

Digitalia

Anno VI, Numero 2 - **2011**

Rivista del digitale nei beni culturali

ICCU-ROMA

Linked archival authority data. Una sperimentazione sui *linked open data* nel sistema informativo regionale *IBC Archivi*

Silvia Mazzini - Francesca Ricci

Regesta.exe, IBC-Soprintendenza per i beni librari e documentari della Regione Emilia-Romagna

Durante l'*International Workshop on Semantic Digital Archives*, tenutosi a Berlino il 29 settembre scorso in occasione della quindicesima *International Conference on Theory and Practice of Digital Libraries (TPDL)*¹, l'Istituto per i beni artistici culturali e naturali della Regione Emilia-Romagna (IBC)² ha presentato i primi risultati di una sperimentazione di "apertura dei dati archivistici" realizzata grazie alla collaborazione di Regesta.exe³. In particolare, tale sperimentazione si è focalizzata sulle descrizioni dei soggetti (enti, persone e famiglie) produttori d'archivio, realizzate sulla piattaforma IBC-xDams⁴ in formato EAC-CPF (Encoded Archival Context-Corporate Bodies, Persons, and Families)⁵.

Anche in Italia si va consolidando l'esperienza nell'applicazione degli standard descrittivi archivistici e nell'utilizzo del web e delle ontologie come strumenti per rappresentare, veicolare, contestualizzare e integrare le descrizioni di risorse documentarie. È ormai usuale e abbastanza diffuso, inoltre, l'utilizzo di link come mezzo per collegare le descrizioni archivistiche sul web ad informazioni "altre", al fine di ampliare l'offerta informativa rivolta agli utenti che accedono agli strumenti di ricerca *on line*. In ambito archivistico gli interventi di inventariazione e descrizione producono sempre più spesso inventari elettronici, spesso consultabili *on line* e fruibili tramite ricerche testuali, navigazione nella struttura e, anche se meno di frequente, tramite l'accesso ai dati codificati in formato standard EAD o EAC-CPF. Obiettivo di EAC-CPF, così come dell'affine EAD (Encoded Archival Description)⁶

¹ <http://sda2011.dke-research.de/> (15.11.2011).

² <http://www.ibc.regione.emilia-romagna.it/> (15.11.2011).

³ <http://www.regesta.com/> (15.11.2011).

⁴ Piattaforma web based realizzata per IBC-Soprintendenza per i beni librari e documentari della Regione Emilia-Romagna da Regesta.exe per la schedatura archivistica EAD ed EAC compliant. Si veda a tal proposito B. Argelli, *Una regione in rete: gli archivi storici dell'Emilia-Romagna e il progetto IBC* in "DigItalia", 3 (2008) n. 2, nonché le informazioni pubblicate in *IBC Archivi*: <http://archivi.ibc.regione.emilia-romagna.it> (15.11.2011).

⁵ DTD per la marcatura in formato XML di record d'autorità archivistici conformi ad ISAAR (CPF): <http://eac.staatsbibliothek-berlin.de/> (15.11.2011).

Progetti

dedicato alla codifica di strumenti di corredo archivistici, è favorire l'interoperabilità tra sistemi informativi archivistici diversi, indipendentemente dai supporti hardware e software utilizzati⁷. Alla diffusione e diversificazione delle modalità di valorizzazione sul web, si è infatti accompagnato lo sviluppo di standard di varia natura: standard internazionali specializzati per la codifica delle diverse risorse, standard per l'interoperabilità e la condivisione dei dati, e "meta-standard" nazionali ed internazionali finalizzati al dialogo tra sistemi informativi diversi.

