

Gli archivi del MiBACT

L'integrazione dei dati archeologici digitali

Sara Di Giorgio
sara.digiorgio@beniculturali.it

¹ Istituto Centrale per il Catalogo Unico delle biblioteche italiane, Roma, Italia

Abstract.

The Central Institute for the Union Catalogue of Italian Libraries and Bibliographic Information is an Institute of the Ministry of Cultural Heritage and Activities and Tourism (MiBACT) that manages the National Library Service, the union catalogue of over 5,800 Italian libraries. It coordinates on behalf of MiBACT major digital cultural heritage projects at national level such as Internet Culturale, the portal of the digital collections of the Italian libraries, and CulturaItalia, the national cross-domain aggregator.

At European level the Institute coordinates and participates to top level European initiatives in the digital cultural heritage sector, such as projects contributing to Europeana and to other research infrastructures like ARIADNE, DARIAH and PARTHENOS.

Keywords: National aggregator, interoperability, CIDOC-CRM, data integration.

1 Introduzione

L'Istituto Centrale per il Catalogo Unico delle biblioteche italiane (ICCU), partecipa al progetto ARIADNE, svolgendo il ruolo di fornitore di contenuti tramite il portale CulturaItalia, l'aggregatore italiano dei contenuti digitali realizzati da archivi, musei e centri di documentazione e ricerca. L'ICCU, in particolare, normalizza i dati, rendendoli interoperabili con alcune delle principali iniziative europee di digitalizzazione, come ARIADNE ed Europeana.

L'Istituto centrale, nell'ambito della strategia complessiva per lo sviluppo dei sistemi della conoscenza e del posizionamento dell'Italia in campo europeo, avviata già da alcuni anni dal MiBACT, attraverso i vari progetti MINERVA [1], tra il 2002 e il 2008, ha maturato una grande esperienza nell'ambito del coordinamento delle politiche e dei programmi di digitalizzazione a livello comunitario, assumendo un ruolo di primo piano nella realizzazione di un network al quale partecipano numerosi ministeri della cultura e centinaia di istituzioni culturali del vecchio continente, tra cui biblioteche, archivi, musei e centri di ricerca.

L'ICCU, impegnato nel dibattito europeo per la digitalizzazione e l'accesso in rete al patrimonio culturale, ha avviato, in particolare, due linee di programmazione

complementari: da un lato facendosi promotore di una digitalizzazione ‘diffusa’, finalizzata a favorire la conoscenza e lo sviluppo delle industrie creative e del riutilizzo delle risorse digitali, intesi come volano di nuova produttività, dall’altro elaborando l’analisi dei servizi e delle opportunità offerte dalle *e-infrastructure* per lo sviluppo di nuovi servizi per la comunità dei ricercatori che lavorano alla valorizzazione e alla tutela dei beni culturali.

Queste esperienze hanno posto delle solide basi per sviluppare la cooperazione tra il settore del patrimonio culturale, le *research infrastructures* e le infrastrutture digitali, le *e-infrastructure* ipotizzando un piano d’azione comune per la ricerca e lo sviluppo futuro.

Nella prima linea strategica – quella della digitalizzazione ‘diffusa’ - si inserisce l’attività dell’Istituto di gestione degli aggregatori nazionali di risorse culturali digitali: il Servizio Bibliotecario Nazionale - SBN, la rete delle biblioteche italiane, Internet Culturale, il portale delle biblioteche italiane (www.internetculturale.it) e CulturalItalia (www.culturalitalia.it), il portale italiano della cultura che integra anche le risorse di SBN e Internet Culturale. Tutti questi aggregatori si integrano poi nel contesto internazionale fornendo le proprie risorse a Europeana, la biblioteca digitale europea, attraverso CulturalItalia e grazie ad una serie di progetti: Linked Heritage, AthenaPlus, EuropeanaPhotography, Partage Plus, Europeana Collections 14-18, Europeana Sounds, Ambrosia.

