

## I dati del SIGECweb

Laura Moro, Direttore dell'Istituto centrale per il catalogo e la documentazione

L'ICCD si è affacciato da poco tempo al mondo dei dati aperti; questo ci consente di poter beneficiare delle esperienze altrui, in una materia che ha ancora molti aspetti sperimentali, soprattutto quando applicata al patrimonio culturale.

I dati prodotti nel Sistema informativo generale del catalogo (SIGECweb) sono le informazioni che descrivono le caratteristiche storiche critiche e tecniche dei beni costituenti il patrimonio culturale. Sono dati quindi che descrivono beni materiali collocati in un determinato contesto geografico e culturale (il rapporto stretto con il bene materiale fa parte della nostra storia, è una condizione; per chi lavora al MiBACT il bene materiale continua ad essere la "risorsa", mentre ci troviamo a doverci confrontare ormai con un mondo in cui la risorsa digitale è più importante del bene stesso. Una limitazione che può essere vista come una condizione che porta ricchezza in un dibattito interdisciplinare). Questi dati sono prodotti da soggetti istituzionali diversi, in momenti storici diversi, e sono frutto della lettura del bene culturale fatta attraverso competenze tecniche specializzate; non siamo quindi in un sistema di produzione di dati di tipo partecipativo né in un'attività di meccanica registrazione di un dato di fatto. Tali dati sono il frutto di un'attività amministrativa che innesta su un'attività scientifica; si pongono quindi a metà strada tra l'adempimento normativo e la ricerca. Un esempio di questo doppio regime può essere trovato nei diversi modi di "chiamare" un luogo della cultura: contenitore fisico (Palazzo ducale di Urbino) o contenitore giuridico (Galleria nazionale delle Marche) o luogo di conservazione (Museo di Palazzo ducale) a seconda della prospettiva da cui viene indagato.

Si tratta comunque di dati primari, ossia il sistema del catalogo è la fonte "ufficiale" e autorevole dei dati descrittivi del patrimonio culturale (intendendo qui i beni storico artistici, esclusi quindi archivi e biblioteche)

Da una recente indagine fatta dall'ICCD, stimiamo che la consistenza dei beni catalogati in sistemi informatizzati sia di circa 9 milioni di unità appartenenti a soggetti pubblici e privati (va notato quindi che stiamo parlando della gestione di dati informativi di beni di cui non si ha la titolarità – questo è un tema: la titolarità di un dato che descrive un bene di cui non si è proprietari o detentori è una titolarità piena?). Di questi, circa tre milioni fanno riferimento alla banca dati del SIGECweb, gestita dall'ICCD, altri 3 milioni risiedono nella banca dati della CEI, che sta portando a compimento il censimento di tutti i beni culturali di proprietà delle diocesi e attuando sue autonome politiche di pubblicazione e valorizzazione del proprio patrimonio informativo; e il restante terzo è "disperso" nelle banche dati regionali e locali che pubblicano autonomamente tali dati e decidendo anche le politiche da adottare per l'apertura dei dati stessi.

Il Sigec web è un sistema “operazionale”, è finalizzato cioè alla produzione dei dati; assolve quindi a funzione di consultazione pubblica solo per gli utenti autorizzati. Per questo motivo l’ICCD ha predisposto una serie di strumenti per la messa a disposizione dei dati attraverso il web.