Anche l'esperienza di *IBC Archivi* (sistema informativo partecipato degli Archivi storici in Emilia-Romagna) si è sviluppata attraverso l'applicazione degli standard internazionali di descrizione archivistica e l'utilizzo dei diversi standard di codifica xml⁸. Nella prima metà del 2010, pertanto, a seguito della pubblicazione della versione definitiva dello standard EAC-CPF, IBC e Regesta.exe hanno realizzato una nuova scheda per la descrizione informatizzata dei soggetti produttori d'archivio puntualmente conforme allo standard ISAAR (CPF) e al nuovo formato⁹. La sperimentazione che qui si intende presentare, muovendo dalle schede descrittive prodotte sulla piattaforma IBC-xDams nel nuovo formato EAC-CPF (o migrate in esso dal precedente standard EAC), si è data come obiettivo la realizzazione di un'ontologia dello standard EAC-CPF e di un'ontologia della scheda descrittiva conforme allo standard, finalizzate alla pubblicazione in "formato" *Linked Open Data* (LOD)¹⁰ di tali descrizioni.

Com'è noto, il graduale passaggio dal web 2.0 al web semantico cui stiamo progressivamente assistendo è orientato alla produzione di contenuti per il web in grado di supportare la condivisione, la valorizzazione e la gestione delle informazioni archivistiche e delle relazioni tra queste intercorse (o intercorribili). *Linked Data*, web semantico e RDF sono le componenti tecnologiche su cui si fonda questo cambiamento, ma -pare evidente- non sono sufficienti a realizzarlo.

È dunque necessario che chi produce e pubblica dati sul web segua nuove strategie, utilizzando queste tecnologie non solo come mezzo per la produzione condivisa di contenuti (con gli strumenti partecipativi del web 2.0) o come mezzo di

⁶ DTD per la marcatura in formato XML di strumenti di corredo archivistici:

<http://www.loc.gov/ead/> (15.11.2011).

⁷ Nel maggio 2008 IBC ha ospitato i lavori del *EAC-CPF Working Group* in occasione del convegno "Standard e formati di scambio per l'interoperabilità dei sistemi archivistici". Per iniziativa di IBC-Soprintendenza per i beni librari e documentari e della Soprintendenza archivistica per l'Emilia Romagna verrà pubblicata la traduzione della *tag library* dello standard in lingua italiana.

⁸ EAG (Encoded Archival Guide) per la codifica delle descrizioni dei soggetti conservatori, EAD per i complessi documentari e EAC-CPF per la descrizione dei soggetti produttori.

⁹ Per ulteriori informazioni si consulti la pagina dedicata: [http://archivi.ibc.regione.emilia-](http://archivi.ibc.regione.emilia-romagna.it/ibc-)

cms/cms.find?flagfind=findnews&type=1&numDoc=147&munu_str=0_3_4_0#nogo (15.11.2011).

¹⁰ Per maggiori informazioni sui *Linked Open Data* (LOD) è possibile consultare il sito ufficiale:

<http://linkeddata.org/> o la pagina ad essi dedicata in Wikipedia:

http://it.wikipedia.org/wiki/Dati_collegati (15.11.2011).

Progetti

rappresentazione delle entità e delle relazioni tra le varie “componenti” della descrizione (ad esempio tramite le ontologie), ma soprattutto come strumento adeguato a qualificare semanticamente i dati e le loro relazioni aprendone i contenuti informativi all’uso di una comunità più vasta e diversificata e realizzando una vera e propria rete semantica aperta.

A partire da queste prime considerazioni, IBC ha scelto di avviare una sperimentazione sull’apertura dei dati archivistici prodotti attraverso gli interventi di riordino e inventariazione realizzati dall’Istituto sul territorio emiliano-romagnolo. A fronte dei diversi possibili campi di applicazione, si è scelto di focalizzare questa prima fase di sperimentazione sulle descrizioni dei soggetti produttori riconoscendo in essi le “entità” archivistiche che più significativamente si sarebbero adattate ad una sperimentazione di “*linked archival data*”. Già il capitolo sesto di ISAAR(CPF)¹¹, infatti, aveva individuato nella descrizione di tali soggetti (e con essi dei contesti di produzione della documentazione) sia una chiave primaria di accesso alla documentazione archivistica sia, contestualmente, un tramite privilegiato per la messa in relazione delle risorse archivistiche a tutte le risorse informative raccolte, prodotte, descritte o pubblicate dai medesimi soggetti o contesti. È stato per questo possibile immaginare, ad esempio, molteplici tipologie di relazioni tra diversi soggetti, tra soggetti e complessi archivistici prodotti, o tra soggetti produttori d’archivio e altri authority esistenti.