Nella seconda linea di ricerca - quella sulle infrastrutture digitali - avviata nel 2009 per approfondire le possibilità offerte nel settore del patrimonio culturale e che si sviluppa in linea con la *Joint Programming Initiative for Cultural heritage* guidata da MiBACT e MIUR, si inseriscono invece alcune iniziative europee che hanno l’obiettivo di armonizzare i programmi di ricerca sul patrimonio culturale digitale usando le opportunità fornite dalle *e-Infrastructure: DCH-RP* (2012-2014).

Il dialogo tra le *e-Infrastructure* e gli istituti europei della cultura è stato inaugurato con il progetto DC-NET (2009-2012) e poi rilanciato con i progetti INDICATE (2010-2012), e DCH-RP che hanno indagato sugli aspetti politici e tecnici che riguardano la relazione tra il settore del patrimonio culturale e le *e-infrastructure*.

Nell’ambito del progetto ARIADNE, l’ICCU coordina il gruppo di lavoro italiano composto da esperti dell’Istituto Centrale del Catalogo e della Documentazione (ICCD) e della Soprintendenza Speciale per il Colosseo, il Museo Nazionale Romano e l’Area Archeologica di Roma (SSBAR) che ha contribuito alla definizione del *framework* semantico di ARIADNE e all’integrazione di banche dati di rilevanza nell’ambito dell’archeologia.

L’ICCU, come fornitore di contenuti, promuove l’apertura dei dati e la condivisione di strumenti terminologici e standard per lo sviluppo di un modello di conoscenza comune per l’integrazione degli archivi digitali archeologici in Europa. Il principale canale di invio dei dati è il portale della cultura italiana, CulturalItalia, promosso dal Ministero dei beni, delle attività culturali e del turismo e gestito dall’ICCU.

2 Il portale CulturaItalia

CulturaItalia [2] svolge il ruolo di aggregatore nazionale ed è il fornitore principale dei contenuti italiani di Europeana e di ARIADNE: tra i suoi obiettivi vi è quello di indirizzare gli utenti di internet, attraverso la consultazione di un Indice [3] che aggrega oltre 3 milioni di dati, verso risorse digitali spesso disperse nel web, il più delle volte non facilmente raggiungibili, e pertanto ignorate dagli stessi studiosi delle diverse discipline.

CulturaItalia è un'iniziativa condivisa con Regioni, Università e altri importanti istituti culturali italiani, sia pubblici che privati, ed è il principale provider di contenuti italiani per ARIADNE a cui ha inviato circa 41 mila 'risorse' archeologiche, provenienti da musei e centri di documentazione statali e regionali.

Il portale è, di fatto, una rete collaborativa di istituzioni pubbliche e private in continua crescita, che recepisce il dibattito europeo e internazionale sull'accesso online del patrimonio culturale per tutti i cittadini e lo trasmette alla rete di istituti culturali a cui è collegato, favorendo la partecipazione degli istituti culturali e di ricerca alle iniziative internazionali per la conoscenza e la ricerca nell'ambito del patrimonio culturale.

CulturaItalia presenta una sezione, denominata MuseiD-Italia dedicata ai musei e permette di ricercare e confrontare opere provenienti da collezioni di diversi istituti, trovare informazioni relative a mostre temporanee e permanenti in tutto il territorio italiano e allo stesso tempo di rintracciare le informazioni sulle strutture che ospitano, fisicamente, il bene culturale. Grazie a questa iniziativa è stato possibile svolgere un lavoro di recupero di risorse esistenti e procedere a nuove digitalizzazioni, dando impulso ad un settore dove risulta digitalizzato appena il 10% dell'intero patrimonio museale, monumentale e archeologico nazionale.

