in un'unica struttura i contenuti semantici di centinaia di schemi di database dei settori museale, archivistico e bibliotecario. Il fatto che un concetto fosse presente in una base di dati ampiamente utilizzata nel settore di riferimento, ha costituito la garanzia per accogliere quel concetto nello schema. Per lo sviluppo del modello concettuale, tuttavia, è stato necessario raggiungere l'*ontological commitment*, l'accordo tra gli esperti di un dominio scientifico sul fatto che i concetti contenuti nell'ontologia risultino di fondamentale importanza per il settore, siano rilevanti e condivisi (Doerr, Ore, e Stead 2007).

È un'ontologia formale, creata con l'obiettivo di permettere lo scambio e l'integrazione delle informazioni tra fonti eterogenee del patrimonio culturale. Offre il modello, l'architettura semantica, per trasformare le diverse fonti in una risorsa globale, che può trovare la sua realizzazione in una biblioteca digitale di vaste dimensioni o in una rete di fonti eterogenee. Definisce la semantica che sottostà ai diversi schemi e alle differenti strutture documentali usate nel patrimonio culturale. Si tratta di una ontologia che descrive la concettualizzazione di un dominio specifico – il patrimonio culturale – ma che presenta categorie e classi di alto livello, come *Temporal entity*, *Period*, *Activity*, *Modification*, *Persistent item*, *Creation*, *Conceptual object*, ecc.

L'obiettivo era modellare uno strumento adatto alla realizzazione dell'interoperabilità semantica tra gli schemi di metadati già strutturati dalle istituzioni della memoria, musei, archivi, biblioteche, che permettesse di rappresentare la conoscenza offerta dalle collezioni di informazioni già costituite, consentisse l'accesso integrato a dati già strutturati, e di navigare all'interno di



collezioni eterogenee e distribuite.¹⁵ Questa ontologia, infatti, non è stata sviluppata allo scopo di strutturare la terminologia relativa agli *item* che fanno parte del patrimonio culturale, compito tra l'altro già assolto, per la Storia dell'arte e l'Architettura, da *Art and Architecture Thesaurus*,¹⁶ ma per cogliere la semantica soggiacente alle strutture diverse di dati che appartengono ad istituzioni eterogenee del settore di riferimento.

La semantica di alto livello risulta più stabile ed è più facile raggiungere un accordo tra settori diversi ma convergenti di quanto non si possa fare prendendo in considerazione la terminologia. Per questo motivo le classi che rappresentano la terminologia sono state mantenute separate rispetto allo schema concettuale, sono state trattate come dati e connesse all'ontologia solo attraverso la proprietà "P2 *has type*", come vedremo in seguito.

Fin dalla fase iniziale, per la rappresentazione di questa ontologia è stato scelto il modello *object-oriented*, costituito da classi raggruppate in gerarchie e collegate l'un l'altra attraverso le proprietà, allo scopo di poterla convertire in altri modelli analoghi e realizzare facilmente future estensioni (Crofts, Doerr, e Gill 2003) ed è stato usato il linguaggio formale per la

¹⁵ Una chiara presentazione della funzione di integrazione e di mediazione tra record realizzati usando schemi di metadati diversi (DC, EAD, MODS, ecc.) svolta dal modello concettuale CIDOC-CRM, con la mappatura tra Dublin Core Collections Application Profile (DCCAP) e CIDOC-CRM, in Lourdi, Papatheodorou, Doerr (2009).

¹⁶ Sviluppato da The Getty Research Institute negli anni Ottanta, AAT fu pubblicato a stampa nel 1990 (Petersen e Getty art history information program 1990). Oggi è consultabile la versione online, e insieme agli altri vocabolari del Getty Institute i suoi dati sono esposti come *Linked Open Data* nella LOD cloud. <http://www.getty.edu/research/tools/vocabularies/aat>.



rappresentazione della conoscenza TELOS (Mylopoulos et al. 1990) che permetteva di definire classi, relazioni, *instances* delle classi, e rappresentare le entità temporali e i cambiamenti nel tempo. Ma la versione a disposizione in rete è da sempre stata presentata volutamente in formato testuale, con *scope notes* che descrivono l'intensione delle classi e delle proprietà.

È stata approvata come standard ISO: 21127 nel settembre 2006, con revisione nel 2014. Attualmente sono disponibili anche alcune versioni equivalenti, una codificata in RDFS, con le definizioni delle classi e delle proprietà in diverse lingue, basata sul draft della versione 6.2.1 del 2012, e una in OWL2, realizzata nel 2009.

La versione più recente di CIDOC-CRM, la 6.2.1, è stata rilasciata nell'ottobre 2015: ICOM-CIDOC. *Definition of the CIDOC Conceptual Reference Model*. Produced by the ICOM/CIDOC Documentation Standards Group, Continued by the CIDOC CRM Special Interest Group Version 6.2.1 October 2015. Current Main Editors: Patrick Le Boeuf, Martin Doerr, Christian Emil Ore, Stephen Stead.¹⁷

In questa ultima versione 6.2.1 CRM presenta 94 classi e 168 proprietà, con un incremento notevole delle proprietà, che erano 149 nella versione 5.0.4, mentre le classi erano 90.¹⁸

¹⁷ http://www.cidoc-crm.org/docs/cidoc_crm_version_6.2.1.pdf

¹⁸ Per avere maggiori dettagli sul processo di elaborazione negli anni di questa ontologia e visionare le diverse versioni che si sono succedute nel tempo: http://www.cidoc-crm.org/official_release_cidoc.html. Per quanto riguarda la fusione che è stata realizzata con l'ontologia ABC, sviluppata per scambiare informazioni nell'ambiente delle biblioteche digitali, per le classi e le proprietà, si può vedere Doerr, Hunter, Lagoze (2003).