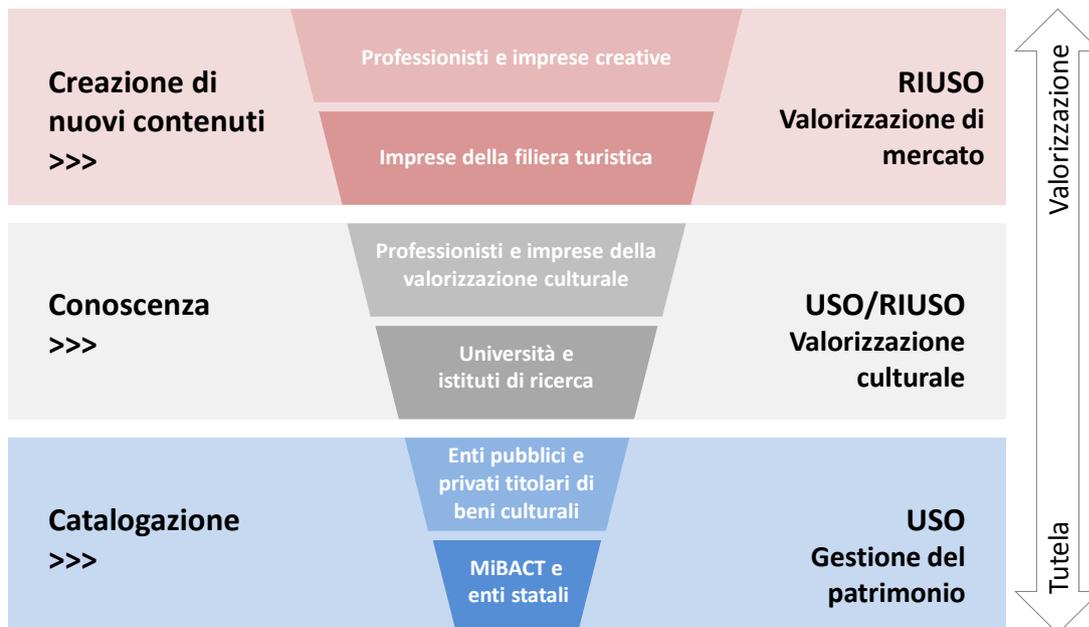
Quello che vi espongo ora quindi è il progetto *OPENiccd* che si attiverà progressivamente nei mesi a venire. Non si tratta solo di trovare uno standard per la trasformazione dei dati in open data e in linked open data. L’apertura dei dati non è un adempimento burocratico (anche se è un obbligo di legge), non si tratta di mettere in piedi dei “macina-caffè” che producono triple rdf in catena di montaggio. Si tratta di creare opportunità, questo è lo spirito degli open data, e quindi bisogna avere consapevolezza di cosa si sta facendo e per chi.

Per chi. E’ la prima cosa che ci siamo chiesti: chi sono gli utilizzatori dei dati che si pubblicano? La domanda è fino a un certo punto retorica, perché il patrimonio culturale ha un bacino di utenza enorme ed enormemente diversificato, andando dal semplice cittadino che guarda agli operatori commerciali che riusano.

Le categorie di utilizzatori sono state individuate distinguendo:

- chi produce la catalogazione e la utilizza rispetto a chi la utilizza soltanto;
- la finalità alla base dell’utilizzo della catalogazione: tutela e/o valorizzazione; uso/riuso
- per quanto riguarda la valorizzazione, la distinzione tra chi utilizza la catalogazione per scopi di mercato e chi per finalità non di mercato.

**OPENiccd** I gruppi di utenza potenziale dei dati del Catalogo dei beni culturali



Per ciascuna categoria di utenti abbiamo immaginato delle modalità specifiche per consentire l'accesso ai dati. Il progetto *OPENiccd* si articola quindi su un "pacchetto" di strumenti diversi per rendere fruibile e utilizzabile la banca dati del catalogo.

## OPENiccd

| piattaforma                               | accesso  | formato                           | licenza                   |
|---|--|-----------------------------------|---------------------------|
| <b>SIGECweb</b>                           | manuale<br>(utenti accreditati)                | XML<br>CSV                        | CC BY NC SA               |
| <b>Browsing web service</b>               | automatizzato<br>(utenti accreditati)          | SOAP (XML)                        | CC BY NC SA               |
| <b>Sito web pubblico</b>                  | manuale<br>(accesso libero)                    | PDF                               | CC BY NC SA               |
| <b>Piattaforma Dkan</b>                   | manuale e<br>automatizzato<br>(accesso libero) | CSV<br>XML<br>JSON                | CC 0<br>CC BY<br>CC BY SA |
| <b>OAI Provider<br/>(SPARQL endpoint)</b> | automatizzato<br>(accesso libero)              | XML<br>Pico/ Dublin Core<br>(RDF) | CC 0<br>CC BY<br>CC BY SA |

### SIGECweb ([www.sigecweb.beniculturali.it](http://www.sigecweb.beniculturali.it))

Il *Sistema Informativo Generale del catalogo* è il sistema operativo attraverso il quale viene gestito l'intero processo della catalogazione di beni culturali di proprietà pubblica e privata: permette, in tempo reale, la diffusione degli standard catalografici, gli aggiornamenti funzionali, l'immediata implementazione dei dati conoscitivi sul patrimonio culturale, la loro fruizione e condivisione con altri sistemi. Il sistema consente di predisporre l'ambiente di lavoro proprio di ciascun operatore, differenziando così le azioni che ognuno avrà a disposizione per effettuare le proprie attività nella gestione dei dati di catalogo.