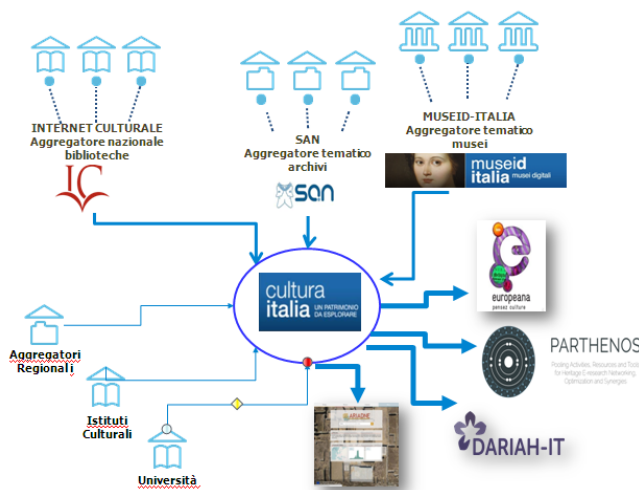


Fig. 1. Il workflow di aggregazione di CulturaItalia

I dati aggregati e rilasciati dai partner con la Licenza Creative Commons “CC0 1.0 Universal Public Domain Dedication” sono disponibili per il riuso nella sezione Dati.CulturaItalia, realizzata nel 2012, che offre uno SPARQL endpoint per la consultazione dei dati in formato Linked Open Data, e un OAI Provider, il repository conforme all’Open Archives Initiative [4], nota anche con l’acronimo OAI, che rende facilmente fruibili gli archivi che contengono i documenti prodotti. Nell’OAI Provider i dati sono accessibili, in formato XML e RDF, con diversi profili: Dublin Core, PICO Application Profile (il profilo applicativo di CulturaItalia in formato Dublin Core Qualified) e CIDOC-CRM nell’implementazione Erlangen CRM/OWL.

I LOD di CulturaItalia sono arricchiti semanticamente con gli authority file del VIAF (Virtual International Authority File: www.viaf.org), con i nomi geografici di GeoNames (www.geonames.org/), con il PICO Thesaurus [5] in SKOS e il DCMI Type vocabulary.

3 Il CIDOC-CRM per l’integrazione dei dati di CulturaItalia

La scelta del CIDOC-CRM [6], nella implementazione OWL Erlangen CRM [7], per l’integrazione dei dati, è stata fatta a seguito di un’attenta analisi dei diversi modelli concettuali disponibili ed è stato individuato il CIDOC in quanto modello concettuale di riferimento a livello internazionale per la codifica dei dati in RDF, in grado anche di garantire l’integrazione e lo scambio dei dati delle collezioni di musei e archivi con quelli provenienti dal settore bibliografico e codificati con il modello FRBRoo [8].

Il lavoro di mapping è stato affidato alla Scuola Normale Superiore di Pisa, che offre un supporto tecnico-scientifico allo staff tecnico del MiBACT per lo sviluppo del progetto CulturaItalia, fin dal suo nascere.

Il documento, elaborato nel 2012 definisce le regole per il mapping tra le risorse culturali digitali strutturate secondo il PICO Dublin Core Application Profile [9] (PICO AP), adottato per le risorse pubblicate nel Portale italiano della Cultura www.culturaitalia.it, ed il CIDOC-CRM. PICO AP è un profilo applicativo Dublin Core e come è già stato rilevato nella letteratura relativa al mapping tra Dublin Core e CIDOC CRM, per ogni valore dell’elemento DC “Type” (che specifica il tipo della risorsa descritta), occorre definire un mapping diverso, relativo ad una entità corrispondente del CIDOC-CRM [10].

Es.: IF DCMIType = Image, THEN la risorsa descritta deve essere mappata come:
CIDOC-CRM entity = E38_Image

Di conseguenza, ogni record codificato secondo il PICO AP produrrà una principale entità del CIDOC-CRM corrispondente, e il mapping di tutti gli altri elementi PICO AP che descrivono la risorsa dipenderà dal mapping di alto livello tra il tipo di risorsa e la corrispondente entità CIDOC-CRM.

Nel PICO AP, l’elemento dc:type è obbligatorio e ripetibile [11]. Un record PICO contiene infatti sempre almeno un valore dal DCMIType Vocabulary [12] (Collection, Dataset, Event, Image, InteractiveResource, MovingImage, PhysicalObject, Service,

Software, Sound, StillImage, Text) o dal PICOType Vocabulary [13] (CorporateBody, PhysicalPerson, Project).

Nel caso in cui un record PICO contenga più di un termine dai vocabolari DCMIType e/o PICOType, nel documento è specificato quale termine deve essere considerato come principale, cioè a quale termine si fa riferimento per il mapping. Il mapping presenta inoltre alcune simulazioni di casi complessi di mapping (tra un record PICO che contiene più di un termine DCMI/PICO Type e un elemento CIDOC-CRM corrispondente), formulati sulla base di alcuni casi reali reperiti tra le risorse di metadati di CulturalItalia.