Principi di modellazione di CIDOC-CRM

Gli sviluppatori di questa ontologia si sono assunti il compito di elaborare un modello concettuale formale, che fosse però compatibile con la conoscenza reperibile nelle fonti prese come riferimento, cioè con la conoscenza storica. Ciascuna descrizione del passato, che sia stata realizzata nel passato, è da considerarsi unica e non sempre è verificabile. Una ontologia destinata a creare un supporto per la conoscenza che deriva dai dati storici, quindi, secondo gli sviluppatori di CIDOC-CRM, non solo dovrà osservare i principi ontologici, ma dovrà anche prendere in considerazione e rispettare i principi epistemologici in base ai quali la conoscenza è stata sviluppata. Questo significa che il modello deve essere ontologicamente fondato, ma abbastanza elastico da essere compatibile con la granularità della conoscenza proveniente dai dati storici, e permettere di accogliere anche informazioni contraddittorie relative alle *instances* (Doerr 2003, 80).

Il riferimento è alla *Open World Assumption* e si assume come principio base il fatto che le informazioni gestite dai sistemi informativi siano incomplete; secondo questa impostazione non si può concludere che siano false le proposizioni che, in base alla conoscenza del momento, non è possibile provare che siano vere; una visione opposta a quella prospettata dalla *Close World Assumption*, che comporta invece l'assunzione della completezza degli elementi in un *set* di individui, e ammette che si possa concludere che una proposizione sia falsa, se non si può provare che sia vera. In genere nel Web prevale la visione della *Open World Assumption*, e la semantica adottata dai linguaggi ontologici, come



RDFS¹⁹ e OWL,²⁰ è ispirata a questa visione (Antoniou e Harmelen 2009).

CRM è stato modellato avendo come riferimento il ragionamento monotonic e per questo motivo esso può permettere la fusione di enormi basi di conoscenza senza che si creino conflitti.

A reasoning form is monotonic if an addition to the set of propositions making up the knowledge base never determines a decrement in the set of conclusions that may be derived from the knowledge base via inference rules. In practical terms, if experts enter subsequently correct statements to an information system, the system should not regard any results from those statements as invalid, when a new one is entered. The CRM is designed for monotonic reasoning and so enables conflict-free merging of huge stores of knowledge. (ICOM/CIDOC 2015, XI)

La Logica monotona prevede che ciascuna proposizione considerata vera e ben formata continui ad essere considerata vera anche dopo che siano stati aggiunti altri assiomi o altre proposizioni considerate vere. Nel caso di CIDOC CRM, ciascuna *instance* CRM e le sue proprietà devono rimanere valide e ben formate anche se nuove *instances* vengono aggiunte, coerenti con il dominio di conoscenze definito dagli esperti. Avere accettato la monotonicità come principio di modellazione logica comporta il fatto che i costrutti esistenti in CRM e le deduzioni realizzate su di essi devono sempre restare valide e ben formate anche se vengono aggiunti nuovi costrutti usando delle

¹⁹ <http://www.w3.org/TR/rdf-schema>. Più completo:

<http://www.w3.org/TR/rdf-mt>

²⁰ www.w3.org/TR/owl2-overview; <http://www.w3.org/TR/2009/REC-owl2-primer-20091027>.



“estensioni” a CRM. Ad esempio, si potrebbe aggiungere come estensione alla classe E7 *Activity* una certa sottoclasse per descrivere una particolare pratica attivata in un certo luogo e in un certo tempo, e nessuna delle relazioni *IsA* o delle eredità delle proprietà sarebbero compromesse (ICOM/CIDOC 2015, XV). Se una inferenza è valida, le conclusioni che ne derivano non potranno mai essere eliminate a causa dell’inserimento di nuove premesse che le contraddicono.

CRM di CIDOC preserva la monotonicità anche quando vengono aggiunte al sistema nuove *instances*: le *instances* esistenti, le loro proprietà e le deduzioni realizzate in base ad esse, rimarranno valide e ben formate. Un esempio presentato nella versione 6.2.1 dell’ontologia: se si descrive correttamente un *item* come istanza di E19 *Physical Object*, e successivamente lo stesso *item* venisse caratterizzato come *instance* di E20 *Biological Object*, il sistema non dovrebbe smettere di trattarlo come *instance* di E19 (ICOM/CIDOC 2015, XV). Si preserva la monotonicità per salvaguardare le opinioni alternative, e questo consente di accumularle. Allora, la conoscenza che è stata integrata usando CRM, può servire come base di ricerca, in cui sono accumulate opinioni alternative circa entità ben definite, laddove invece le conclusioni circa la verità sono l’obiettivo delle costruzioni *open-ended*. Si possono così accumulare opinioni alternative nella base di conoscenza, e anche informazioni incomplete, in quanto tutte le proprietà devono essere considerate opzionali e ripetibili per quanto riguarda il loro *Domain* e il loro *Range*. Questo significa che una volta inseriti nella base di conoscenza come *instances* di E21 *Person*, ad esempio, sia El Greco che King Arthur, considerati come esistenti, opinioni alternative riguardanti ad esempio proprietà come il luogo di nascita devono essere aggiunte senza sottoporre ad un vaglio di validità le proprietà già esistenti (ICOM/CIDOC 2015, XV).