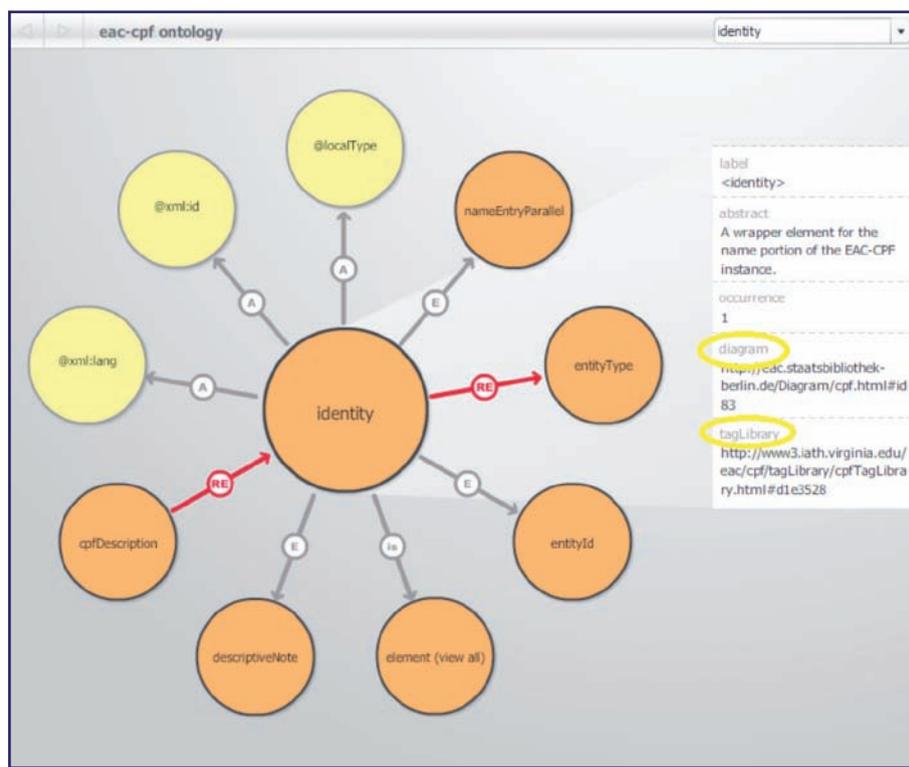
Il primo passo in questa direzione è stata l’“apertura” dello standard e la realizzazione e pubblicazione di un’ontologia per EAC-CPF in un formato RDF. La *tag library* dello standard ha una struttura abbastanza semplice, articolata in 90 elementi e 30 attributi. Sono state, quindi, analizzate le relazioni tra i singoli elementi e gli attributi ed è stata sostanzialmente prodotta una nuova rappresentazione dello standard, atta a fornire uno strumento alternativo di navigazione nella struttura dello standard, che ne mostrasse le relazioni e puntasse alle specifiche della *tag library* ufficiale e al *diagram* dello schema xml, per le specifiche tecniche di ciascun elemento. È stata cioè prodotta una formalizzazione (una “vista”) diversa del medesimo standard archivistico di cui si mostra di seguito una esemplificazione sotto forma di grafo¹².

Questo primo step, benché molto utile come primo approccio al problema e come interessante strumento di studio e approfondimento dello standard, non poteva essere utilizzato per “aprire” al mondo dei LOD le schede dei record di autorità codificati. È stato pertanto necessario realizzare una seconda ontologia

¹¹ Standard internazionale per i record d’autorità archivistici di enti, persone, famiglie: http://media.regesta.com/dm_0/ANAI/anaiCms/ANAI/000/0111/ANAI.000.0111.0001.pdf (15.11.2011).