Sulla base delle risorse digitali che al momento sono aggregate in CulturalItalia e del dominio del PICO AP, il termine DCMIType = Physical Object è mappato come: entità CIDOC-CRM = E22_Man_Made_Object (e non come: E19_Physical_Object). Inoltre quando DCMIType = Collection, generalmente il record PICO contiene molti altri termini DCMI/PICO Type. In tutti i casi in cui è presente il termine DCMIType “Collection”, la risorsa PICO sarà sempre mappata come: entità CIDOC-CRM = E78_Collection.

Sulla base del mapping di alto livello sono state identificate 12 entità CIDOC-CRM corrispondenti ai vari tipi possibili di record PICO AP.

Come già esplicitato, il mapping varia a seconda dell’entità CIDOC-CRM a cui corrisponde ciascun record PICO AP: infatti, con il variare dell’entità CRM definita dal mapping di alto livello, variano anche le proprietà e le entità CRM a cui corrispondono i singoli elementi del PICO AP che descrivono ogni risorsa.

Ad es.:

- Per una risorsa PICO con DCMIType= “PhysicalObject” (= crm:E22.Man_Made_Object), l’elemento PICO AP <pico:author> è mappato come segue:
 - E22.Man-Made_Object -> P108i.was_produced_by -> E12.Production
 - E12.Production -> P14.carried_out_by -> E39.Actor
 - E39.Actor -> P14.1.in_the_role_of -> E55.Type = author
- Per una risorsa PICO con DCMIType= “StillImage” (= crm:E38.Image), lo stesso elemento PICO AP <pico:author> viene mappato invece così:
 - E38.Image -> P94i.was_created_by -> E65.Creation_Event
 - E65.Creation_Event -> P14.carried_out_by -> E39.Actor
 - E39.Actor -> P14.1.in_the_role_of -> E55.Type = author

Poiché le entità CRM sono ordinate in poli-gerarchia e poiché le entità figlie in CRM ereditano le proprietà relative dalle entità madri, è possibile raggruppare le 12 entità identificate (e di conseguenza i mapping da implementare) in 4 gruppi principali, corrispondenti alle 4 tabelle di mapping nei paragrafi a seguire:

1. E18.Physical_Thing comprende i mapping di: E22.Man_Made_Object e E78.Collection
2. E28.Conceptual_Object comprende i mapping di: E73.Information_Object, E29.Design_or_Procedure,

- E33.Linguistic_Object, E36.Visual_Item, E38.Image
3. E39.Actor comprende i mapping di:
E40.Legal_Body, E21.Person
 4. E5.Event

Il documento di mapping, disponibile nella sezione di CulturaItalia dedicata alla documentazione tecnica [14], presenta le tabelle in cui sono indicati i vari elementi ed encoding del PICO AP: 1) DC terms; 2) PICO AP elements / refinements; 3) PICO AP encoding schemes mappati in CIDOC-CRM. Nel documento sono inserite a titolo esemplificativo delle indicazioni per l'implementazione e degli esempi di implementazione.

4 I dati archeologici inviati ad ARIADNE tramite CulturaItalia

L'integrazione dei dati messi a disposizione dai partner in ARIADNE avviene su due livelli: da una parte tramite le risorse digitali presentate nell'ARIADNE Registry, dall'altra con l'integrazione sperimentale di risorse di archivi di numismatica utilizzando la tecnologia dei LOD.

Il Registry, in particolare, descrive le risorse digitali come report di scavo, database di monumenti e siti archeologici, collezioni di immagini 2D e 3D ed è alimentato direttamente dai partner, che provvedono a fornire le informazioni necessarie. È perciò integrato nel portale ARIADNE [15] e offre a tutti i ricercatori la possibilità di accedere a una enorme quantità di informazioni sulle risorse archeologiche messe a disposizione, in tutta Europa, dai partner che costituiscono il Consorzio.