Una categorizzazione utilizzata in CIDOC-CRM riguarda il concetto di identità attraverso il tempo, e la distinzione che viene applicata alle *instances* delle classi a seconda del modo in cui esse sono presenti nel tempo. La distinzione filosofica è esercitata tra *endurants*, le entità che “sono” nel tempo, le cui parti sono tutte presenti nel momento in cui esse sono presenti (una persona, il foglio che sto leggendo) e *perdurants*, le entità che “divengono” nel tempo e che nel momento in cui sono presenti, sono solo parzialmente presenti, e le loro fasi precedenti o passate possono non essere presenti, come nel caso di azioni e di processi (lo scorrere della vita; le porzioni di testo che ho già lette su quel foglio). Gli *endurants* partecipano al divenire e allo svolgersi dei *perdurants*. Queste sono categorie adottate anche dall'ontologia fondazionale DOLCE.

In CIDOC-CRM, ad esempio, le istanze della classe E2 *Temporal Entity*, che in realtà è una classe astratta e non ha istanze dirette, ma raggruppa le istanze con una componente temporale delle classi E4 *Period*, E5 *Event* e E3 *Condition State*, sono da considerarsi *Perdurants* (es.: l'età del bronzo), mentre le istanze della classe E77 *Persistent Item*, e delle sue sottoclassi E39 *Actor* e E70 *Thing* (e, ancora, delle sottoclassi di queste ultime), sono da considerarsi *Endurants*, in quanto, pur avendo una esistenza limitata nel tempo, preservano la loro identità nel corso degli eventi: possono essere entità fisiche (es.: Leonardo da Vinci) o concettuali (es.: Prima legge della Termodinamica).

La dichiarazione di disgiunzione delle classi è espressa nel caso in cui le classi non abbiano *instances* in comune, in alcuno dei mondi possibili. E2 *Temporal entities* è disgiunta da E77 *Persistent item*. E18 *Physical thing* è dichiarata nell'ontologia disgiunta da E28 *Conceptual object*. In quest'ultimo caso la distinzione è operata tra *items* materiali e immateriali.



Caratteristiche di CIDOC-CRM

CIDOC-CRM dichiara una classe solo se questa è necessaria come *Domain*²¹ o come *Range*²² di una proprietà, oppure se è un concetto-chiave per il settore. La maggior parte delle classi in CRM sono concetti primitivi, il cui significato non deve essere dedotto logicamente da altri concetti.

Il centro propulsore di questa ontologia, tuttavia, è costituito dalle proprietà o relazioni, e non dalle classi. Avere affidato il ruolo primario alle proprietà collocando al centro dello schema le relazioni tra le entità, invece di fondarsi sulla definizione della natura degli individui coinvolti, sembra contrastare con l'impostazione adottata in tutti i modelli *object-oriented* e rivela invece l'attenzione di natura epistemologica che questo sistema ha rivolto alle modalità della produzione della conoscenza, come sottolineato più volte da Martin Doerr (Doerr 2003, 84 e 90).

Le proprietà possono essere presentate attraverso il raggruppamento logico in gruppi. Il primo e fondamentale raggruppamento riguarda le proprietà che si possono sussumere sotto i concetti di *Participation and Spatio-temporal Reasoning* che, secondo i modellatori di questa ontologia, caratterizzano la semantica della struttura dei dati nell'area culturale e permettono di fare deduzioni valide sugli attori di un evento e sulle loro relazioni senza considerare la certezza, ad esempio, della datazione. La modellazione esplicita degli eventi permette di integrare più facilmente i contenuti culturali diversi. La presenza

²¹ Il *Domain* limita gli individui cui la proprietà si applica: ad esempio, la proprietà <avere figli> si applica al *Domain* Mammiferi.

²² Il *Range* definisce quali individui la proprietà può avere come suoi valori, è una restrizione che limita gli individui che possono essere i valori di quella proprietà.



di entità differenti – persone, luoghi, attività – che convergono in un evento, dovrebbe permettere di stabilire l'interazione tra quelle entità nello stesso momento e lo scambio di informazioni tra gli attori, o l'influenza di una entità sull'altra. Le proprietà centrali sono P11 e P12.

“P11 had participant (participated in). Domain: E5 Event; Range: E39 Actor.

[...] This property describes the active or passive participation of instances of E39 Actors in an E5 Event. It connects the life-line of the related E39 Actor with the E53 Place and E50 Date of the event. The property implies that the Actor was involved in the event but does not imply any causal relationship. The subject of a portrait can be said to have participated in the creation of the portrait. Examples: Napoleon (E21) *participated in* The Battle of Waterloo (E7) [...]” (ICOM/CIDOC 2015, 48).