¹² Il risultato è visibile in: <http://archivi.ibc.regione.emilia-romagna.it/ontology/EAC-CPF/genericEAC/RelationBrowserCIA.html> (15.11.2011).

Progetti



orientata non alla descrizione dello standard in RDF, ma alla definizione di un nuovo *data model* basato sullo schema stesso, ma semanticamente diverso, rivolto alla rappresentazione in formato RDF delle schede descrittive dei soggetti produttori pubblicate in *IBC Archivi*. In questa seconda ontologia, si è scelto di includere per quanto possibile, porzioni di altri standard, quali *Skos* (Simple Knowledge Organization System)¹³, *Foaf* (Friend of a Friend)¹⁴, *DC* (Dublin Core)¹⁵, *Bio* (bio-graphical ontology)¹⁶, *Viaf* (The Virtual International Authority File)¹⁷ e *Gn* (Geo-Names)¹⁸. Di seguito si mostra una rappresentazione di questa seconda ontologia prodotta¹⁹.

Per molti anni IBC ha sperimentato e promosso fattivamente l'utilizzo degli standard internazionali di descrizione archivistica e dei sistemi di codifica, per descrivere soggetti conservatori d'archivio, archivi storici e soggetti produttori presenti sul territorio regionale.

¹³ <http://www.w3.org/2004/02/skos/core#> (15.11.2011).

¹⁴ <http://xmlns.com/foaf/0.1/> (15.11.2011).

¹⁵ <http://purl.org/dc/elements/1.1/> e <http://purl.org/dc/terms> (15.11.2011).

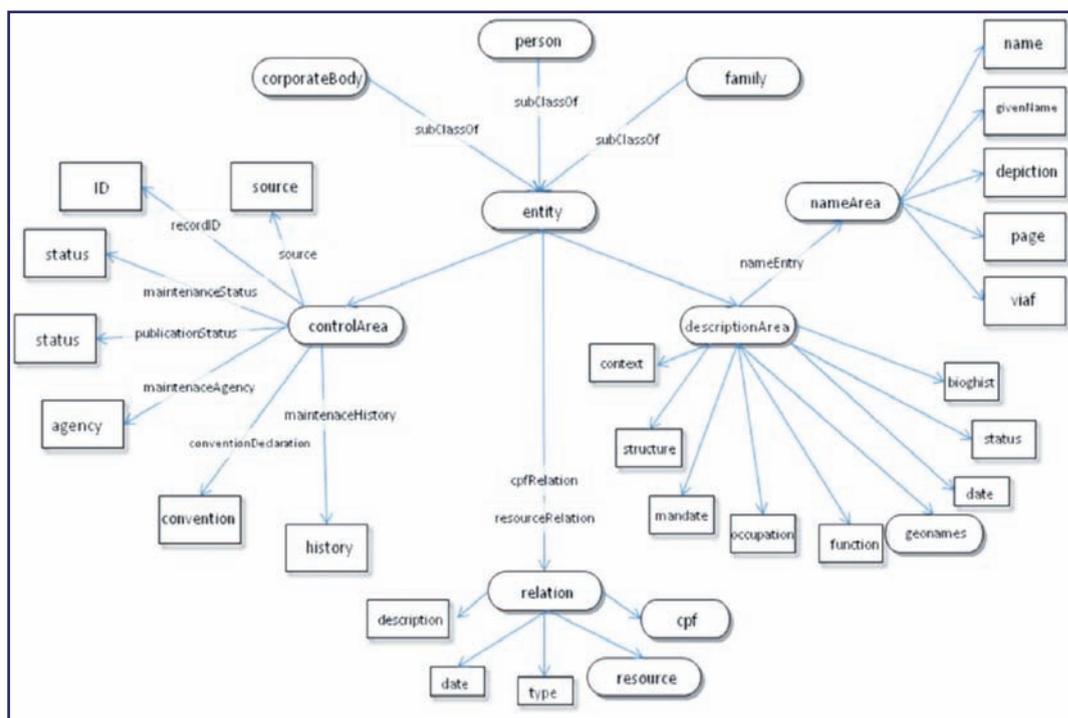
¹⁶ <http://purl.org/vocab/bio/0.1/> (15.11.2011).