I dataset di archeologia aggregati in CulturaItalia e messi a disposizione di ARIADNE sono:

- MuseiD-Italia (109 collezioni digitali, collegate a 17.941 record)
- Fototeca della SSBAR, FOTOSAR (1.567 record)
- Direzione Regionale Campania (3.252 record)
- Direzione Regionale Campania/Museo archeologico Nazionale (1.489 record)
- Direzione Regionale Lombardia (1.388 record)
- Regione Umbria (1.624 record)
- Regione Marche (17.119 record)
- Regione Lombardia (7.150)
- Regione Calabria (2.232)
- Regione Emilia Romagna (4.157)

Per l'integrazione dei dati è stato fatto un mapping tra il profilo dei dati di CulturaItalia (PICO AP) e il profilo del Registry (ACDM). I dati pubblicati nell'OAI provider di CulturaItalia, sono stati perciò integrati e pubblicati nel portale ARIADNE.

L'Indice di CulturaItalia viene costantemente alimentato e nuovi contenuti provenienti da archivi di archeologia sono in via di acquisizione, come ad esempio quelli dell'Archivio di Documentazione Archeologica della Soprintendenza Speciale per il Colosseo, quelli del Museo di Archeologia Nazionale e dell'Area Archeologica

di Roma, che contiene i documenti prodotti dalla Soprintendenza a partire dal 1870, data della sua fondazione. Infine è in via di acquisizione il materiale scientifico preparatorio alla redazione della ‘Carta Archeologica’ di Roma, che raccoglie le informazioni dei ritrovamenti, risalendo almeno fino al Cinquecento ed un cospicuo nucleo di documentazione in copia tratto dagli archivi pubblici e privati e dalle collezioni di maggior interesse per lo studio della topografia di Roma antica. Una volta inglobati i dati, questi saranno resi disponibili nel Registry di ARIADNE.

Tramite CulturalItalia, è stato inoltre reso disponibile un sottoinsieme di 6.155 dati numismatica per testare l'integrazione a livello di elemento di set di dati archeologici tramite la tecnologia dei LOD. Sono state individuate in MuseiD-Italia le seguenti collezioni di monete provenienti da musei e centri italiani:

- Museo archeologico nazionale di Venezia
- Museo archeologico nazionale di Crotone
- Museo Archeologico Nazionale - Reggio di Calabria
- Museo Archeologico Nazionale di Altamura
- Collezione numismatica della Regione Umbria

I dati sono stati forniti nel formato CIDOC-CRM e integrati con quelli provenienti dagli archivi messi a disposizione degli altri partner per costituire un primo gruppo di dati standardizzati e collegati attraverso un modello concettuale in grado di catturare la complessità della conoscenza in ambito archeologico su cui è possibile eseguire *query* in chiave semantica.

5 Il contributo per lo sviluppo dell'integrazione semantica dei dati archeologici

L'ICCU ha costituito un gruppo di lavoro al quale partecipano esperti dell'Istituto Centrale del Catalogo e della Documentazione (ICCD) e del PIN, coordinatore di ARIADNE, per svolgere attività di mapping finalizzate a progettare l'evoluzione del CIDOC-CRM verso un'estensione specifica in grado di ‘catturare’ i concetti specifici nel campo archeologico. Infatti il sistema di schedatura dell'ICCD, utilizzata a livello nazionale per la catalogazione del patrimonio archeologico italiano, è risultato, a livello europeo, come il più completo e il più ‘efficace’ nel descrivere il mondo estremamente complesso e articolato dei beni archeologici.

Nell'ambito di questo gruppo di lavoro sono stati realizzati i seguenti documenti di mapping:

- Scheda reperto archeologico RA con il CIDOC-CRM [16]
- Scheda Monumento Archeologico/Complesso archeologico MA/CA con il CIDOC-CRM [17]
- ICCD_Thesaurus_RA (Thesaurus per la definizione dei beni archeologici) con AAT [18]

Il mapping con la scheda RA è stato particolarmente significativo per la relazione

nel modello CIDOC tra l'oggetto archeologico e il luogo di reperimento e i dati di scavo, mentre il mapping eseguito tra la scheda MA/CA e il CIDOC CRM ha contribuito a mettere in relazione il modello ontologico con i monumenti archeologici e le loro componenti.