“P12 occurred in the presence of (was present at). Domain: E5 Event; Range: E77 Persistent Item.

[...]. This property describes the active or passive presence of an E77 Persistent Item in an E5 Event without implying any specific role. It connects the history of a thing with the E53 Place and E50 Date of an event. For example, an object may be the desk, now in a museum on which a treaty was signed. The presence of an immaterial thing implies the presence of at least one of its carriers.” (ICOM/CIDOC 2015, 49). (*Persistent Item* possono essere entità fisiche, come persone e oggetti, e entità concettuali, come idee e concetti).

Un altro significativo insieme riguarda le proprietà raggruppabili sotto il concetto di *Influence*, che concerne le influenze reciproche



delle diverse attività umane. Significativa la proprietà “*P15 was influenced by (influenced)*. Domain: E7 Activity; Range: E1 CRM Entity”, una proprietà di alto livello, che definisce le relazioni tra ogni attività (E7) e qualsiasi entità (E1) che possa avere avuto una influenza su di essa. La classe E1 comprende tutto ciò che ricade nell’universo del discorso di CIDOC-CRM. Di fondamentale importanza, inoltre, la proprietà “*P67 refers to (is referred to by)*. Domain: E89 Propositional Object; Range: E1 CRM Entity”, che mette in evidenza la relazione di *aboutness* che si instaura tra una instance immateriale appartenente alla classe E89 *Propositional Object*, cioè proposizioni circa oggetti reali o immaginari che costituiscono argomenti di discorso, e qualsiasi entità (E1).

Il principio guida per l’integrazione delle informazioni provenienti da fonti diverse è la modellazione esplicita degli eventi: *Temporal entities*, cioè i periodi temporali, le condizioni particolari nel tempo, assumono quindi un ruolo centrale nello schema. L’obiettivo è quello di integrare le informazioni da fonti diverse in un contesto storico che i singoli oggetti del patrimonio culturale illustrano e dal quale deriva la loro rilevanza, facendo emergere i collegamenti con le persone, i luoghi, le utilizzazioni con cui essi sono stati in contatto.

L’informazione viene considerata evento-centrica: la rappresentazione degli “eventi” permette di collegare i fatti, le idee, le persone, i luoghi, gli oggetti in una rappresentazione coerente della storia. Il passato viene analizzato sulla base della considerazione dell’evoluzione storica come una serie di “eventi discreti” nello spazio e nel tempo, che riguardano *Persistent items (Endurants)*, come gli oggetti e le persone, coinvolti contemporaneamente, e le entità concettuali, mentre gli eventi in sé sono considerati *Temporal entities (Perdurants)*. Solo questi ultimi sono direttamente connessi allo spazio e al tempo nell’ontologia.



Gli *Actors* (classe E39, tutti coloro che agiscono, persone o enti, sottoclasse di E77 *Persistent item*) sono correlati agli oggetti fisici e agli oggetti concettuali attraverso le *Temporal entities*. Le entità sono denominate attraverso identificatori (*Appellations*).

Altre caratteristiche peculiari dell'ontologia CIDOC-CRM sono: la *Multiple instantiation* (istanziamento molteplice e non-esclusività di classi e proprietà), per cui una *instance* (un individuo) di una Classe può essere allo stesso tempo *instance* di altre classi, ed essere descritta attraverso le proprietà di tutte classi cui appartiene: un oggetto particolare può essere sia una *instance* di E20 *Biological Object* che di E22 *Man-Made Object*; la *Multiple inheritance* (ereditarietà multipla), in base alla quale una classe può discendere da più di una superclasse ed ereditare, quindi, le Proprietà di diverse superclassi: ad es., *Person* è sottoclasse sia di *Actor* che di *Biological object*, e quindi eredita le proprietà di entrambe.

Nell'ontologia CIDOC-CRM le classi che accolgono la terminologia specifica sono state separate dallo schema concettuale. La proprietà "P2 *has type (is type of)*", che prevede come *Domain* E1 CRM Entity, cioè il complesso delle entità che ricadono nel settore coperto dalla definizione di patrimonio culturale data da ICOM, e come *Range* tutti i valori previsti dalla classe E55 Type, è ereditata da ciascuna classe di CIDOC-CRM (eccetto E59 *Primitive values*). Questo crea un meccanismo generale che permette la specializzazione della classificazione delle *instances* di CRM a qualsiasi livello di dettaglio, attraverso un link a vocabolari controllati esterni, a thesauri, a schemi di classificazione o ontologie di dominio, che forniscono di volta in volta i termini specifici. Le istanze di E55 possono essere collegate alla proprietà "P127 *has broader term (has narrower term)*" e formano gerarchie di termini (ICOM/CIDOC 2015, XVI).



