

DATI.CULTURAITALIA.IT, UN PROGETTO PILOTA DEDICATO AI DATI APERTI E AI LINKED OPEN DATA

Il progetto pilota dati.culturaitalia.it è stato avviato nel 2012 per costruire una sezione di CulturaItalia dedicata ai Dati aperti e fornire un servizio per i Linked Open Data (LOD), che renda progressivamente disponibili insiemi di dati aperti aggregati dal portale CulturaItalia (<http://www.culturaitalia.it/>). CulturaItalia è il portale della cultura italiana, gestito dall'Istituto Centrale per il Catalogo Unico delle biblioteche italiane (ICCU) del Ministero dei Beni e delle Attività Culturali e del Turismo e in quanto aggregatore nazionale è il fornitore privilegiato dei contenuti culturali italiani digitali a Europeana (<http://www.europeana.eu/>) il portale europeo che dà accesso al patrimonio culturale di musei, biblioteche e archivi di tutti i paesi membri. Una prima versione di questo servizio è disponibile online dal 2013 (<http://dati.culturaitalia.it/>) come una sezione, di CulturaItalia dedicata al LOD. Sono attualmente disponibili in formato LOD il Thesaurus PICO¹, adottato dal portale per facilitare la navigazione dei dati (metadati) aggregati e indicizzati e i dataset relativi a: Accademia di Santa Cecilia, progetto ArtPast, Digibess, Internet Culturale, Polo Museale Fiorentino, Regione Marche e Anagrafe delle Biblioteche Italiane. Tutti i dati sono disponibili con la licenza Creative Commons CC0 1.0 – Universal Public Domain Dedication. A questo primo insieme di dati ne verranno aggiunti di nuovi, man mano che i fornitori di contenuti di CulturaItalia adottano la licenza CC0, in accordo con la politica di diffusione dei dati raccomandata dall' ICCU e da Europeana.

Nella sezione dati.culturaitalia.it è possibile avere accesso ai dati tramite uno SPARQL endpoint e un provider OAI.

I metadati aggregati da CulturaItalia e codificati secondo lo schema XML PICO Dublin Core Application Profile (PICO AP)² sono stati mappati in due differenti schemi RDF:

- il Data Model Europeana (EDM)
- il CIDOC Conceptual Reference Model (CIDOC-CRM).

I documenti di mapping sono stati elaborati da Emilia Masci, della Scuola Normale Superiore di Pisa, che offre un supporto tecnico-scientifico al gruppo di lavoro dell'ICCU impegnato nello sviluppo di CulturaItalia, fin dalla sua progettazione e realizzazione.

¹ Thesaurus PICO 4.3 (http://purl.org/pico/thesaurus_4.3.0.skos.xml).

² Pico Application Profile (PICO-AP: <http://purl.org/pico/picoap1.0.xml>).



Fig. 1 – Home page della sezione dati.culturaitalia.it.

Il documento di mapping tra PICO AP e EDM³ è stato elaborato nel 2012 per codificare i dataset aggregati da CulturaItalia secondo il profilo EDM e renderli disponibili ad Europeana via OAI-PMH.

In aggiunta alla mappatura con EDM, il gruppo di lavoro tecnico e scientifico per il progetto CulturaItalia ha deciso di fornire una mappatura tra PICO AP e un modello concettuale di riferimento riconosciuto come standard ufficiale a livello internazionale, che possa essere codificato in RDF e che sia largamente usato per la documentazione di Beni Culturali, come alcuni modelli adottati da musei, biblioteche e archivi: dopo una analisi accurata è stato scelto CIDOC-CRM, poiché è sembrato il modello di riferimento più appropriato per questo scopo. Più in particolare, è stata scelta per il mapping l'implementazione OWL cosiddetta Erlangen CRM⁴. Le Triple RDF mappate in Erlangen CRM sono state poi arricchite con link a URI che identificano le istanze di risorse RDF riconosciute a livello internazionale per i nomi geografici

³ M.E. Masci 2012, *Definizione del mapping tra PICO AP e EDM – version 2.0* (http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/mapping/pico_edm/Mapping-PICO-EDM-2.0.pdf).

⁴ M.E. Masci 2013, *Mapping tra PICO AP e CIDOC-CRM – version 1.0* (http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/documenti/mapping/pico_cidoc/mapping_PICO_CIDOC-CRM_ITA-ENG.pdf).

OAI 2.0 Request Results

[Home](#) | [Identify](#) | [ListRecords](#) | [ListSets](#) | [ListMetadataFormats](#) | [ListIdentifiers](#)

You are viewing an HTML version of the XML OAI response. To see the underlying XML use your web browsers view source option: [page](#).