Possono accedere alla banca dati del sistema gli utenti accreditati per finalità istituzionali e di ricerca. Gli utenti potenziali di questo sistema sono le amministrazioni pubbliche, gli istituti di ricerca e i soggetti singoli che svolgono attività di catalogazione

I dati sono resi disponibili in formato xml e csv e sono rilasciati secondo la licenza *Creative Commons Attribuzione-Non commerciale-Condividi allo stesso modo* (CC-BY-NC-SA).

**Browsing web service** (servizi in fase di test)

Al fine di consentire l'interoperabilità tra sistemi informativi sono stati predisposti dei servizi web per l'accesso automatizzato alle informazioni. In particolare il servizio di browsing consente, ad un sistema associato ad un ente registrato, di ricercare dati sulla BD del SIGECweb ed ottenere insiemi di informazioni coerenti con i livelli di visualizzazione dei dati connessi alle autorizzazioni assegnate all'Ente, ovvero al sistema interoperante, per i propri fini istituzionali. Il servizio è al momento in fase di sperimentazione.

Possono accedere al servizio i sistemi accreditati di istituzioni pubbliche o di pubblico interesse sulla base di specifici accordi. Per i documenti tecnici relativi ai servizi web rilasciati si rinvia all'indirizzo <http://www.iccd.beniculturali.it/index.php?it/394/interoperabilit>.

I dati sono resi disponibili in formato soap (xml) e sono rilasciati secondo la licenza *Creative Commons Attribuzione-Non commerciale-Condividi allo stesso modo* (CC-BY-NC-SA).

**Sito web pubblico** ([www.catalogo.beniculturali.it](http://www.catalogo.beniculturali.it))

Il *Catalogo generale dei beni culturali* è il sito pubblico che permette di ricercare informazioni sui beni culturali catalogati attingendo le informazioni in modo dinamico dal SIGECweb. Attualmente sono consultabili le schede con immagini relative a beni culturali di proprietà pubblica fino ad oggi conferiti al Sistema informativo generale del catalogo: monumenti, collezioni, raccolte, oggetti di interesse artistico o storico, reperti e siti archeologici, beni scientifici e naturalistici. Non include al momento tutti i dati conoscitivi prodotti e raccolti dai centri di documentazione regionali, che vengono gestiti e valorizzati anche a livello territoriale. Percorsi guidati permettono di effettuare ricerche, visualizzare anteprime e accedere alla scheda di catalogo del singolo bene di interesse.

L'accesso al sistema è libero e le informazioni sono a disposizione di tutte le tipologie di utenti che vogliono conoscere le principali caratteristiche dei beni pubblici catalogati (definizione e classificazione, descrizione, localizzazione, attributi storici, caratteristiche tecniche). L'esplorazione dei contenuti si effettua attraverso una navigazione a "faccette" sulle chiavi di ricerca "chi, cosa, dove"; strumenti di ricerca libera e guidata consentono di muoversi all'interno della banca dati.

Le schede di catalogo sono disponibili in formato pdf.

Poiché il sito contiene contenuti diversificati, anche redazionali, le informazioni vengono rilasciate con licenza *Creative Commons Attribuzione-Non commerciale-Condividi allo stesso modo* (CC-BY-NC-SA). In futuro, in relazione alle informazioni che saranno effettivamente presenti e agli utilizzi riscontrati, i contenuti potranno essere rilasciati con licenza meno restrittiva.

**Open data** (indirizzo da definire [www.catalogo.beniculturali.it/opendata](http://www.catalogo.beniculturali.it/opendata))

I dati aperti del Catalogo generale dei beni culturali saranno resi disponibili attraverso due strumenti:

- Un'applicazione, sviluppata sulla piattaforma open source Dkan, in corso di implementazione, che espone dataset riferiti a diverse tipologie di contenuti: dati di catalogo, statistiche della

catalogazione, soggetti produttori, vocabolari, schemi di tracciati schedografici standard. I dataset saranno via via resi disponibili, anche in base alle richieste dei fruitori che saranno monitorate al fine di incontrare le esigenze degli utenti.

- Un [Harvester OAI Provider](#), che gestisce le richieste secondo il protocollo OAI-PMH, sviluppato dalla Open Archives Initiative e utilizzato per il recupero (o harvesting) dei metadati dei record appartenenti ad un archivio. Le richieste di dati relativi alle sole schede di catalogo pubblicate possono essere gestite sulla base di diversi tipi di interrogazione. Attualmente i dati sono pubblicati secondo gli standard base [PICO](#) e [Dublin Core](#) in formato XML.

L'accesso al sistema è libero e le informazioni sono a disposizione degli utenti che vogliono riutilizzare i dati.