Lo standard *Simple Knowledge Organization System* (SKOS),²³ realizzato da Semantic Web Deployment Working Group (W3C), offre un modello per esprimere i contenuti presenti in thesauri, sistemi di classificazione, liste di intestazioni per soggetti, come dati leggibili dalle macchine e permette di trasferire i sistemi elaborati dalla Library Science (KOS) nel contesto delle tecnologie del Semantic Web e di renderli condivisibili e riutilizzabili. I concetti e gli schemi di concetti sono identificabili tramite URI, e questo permette a chiunque di riferirsi ad essi in modo non ambiguo.

Breve rassegna sull'utilizzazione di CIDOC-CRM

La rassegna che segue non ha pretesa di esaustività, ma si offre come un primo ragguaglio sull'impiego di CIDOC-CRM.

Una delle più grandi collezioni di oggetti del patrimonio artistico gestita attraverso CIDOC è Classical Art Research Online Research Services (CLAROS),²⁴ servizio coordinato dall'Università di Oxford, che raggruppa partner di diverse nazioni, musei e istituti di ricerca di Atene, di Parigi, di Roma, di Heidelberg, di Colonia ecc. CLAROS ha realizzato un'infrastruttura per la consultazione e la ricerca simultanea di collezioni d'arte antica greca e romana, sculture, ceramiche, camei, fotografie, sarcofagi, statue, pitture, monete ecc. I record prodotti dalle fonti federate, ciascuna con i propri database realizzati con l'impiego di tecnologie e schemi diversi, dal 2007 vengono raccolti e resi interoperabili utilizzando le tecnologie del Semantic Web. Dal 2010 sono stati resi interoperabili e simultaneamente consultabili attraverso un'interfaccia unica,

²³ <http://www.w3.org/TR/2009/REC-skos-reference-20090818>

²⁴ <http://clarosnet.org/XDB/ASP/clarosHome>



utilizzando CIDOC-CRM, *dataset* che si trovano geograficamente distribuiti, mettendo a disposizione, anche come *open data*, venti milioni di record. I partner decidono quali porzioni dei propri record rendere accessibili e aperti, e come mappare ogni componente con uno dei termini di CIDOC. Il database realizzato genera asserzioni in RDF nel formato XML; i dati sono poi linkati con gli altri dati del Semantic Web utilizzando la collezione di dati aperti *Geonames*.²⁵ Per una porzione degli oggetti, in particolare sculture e vasi greci, è possibile attivare la ricerca per immagini.

A partire dal 2007, il sistema *Collection online* del British Museum²⁶ ha permesso agli utenti remoti di fruire di tre milioni e mezzo di oggetti digitali delle collezioni del British Museum. CIDOC è stato utilizzato per organizzare i metadati per la ricerca. Dal 2011 i dati sono esposti anche in RDF come dati aperti, organizzati tramite CIDOC-CRM, e interrogabili grazie a un *endpoint* SPARQL,²⁷ il servizio per *Linked Data* realizzato dal British Museum e Andrew W. Mellon Foundation nell'ambito del progetto Research Space,²⁸ basato sul *mapping* tra i dati del British Museum e l'ontologia CIDOC-CRM (Oldman e CRM Labs 2014). Research Space è un progetto volto ad offrire un ambiente collaborativo per la ricerca nel settore delle Scienze umane e del patrimonio culturale, usando le tecnologie del Semantic Web e CIDOC-CRM per integrare dati eterogenei. È un sistema basato

²⁵ <http://www.geonames.org>

²⁶ http://www.britishmuseum.org/research/collection_online/search.aspx

²⁷ <http://collection.britishmuseum.org>. Sito consultato l'ultima volta il 22 aprile 2016. Attualmente (giugno-luglio 2016), l'accesso al sito è discontinuo. Il dataset British Museum Collection sarà offerto su Datahub <https://datahub.io/dataset/british-museum-collection>

²⁸ <http://www.researchspace.org>



su una interfaccia del tutto innovativa, che permette di ricercare oggetti, persone, luoghi, eventi, periodi temporali e concetti, basandosi sulle relazioni esistenti tra questi, e offre un contesto coordinato alla ricerca, rispettando le peculiarità dei singoli contenuti. Un sistema open source per il *mapping* dei dati con l'ontologia CIDOC, *3M Mapping Memory Manager*,²⁹ realizzato da FORTH (Grecia) e Delving BV (Olanda), rende possibile per chiunque mappare i dati provenienti da fonti diverse con le relazioni semanticamente ricche e le entità di CIDOC, e creare *Linked data* qualitativamente migliori.

Altri progetti nell'ambito degli studi sul patrimonio culturale che stanno impiegando l'ontologia CIDOC-CRM sono elencati sul sito di ResearchSpace. Oltre a CLAROS sono citati *Arches project*,³⁰ progetto realizzato attraverso la collaborazione tra Getty Conservation Institute (GCI) e World Monuments Fund (WMF) con l'obiettivo di realizzare un sistema informativo basato su Web e con riferimenti geospaziali per il patrimonio culturale immobile, come siti archeologici e paesaggi; il Museum of New Zealand³¹; il Russian Linked Culture Cloud,³² progetto di *open data* che riguarda il patrimonio culturale russo.