Datestamp of response 2013-02-28T13:21:05Z

Request URL http://194.242.241.172:8080/metaoaiicat/OAIHandler

Request was of type ListRecords.

OAI Record: oai:culturaitalia.it:ANSC00002595

OAI Record Header

OAI Identifier oai:culturaitalia.it:ANSC00002595 [pico](#) [edm](#) [cidoc_crm](#) [oai_dc](#) [formats](#)

Datestamp 2012-02-01T18:04:53Z

setSpec accademiaSCecelia [Identifiers](#) [Records](#)

PICO Metadata (pico)

```
<pico:record>
  <dc:identifier> ANSC.00001.00045.00005</dc:identifier>
  <dc:title xml:lang="it">Molis</dc:title>
  <dc:type xsi:type="dcterms:DCMIType">Sound</dc:type>
  <dc:type xml:lang="it">canto narrativo</dc:type>
  <dcterms:isPartOf xml:lang="it">name=raccolta,value=028, a cura di Sergio Liberovici</dcterms:isPartOf>
```

Fig. 2 – Record di esempio con la possibilità di visualizzare i vari formati PICO, EDM, CIDOC-CRM, OAI-DC.

e authority file per i nomi personali e aziendali, come GeoNames e il Virtual International Authority File (VIAF). L'endpoint SPARQL di dati.culturaitalia.it è stato sviluppato utilizzando il software open source Virtuoso (software OpenLink). Esso fornisce l'accesso a triple RDF strutturato secondo il formato SKOS, per il PICO Thesaurus e secondo CIDOC-CRM per i dataset rilasciati dai fornitori di contenuti di CulturaItalia con la licenza CC0. L'applicazione permette di recuperare i dati attraverso tre interfacce di interrogazione:

- ricerca testuale;
- query SPARQL;
- query iSPARQL.

La prima consente di effettuare ricerche a testo libero su tutte le triple contenute nel repository, la seconda permette di recuperare i dati tramite una query SPARQL e fornisce alcuni esempi di query, la terza è un'interfaccia di interrogazione ancora più complessa per utenti con un profilo tecnico avanzato.

La sezione dati.culturaitalia.it dà accesso inoltre al Provider OAI di CulturaItalia, sviluppato utilizzando il software Open Source OAICat. Esso contiene metadati strutturati secondo diversi schemi:

- OAI-DC (XML): schema OAI-PMH adottato dall’Open Archives Initiative Protocol for Metadata Harvesting;
- PICO (XML): PICO Application Profile, il profilo applicativo di CulturaItalia;
- EDM (RDF): European Data Model, adottato dal portale Europeana;
- CIDOC (RDF): CIDOC-Conceptual Reference Model nell’implementazione Erlangen CRM/OWL.

Nel prossimo futuro verranno pubblicati nuovi dataset messi a disposizione dei partner di CulturaItalia con la licenza CC0 e le triple RDF saranno linkate con URI esterne, a ulteriori applicazioni semantiche, Authority files e Vocabolari. Inoltre l’applicazione sarà migliorata con lo sviluppo di interfacce web innovative che saranno appositamente progettate per favorire il browsing delle risorse dagli utenti.

SARA DI GIORGIO

Istituto Centrale del Catalogo Unico delle biblioteche italiane
MiBACT

ABSTRACT

The pilot project “dati.culturaitalia.it” started in 2012 to build up a Linked Open Data (LOD) Service that will progressively make available open data-sets from the web-portal CulturaItalia, the Italian national aggregator for Europeana. The application, online since 2013 (<http://dati.culturaitalia.it/>), was designed to allow the resources aggregated by CulturaItalia to be involved into large semantic networks after exposing, sharing and connecting data according to LOD principles. CulturaItalia is the Portal of Italian Culture managed by the Istituto Centrale per il Catalogo Unico (ICCU), in which are involved cultural institutions – museums, archives and libraries – from national, regional and local levels. Metadata aggregated by CulturaItalia, encoded according to the XML schema PICO Dublin Core Application Profile (PICO AP), were mapped into two different RDF schemas: the European Data Model (EDM) and the CIDOC Conceptual Reference Model (CIDOC-CRM). In particular, the OWL implementation named Erlangen CRM was chosen for implementing the mapping. RDF triples mapped into Erlangen CRM were then enriched with links to URIs identifying instances of internationally established RDF resources for geographic names and authority files for personal and corporate names, such as GeoNames and Virtual International Authority File (VIAF).