I dataset sono resi disponibili, per un accesso sia umano che automatizzato, nei formati xml, csv e json, con la licenza *Creative Commons citazione della fonte e Attribuzione-Condividi allo stesso modo* (CC-BY-SA). Alcuni dataset saranno messi a disposizione con licenze meno restrittive in relazione ai loro specifici contenuti.

Si aprono a questo punto del ragionamento una serie di questioni. In primo luogo è necessario porre il problema sulla legittimità di aprire i dati descrittivi relativi a beni dei quali non si è titolari; un conto infatti è consentire l'accesso alla banca dati da parte di utenti accreditati che utilizzano i dati per fini istituzionali, tutt'altro è l'accesso pubblico e il riuso di dati descrittivi di beni di proprietà privata. Anche negli interventi che mi hanno preceduto è stato specificato che la PSI si riferisce a beni che appartengono al dominio pubblico (come musei, archivi e biblioteche), da che sembrerebbe escludersi, almeno in prima istanza, il patrimonio privato.

Ragionando poi in termini di open data, non si può avviare una gestione condivisa del processo: quali sono le aspettative degli utenti e potenziali riutilizzatori? Quali sono i dati di maggiore interesse? Aggregati come?

Inoltre si pone la questione di come costruire l'ontologia sulla quale si mappano i dati per la creazione degli RDF: ha senso che il ministero crei una sua ontologia che possa descrivere al meglio le interconnessioni esistenti tra le varie categorie di beni culturali, come opere d'arte, architetture/contenitori, biblioteche, archivi, attività culturali? Oppure è meglio utilizzare standard già esistenti, come ha fatto ad esempio ICCU per alimentare il suo Sparql Endpoint? Ovviamente l'utilizzo di uno standard internazionale permetterebbe di semplificare l'accesso (come richiesto per esempio dal TDLAB – Laboratorio per il turismo digitale), ma un'ontologia propria, o quanto meno che sorga sotto il coordinamento del MiBACT, permetterebbe di valorizzare al meglio le interconnessioni tra i beni.

Infine una riflessione sul tema delle licenze. Ritengo che la CC0 "by default" sia una falsa soluzione per un falso problema. Ogni dataset potrà avere la licenza più appropriata ai contenuti che trasferisce, senza pregiudizi né nel senso della restrittività né verso opposto. Per dati relativi al patrimonio culturale, autorevoli in quanto provenienti o controllati dal MiBACT, ritengo che non si

possa prescindere dalla licenza CC-BY o CC-BY-SA. Inoltre bisogna distinguere tra dati e metadati. La scheda di catalogo, ad esempio, oltre ai dati anagrafici del bene, che potremmo definire “oggettivi” e quindi assimilabile ad un mero dato, ha dei contenuti scientifici che sono il frutto di ricerche ed elaborazioni storico-critiche (basti pensare al solo problema delle attribuzioni). Non si tratta quindi di porre una didascalia sotto a un’opera (che per altro tutto è meno che oggettiva), si tratta piuttosto di qualificare le relazioni culturali di un bene con il suo contesto.

In conclusione mi sembra si tratti di attivare un processo condiviso tra chi pubblica i dati e chi li riutilizza: a cosa servono questi dati? Che ri-uso se ne fa? Quali sono i nuovi contenuti che vengono creati? Un confronto su questi aspetti aiuterebbe la pubblica amministrazione a prendere coscienza delle operazione che si accinge a fare, che non è solo quella di trasformare dei dati secondo protocolli standard e metterli su un indirizzo pubblico. Anche se ci esprimiamo per sigle non va perso di vista che si sta facendo un’operazione culturale e non solo meramente tecnica.

Per questo motivo l’ICCD sta predisponendo un piano di comunicazione per presentare il progetto OPENiccd, che non significa il convegno d’ordinanza, ma piuttosto un programma di iniziative volto ad un coinvolgimento diretto e indiretto dei molteplici target di utenza individuati. Al momento l’ICCD, ma come tutto il ministero e la pubblica amministrazione in generale, è schiacciato sulle attività di gestione; tutte le nostre sempre minori risorse sono indirizzate a mantenere in piedi i sistemi informativi e a gestire i processi. Dovremmo invece riconquistare un ruolo di maggior indirizzo, che non significa solo dettare regole, ma fare delle vere e proprie azioni di *innovation scouting* finalizzate a ridurre le distanze tra l’istituzione e i vari target per avvicinare potenziali, creatori e promotori di innovazione, attraverso la creazione di una community per favorire l’incontro e la collaborazione con soggetti che a vario titolo potrebbero essere interessati.

Testo rilasciato con licenza Creative Commons Attribuzione – Condividi allo stesso modo (CC BY SA)