L'aggregazione di contenuti culturali digitali attraverso la raccolta dei metadati prodotti da istituzioni eterogenee, che hanno descritto i propri oggetti utilizzando schemi spesso differenti, e la possibilità di consultarli attraverso un unico punto d'accesso, è stato l'obiettivo di due progetti europei, coordinati dall'Istituto

²⁹ <http://www.researchspace.org/home/mapping>

³⁰ http://www.getty.edu/conservation/our_projects/field_projects/arches

³¹ <http://collections.tepapa.govt.nz>

³² <http://culturecloud.ru/resource/rm-lod:Start>



centrale per il catalogo unico ICCU³³ di MIBACT, per la fornitura dei contenuti digitali a Europeana.³⁴ Dei due progetti, Athena³⁵ e Linked Heritage,³⁶ il primo, sviluppato nell'ambito del programma *eContentPlus*, ha avuto come obiettivo il sostentamento della partecipazione di più di cento musei e istituzioni culturali al portale Europeana. Tra il 2008 e il 2011 Athena ha svolto infatti la funzione di aggregazione di più di quattro milioni di contenuti digitali presenti nei musei di ventitré stati membri dell'UE (Piccininno e Vassallo 2013).

Nell'ambito del progetto è stata realizzata la piattaforma *web-based* MINT,³⁷ per mappare i metadati e facilitare l'aggregazione dei contenuti per Europeana. MINT ha utilizzato lo schema per l'*harvesting* dei metadati *Lightweight information describing objects*, LIDO (McKenna, Rohde-Enslin, Stein 2011),³⁸ rilasciato nel 2010 per agevolare la distribuzione in un formato standard delle informazioni relative agli oggetti museali, la loro aggregazione nei *repositories* e nei portali culturali e la connessione dei dati nel Web, e viene impiegato in particolare per aggregare i dati museali e inviarli ad Europeana. LIDO è basato su CIDOC-CRM e presenta 14 gruppi di informazioni, tre dei quali sono obbligatori: *Object-Work Type* (la tipologia dell'oggetto conservato in un

³³ <http://www.iccu.sbn.it/opencms/opencms/it>

³⁴ <http://www.europeana.eu/portal>

³⁵ <http://www.athenaeurope.org/index.php?it/1/home>

³⁶ <http://www.linkedheritage.eu>

³⁷ La piattaforma MINT oggi viene utilizzata anche per i progetti di Europeana Photography, Europeana Fashion, AthenaPlus, LoCloud, EUscreenXL, Europeana Sounds, Europeana Connect, Indicate, Europeana Awareness, Europeana Creative, Ambrosia e Europeana Space.

³⁸ Una presentazione di LIDO corredata della definizione nello schema XML e del documento con le specifiche, si trova sul sito di ICOM, all'URL: <http://network.icom.museum/cidoc/working-groups/lido/what-is-lido>



museo), *Title-Name* (Identificazione) e *Record* (Metadati amministrativi). Se gli oggetti museali non presentano un riferimento ad un autore o ad una data di creazione, LIDO, grazie al fatto che è fondato su CIDOC-CRM, permette di collegarli ad un evento temporale, come il momento della loro creazione, del loro ritrovamento, o della loro utilizzazione, cioè all'evento al quale hanno partecipato attori, che si è determinato in un luogo definito e in un momento preciso (Stein 2011; McKenna, Rohde-Enslin, e Stein 2011). Offre la possibilità di presentare i dati in diverse lingue, ed è utilizzabile per un ampio spettro di oggetti del patrimonio culturale, da quelli artistici ed architettonici a quelli di storia naturale. Diversi profili di LIDO: per l'Archeologia, l'Architettura, le Arti (tra cui la fotografia e la grafica), la Storia culturale, l'Etnologia, i Documenti scritti, la Storia naturale, la Botanica, la Paleontologia e la Mineralogia.

Tra i *dataset* che MIBACT ha reso fruibili come dati aperti, la sezione del portale CulturaItalia dedicata alla esposizione dei Linked Open Data³⁹ raggruppa dati aperti rilasciati dai diversi partner a CulturaItalia, con licenza CC0 per il pubblico dominio per quanto riguarda i *dataset* del Ministero, e altre licenze CC per quanto riguarda gli altri partner. I dati sono accessibili sia tramite un OAI provider, in RDF e XML, sia attraverso uno *endpoint* SPARQL, in RDF. Nel caso dell'OAI provider, i dati sono stati strutturati adottando gli schemi OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), PICO (applicativo di CulturaItalia), Europeana Data Model e CIDOC-CRM, non tutti, però, visualizzabili. I dati fruibili attraverso l'*endpoint* SPARQL sono stati strutturati utilizzando CIDOC-CRM nella versione Erlangen CRM/OWL. I record delle risorse culturali

³⁹ <http://dati.culturaitalia.it>



digitali strutturate secondo il PICO *Application Profile*, adottato dal portale CulturaItalia, vengono trasformati in record strutturati in base a CIDOC-CRM attraverso un'attività di *mapping*.⁴⁰

Infine, è necessario segnalare l'importante attività che ha riguardato l'ontologia CIDOC-CRM, intrapresa a partire dal 2003, con l'obiettivo di sviluppare un modello unico di descrizione per i settori museale e bibliografico, armonizzandone le peculiarità. Il modello FRBRoo (FRBR object-oriented 2012)⁴¹ che ne è derivato, e che si affianca alla famiglia di FRBR, al posto della originaria metodologia entità-relazioni, ha accolto quella *object-oriented*, ristrutturando le entità e le relazioni dell'originario FRBR come classi e proprietà di una ontologia formale. FRBRoo⁴² è una ontologia unica di riferimento per musei e biblioteche, in cui sono combinate le caratteristiche di FRBR e di CIDOC *Conceptual Reference Model*, e gli attributi dichiarati nel modello FRBR, cioè le caratteristiche presentate da ciascuna entità, sono stati tradotti in una o più proprietà di CIDOC CRM.