¹⁷ <http://viaf.org/ontology/1.1/#> (15.11.2011).

¹⁸ <http://www.geonames.org/ontology#> (consultato il 15.11.2011).

¹⁹ <http://archivi.ibc.regione.emilia-romagna.it/ontology/EAC-CPF/eac-cpf.rdf> (15.11.2011).

Progetti



Attualmente il sistema informativo partecipato *IBC Archivi* pubblica le descrizioni di 389 istituzioni archivistiche, 2230 archivi storici e 185 soggetti produttori. Anche grazie a questa rete capillare di descrizioni sul territorio è parso possibile immaginare una rete di link in grado di mostrare, in modo automatico, dinamico e progressivamente sempre più articolato, risorse afferenti gli archivi (e non solo) tra loro collegate (o collegabili). Come accennato in precedenza, il campo di applicazione prescelto per questo progetto sono le schede descrittive dei soggetti produttori d'archivio realizzate nell'ambito del sistema informativo *IBC Archivi*. Le schede dei soggetti produttori d'archivio realizzate e conservate nella piattaforma di *back office* IBC-xDams (400 soggetti circa, tra enti, persone e famiglie), pertanto, hanno costituito il *test bed* del progetto. La prima sperimentazione è stata realizzata sui soggetti produttori delle Carte Andrea Costa (conservate presso l'Archivio storico del Comune di Imola) descritte in *IBC Archivi* tramite la piattaforma IBC-xDams²⁰.

Tali record, in particolare, si prestavano ad essere scelti come caso di studio perchè presentano una descrizione abbastanza analitica e numerose relazioni tra loro e con altri soggetti descritti, nonchè con diverse tipologie di risorse presenti in *IBC Archivi* e in altri sistemi informativi locali, nazionali e di ambiti

²⁰ <http://archivi.ibc.regione.emilia-romagna.it/ead-str/IT-ER-IBC-AS00209-0004118> (15.11.2011).

Progetti

con termini (bibliografici, museali, storiografici e culturali in genere). Tali dati sono stati inoltre messi in relazione all'ontologia *Geo-Names*.

I risultati di questo *test bed* e tutti i materiali prodotti nel corso della sperimentazione possono essere consultati nel "Semantic Environment for *IBC Archivi* Authority Records"²¹.

Resta ancora molto da fare sia dal punto di vista tecnico (in particolare su *query* e *reasoning*), sia dal punto di vista istituzionale (licenze di utilizzo, liberatoria da parte degli altri soggetti coinvolti, predisposizione di un ambiente per la pubblicazione dei dati, ecc.). Tuttavia l'esperienza condotta sulle due ontologie e la sperimentazione sui soggetti produttori delle Carte Andrea Costa dimostrano come i record di autorità possano effettivamente essere i primi dati archivistici da "aprire", in quanto intrinsecamente - e per loro stessa natura - potenziali punti di connessione tra risorse.

The screenshot displays the 'Linked Data API - IBC' interface for the record of Andrea Costa (1851-1910). The header identifies the institution as 'IBC Archivi' (Istituto per i Beni Artistici, Culturali e Naturali) and the project as 'Semantic Environment for IBC Archivi Authority Records'. The main content area shows a list of metadata fields:

- control**
 - maintenance status: new
 - publication status: approved
 - record ID: IT-ER-IBC-SP00001-0000054
- description**
 - bio birth: date 29 novembre 1851, luogo http://geonames.org/3175537
 - bio death: date 19 gennaio 1910, luogo http://geonames.org/3175537

On the right, there is a map of the Bologna region with a 'View' button below it. The interface also includes a search bar and various output format options (json, ttl, plaintext, rdf, xml, Mini, text).

²¹ <http://archivi.ibc.regione.emilia-romagna.it/ontology/semanticEnvironment/index.html> (15.11.2011).