Il mapping con il thesaurus 'definizione dell'oggetto' RA utilizzato nella catalogazione ICCD, con il Thesaurus AAT ha reso possibile l'integrazione multilingue dei dati italiani e la pubblicazione in formato SKOS del thesaurus.

BIBLIOGRAFIA

1. MINERVA (2002-2005), MINERVA Plus (2004-2006) e MINERVA eC (2006-2008); tutte le informazioni sono reperibili sul sito web www.minervaeurope.org
2. Portale CULTURAITALIA
<http://www.culturaitalia.it/opencms/index.jsp?language=it>
3. CULTURAITALIA
http://www.culturaitalia.it/opencms/ricerca.jsp?q=* &cat=indice&language=it
4. <http://www.openarchives.org/>
5. Il Thesaurus PICO
http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/thesaurus/4.3/thesaurus_4.3.0.skos.xml
6. CIDOC-CRM nella versione ufficiale del 2012: 5.0.4, Novembre 2011 e versione 5.1-draft, Novembre 2012
7. Erlangen CRM version: <http://erlangen-crm.org/current-version>
8. http://www.cidoc-crm.org/frbr_drafts.html
9. PICO Application Profile: <http://purl.org/pico/picoap1.0.xml>
10. Riferimenti maggiori utilizzati per il mapping: Doerr M., Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM, Technical Report 274, ICS-FORTH, July 2000: http://www.cidoc-crm.org/docs/dc_to_crm_mapping.pdf; Kakali K., Doerr M., Papatheodorou C., Stasinopoulou T., DC.type mapping to CIDOC/CRM, DELOS WP5-Task5.5, Department of Archives and Library Science / Ionian University, 26/01/2007: http://www.cidoc-crm.org/docs/WP5-T5_5-DC2CRMmapping-060728v0_2-final.doc ; I. Lourdi, C. Papatheodorou, M. Doerr, Semantic Integration of Collection Description. Combining CIDOC/CRM and Dublin Core Collections Application Profile, «D-Lib magazine», July/August 2009, vol. 15 n. 7/8, ISSN: 1082-9873: <http://www.dlib.org/dlib/july09/papatheodorou/07papatheodorou.html>; M. Doerr, Updated graphical representation of the harmonized EDM-CRM-FRBRoo-DC-ORE models, September 2011: http://www.cidoc-crm.org/docs/EDM-DC-ORE-CRM-FRBR_Integration_ORE_fix.ppt
11. PICO AP/dc:type:
<http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/picoap/picoap1.0.xml#type>
12. DCMI Type Vocabulary: <http://dublincore.org/documents/dcmi-type-vocabulary>
13. PICO Type Vocabulary:
<http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/picoap/picoap1.0.xml#PICOType>

14. Masci, M.E., Mapping tra PICO Application Profile e CIDOC Conceptual Reference Model, (2013)
http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/mapping/pico_cidoc/mapping_PICO_CIDOC-CRM_ITA-ENG.pdf
15. Portale ARIADNE <http://portal.ariadne-infrastructure.eu/>
16. Felicetti A., Scarselli T., Mancinelli M. L., Niccolucci F., Mapping ICCD Archaeological Data to CIDOC CRM: the RA Schema; il contributo, presentato al workshop CRMEX 2013. Practical Experiences with CIDOC CRM and its Extensions tenutosi a Malta il 26 settembre del 2013, è disponibile, insieme alla relativa presentazione, sul sito: <http://www.ontotext.com/CRMEX>.
17. Ronzino, P., Amico, N., Felicetti, A. & Niccolucci, F., 2013. European standards for the documentation of historic buildings and their relationship with CIDOC-CRM. In V. Alexiev, V. Ivanov, & M. Grinberg, eds. Practical Experiences with CIDOC CRM and its Extensions (CRMEX). Valletta, Malta, pp. 70–79.
Available at: <http://eurws.org/Vol-1117/paper7.pdf>
18. Felicetti A., Galluccio I., Luddi C., Mancinelli M.L., Scarselli T., Madonna A.D., Integrating terminological tools and semantic archaeological information: the ICCD RA Schema and Thesaurus, presentato a TPDL 2015