⁴⁰ http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/mapping/pico_cidoc/mapping_PICO_CIDOC-CRM_ITA-ENG.pdf, realizzato da Maria Emilia Masci della Scuola Normale Superiore di Pisa.

⁴¹ Per una introduzione, si può ricorrere a Le Boeuf (2012). L'argomento è troppo vasto e complesso per essere trattato brevemente in questo articolo, e gli dedicherò il dovuto approfondimento in un prossimo lavoro.

⁴² Rappresentazione grafica di FRBRoo: http://www.cidoc-crm.org/frbr_inro.html



Bibliografia

- Antoniou, Grigoris, e Frank van Harmelen. 2009. Web Ontology Language: OWL. In *Handbook on Ontologies*, a cura di Steffen Staab e Rudi Studer, 91–110. International Handbooks on Information Systems. Springer Berlin Heidelberg.
http://link.springer.com/chapter/10.1007/978-3-540-92673-3_4.
- Baader, Franz, Ian Horrocks, e Ulrike Sattler. 2009. Description Logics. In *Handbook on Ontologies*, a cura di Steffen Staab e Rudi Studer, 21–43. International Handbooks on Information Systems. Springer Berlin Heidelberg.
http://link.springer.com/chapter/10.1007/978-3-540-92673-3_1.
- Baader, Franz, e Werner Nutt. 2003. Basic description logics. In *The description logic handbook: theory, implementation, and applications*, a cura di Franz Baader, Diego Calvanese, e Deborah L. McGuinness, 47–100. Cambridge: Cambridge University Press.
- Biagetti, Maria Teresa. 2010. *Le Ontologie*. Numero speciale monografico di AIDAinformazioni. Vol. 28. 1-2. AIDAinformazioni. Roma: Associazione italiana per la documentazione avanzata.
- Calvanese, Diego, Pietro Liuzzo, Alessandro Mosca, José Remesal, Martin Rezk, e Guillem Rull. 2016. Ontology-Based Data Integration in EPNet: Production and Distribution of Food during the Roman Empire. *Engineering Applications of Artificial Intelligence* 51 (maggio): 212–29. doi:10.1016/j.engappai.2016.01.005.



- Calvanese, Diego, Alessandro Mosca, Jose Remesal, Martin Rezk, e Guillem Rull. 2015. A 'historical case' of Ontology-Based Data Access. In *Proc. of Digital Heritage 2015 (DH 2015)*, 291–98. IEEE.
doi:10.1109/DigitalHeritage.2015.7419510.
- Civili, Cristina, Marco Console, Giuseppe De Giacomo, Domenico Lembo, e Maurizio Lenzerini. 2013. MASTRO STUDIO: Managing Ontology Based Data Access applications. In *Proceedings of the VLDB Endowment*, 6.12:1314–17.
- Crofts, Nick, Martin Doerr, e Tony Gill. 2003. The CIDOC Conceptual Reference Model: A standard for communicating cultural contents. *Cultivate Interactive*.
http://www.cidoc-crm.org/docs/martin_a_2003_comm_cul_cont.htm.
- Doerr, Martin. 2003. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24 (3): 75–92.
doi:10.1609/aimag.v24i3.1720.
- . 2009. Ontologies for Cultural Heritage. In *Handbook on Ontologies*, a cura di Steffen Staab e Rudi Studer, 463–86. International Handbooks on Information Systems. Springer Berlin Heidelberg.
http://link.springer.com/chapter/10.1007/978-3-540-92673-3_21.
- Doerr, Martin, Jane Hunter, e Carl Lagoze. 2003. Towards a Core Ontology for Information Integration. *Journal of Digital Information* 4 (1).
<https://journals.tdl.org/jodi/index.php/jodi/article/view/92>.



- Doerr, Martin, Christian-Emil Ore, e Stephen Stead. 2007. The CIDOC Conceptual Reference Model: A New Standard for Knowledge Sharing. In *Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual Modeling - Volume 83*, 51–56. ER '07. Darlinghurst, Australia, Australia: Australian Computer Society, Inc.
<http://dl.acm.org/citation.cfm?id=1386957.1386963>.
- Gaio, Silvia, Stefano Borgo, Claudio Masolo, Alessandro Oltramari, e Nicola Guarino. 2010. Un'introduzione all'ontologia DOLCE. A cura di Maria Teresa Biagetti. *Le ontologie, AIDAinformazioni*, 28 (1–2): 107–25.
- Guarino, Nicola. 1998. Formal Ontology and Information Systems. In *Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*, Versione online corretta, 3–15. Amsterdam: IOS Press.
<http://www.lirmm.fr/~mugnier/DEA/guarino98formal.pdf>.
- ICOM/CIDOC. 2015. Definition of the CIDOC Conceptual Reference Model. Version 6.2.1. A cura di Patrick Le Boeuf, Martin Doerr, Christian-Emil Ore, e Stephen Stead. ICOM/CIDOC CRM Special Interest Group.
http://www.cidoc-crm.org/docs/cidoc_crm_version_6.2.1.pdf.
- Johnson-Laird, P. N., Sangeet S. Khemlani, e Geoffrey P. Goodwin. 2015. Logic, Probability, and Human Reasoning. *Trends in Cognitive Sciences* 19 (4): 201–14.
doi:10.1016/j.tics.2015.02.006.
- Le Boeuf, Patrick. 2012. An example of library and museum cooperation: the FRBROO conceptual model. presentato al CIDOC Annual Conference – Enriching Cultural Heritage, Helsinki, giugno.



- http://network.icom.museum/fileadmin/user_upload/minisites/cidoc/ConferencePapers/2012/le-boeuf-keynote.pdf.
- Lourdi, Irene, Christos Papatheodorou, e Martin Doerr. 2009. Semantic Integration of Collection Description: Combining CIDOC/CRM and Dublin Core Collections Application Profile. *D-Lib Magazine* 15 (7/8). doi:10.1045/july2009-papatheodorou.
- McKenna, Gordon, Stefan Rohde-Enslin, e Regine Stein. 2011. Lightweight Information Describing Objects (LIDO): the International Harvesting Standard for Museums. Repro Stampa Ind, Grafica. <http://obs-traffic.museum/lightweight-information-describing-objects-lido-international-harvesting-standard-museums>.
- Mylopoulos, John, Alex Borgida, Matthias Jarke, e Manolis Koubarakis. 1990. Telos: Representing Knowledge About Information Systems. *ACM Trans. Inf. Syst.* 8 (4): 325–362. doi:10.1145/102675.102676.
- Nardi, Daniele, e Ronald Brachman. 2003. An Introduction to Description Logics. In *The Description Logic Handbook: Theory, Implementation, Applications*, a cura di Franz Baader, Diego Calvanese, e Deborah L. McGuinness, 1–40. Cambridge: Cambridge University Press.
- Oldman, Dominic, e CRM Labs. 2014. The CIDOC Conceptual Reference Model (CIDOC-CRM): PRIMER. Version 1.2. <http://www.researchspace.org/home/rsandcrm>.
- OWL 2 Web Ontology Language Primer: W3C Recommendation 27 October 2009. 2009. <https://www.w3.org/TR/2009/REC-owl2-primer-20091027/>.



- OWL Web Ontology Language Overview (W3C Recommendation 10 February 2004). 2004. <https://www.w3.org/TR/owl-features/>.
- Petersen, Toni, e Getty art history information program, a c. di. 1990. *Art & Architecture Thesaurus*. New York: Oxford University Press.
- Piccininno, Marzia, e Valentina Vassallo. 2013. Il flusso di lavoro nei progetti di aggregazione di contenuti culturali digitali. Buone pratiche e controllo della qualità. *DigItalia* 8 (1): 131–45.
- Pollock, John L. 1987. Defeasible reasoning. *Cognitive Science* 11 (4): 481–518.
- Staab, Steffen, e Rudi Studer. 2009. *Handbook on Ontologies*. 2nd ed. Berlin: Springer. <http://www.springer.com/us/book/9783540709992>.
- Stein, Regine. 2011. Introduction to LIDO 2011. presentato al Linked Heritage Training Workshop, Roma, settembre. <http://www.linkedheritage.eu/getFile.php?id=180>.
- Strasser, Christian, e G. Aldo Antonelli. 2015. Non-monotonic Logic. In *The Stanford Encyclopedia of Philosophy*, a cura di Edward N. Zalta, Fall 2015. <http://plato.stanford.edu/archives/fall2015/entries/logic-nonmonotonic/>.



BIAGETTI MARIA TERESA, Sapienza Università di Roma.
mariateresa.biagetti@uniroma1.it

Biagetti, M.T.. "An ontological model for the integration of cultural heritage information: CIDOC-CRM". JLIS.it. Vol. 7, n. 3 (September 2016): Art. #11930. DOI: 10.4403/jlis.it-11930.

ABSTRACT: The article aims at presenting the formal ontology CIDOC-Conceptual Reference Model, used for allowing information and documentation exchange and integration into the diverse sources of cultural heritage, and which is the leading ontology in the field since years. The ontology logical foundations are examined, entering into the details of the Description Logics, and also the theoretical configuration underlying the schema modelling. The relation (properties) structure - which is the ontology core - is presented along with the mechanisms proposed for the integration of cultural heritage information. In the end, this essays offers a brief and non-exhaustive overview of the information systems actually adopting the CIDOC-CRM ontology, from CLAROS to CulturaItalia.

KEYWORDS: Ontologies; Cultural Heritage; Semantic Web; CIDOC-CRM; Content integration.



Date submitted: 2016-06-10

Date accepted: 2016-07-09

Date published: 2016-09-15