

# Eterogeneità naturale delle risorse informative: condividere standard per condividere risorse

*Pierre Piccotti*

## Introduzione

L'integrazione fra risorse informative tra loro diverse è un tema che il CNBA si trova ad affrontare tra i primi in Italia, probabilmente per la natura eterogenea dei supporti e della struttura dei documenti, tradizionali e digitali, che le biblioteche di architettura possiedono, affrontano e trattano nel lavoro quotidiano.

Il problema è aperto, e non sembra che si possano intravedere soluzioni in tempi brevi, a causa delle difficoltà sia tecniche sia operative comuni a tutte le biblioteche italiane.

Un esempio emblematico può venire da una recentissima iniziativa nel campo degli spogli di periodici, settore in cui il CNBA opera da parecchi anni con una realizzazione ormai consolidata<sup>1</sup>.

Sembra che si possa registrare un rinnovato interesse per gli archivi di spoglio di ambito nazionale, in correlazione con i progetti che gravitano intorno alla Biblioteca Digitale Italiana (BDI). Si è appena concluso infatti un incontro sul progetto RES<sup>2</sup>: la BNCF<sup>3</sup> ha organizzato un seminario per lanciare l'idea/progetto di realizzare un database (eterogeneo, cioè risultante da provenienze di produttori diversi) degli spogli.

Sono stati presentati un censimento e uno studio di fattibilità<sup>4</sup> che rimangono in qualche modo autoreferenziali: difatti, nell'affrontare questo tema non si è pensato di consultare gli attuali produttori di spogli in Italia (ACNP, ESSPER, Archinet<sup>5</sup>/CNBA). Il problema principale però è che a tutti questi reali produttori di spogli si chiederebbe di "cedere" il lavoro già fatto, riversandoli nel database citato.

A questo seminario ha partecipato anche la direzione dell'ICCU<sup>6</sup>, che si è dimostrata favorevole ad una iniziativa in qualche modo esterna, ma che si basa su un budget consistente: in buona sostanza, dati e finanziamento dovrebbero essere assicurati dai partecipanti al progetto stesso. Lascia quantomeno perplessi questa linea su un progetto al quale si viene chiamati da parte degli organi centrali nazionali.

E' possibile però tentare un approccio profondamente diverso, e questo è il messaggio che Archinet/CNBA lancia in questa occasione.

## Il contesto

Partiamo, concretamente, dall'ambito disciplinare che ci è più vicino e noto.

Quali sono le risorse informative (intese in senso generale) per l'architettura?

Laura Casagrande dell'Università Iuav di Venezia e Ezio Tarantino dell'Università degli studi di Roma "La Sapienza" in varie occasioni avevano fatto dei censimenti dai quali emergeva una realtà piuttosto disomogenea: come "fonti informative" infatti si possono intendere

---

<sup>1</sup> Si veda al sito CNBA: < [http://www.cnba.it/spoglio\\_intro.php](http://www.cnba.it/spoglio_intro.php) >

<sup>2</sup> Informazioni più dettagliate alla pagina: <<http://www.bnfc.firenze.sbn.it/progetti/RES/>>

<sup>3</sup> BNCF Biblioteca Nazionale Centrale di Firenze

<sup>4</sup> Sia il censimento che lo studio di fattibilità sono stati condotti su appalto esterno, non sono stati prodotti all'interno della BNCF stessa: già questo fatto può sollevare qualche dubbio sulla capacità effettiva di gestione di una simile iniziativa.

<sup>5</sup> Archinet <http://archinet.iuav.it> associazione fra CNBA, Iuav e Quasco il cui obiettivo è la distribuzione delle risorse informative per l'architettura, nella fattispecie, distribuisce per l'Italia Urbadisc/Urbadoc

<sup>6</sup> ICCU Istituto Centrale per il Catalogo Unico

- principalmente le biblioteche con i loro posseduti (eterogenei) e cataloghi (in genere differenziati per tipologia di risorsa specifica),
- alcune banche dati di spogli (sia nazionali che straniere),
- siti di settore (di cui alcuni anche a carattere iconografico),
- poco o quasi nulla relativamente al full-text (i periodici elettronici, nella loro accezione comune, di fatto scarseggiano).

E quali sono i principali problemi emersi?

- non risultano esserci esperienze consolidate di sistemi di integrazione: ad oggi, in Italia, l'unica che si può citare è ADA del Politecnico di Milano, che sarà l'oggetto di una delle presentazioni seguenti;
- le tipologie delle risorse informative sono varie: sia risorse primarie (full text e/o immagini), sia risorse derivate (record bibliografici, spogli, etc.);
- anche dal punto di vista disciplinare la situazione è eterogenea: cartografia, fotografia, progetti di architettura e urbanistica, design e arte più in generale.

## Le attuali esperienze

In questi anni ho seguito da vari punti di vista le problematiche dell'integrazione, partecipando in prima persona alla presentazione di diversi progetti: ExPract, UrPlaNet, MAPS, Arcipelago.it, Archivi Architettura, Gaudì, Archibald,...ed altri.

Diversi di questi sono poi falliti, o perché non finanziati, o per rinuncia dei partner.

Fra tutti, tre però hanno avuto uno sbocco: CriDaup, Urbadisc (ora Urbadoc) e MAIA.

Due parole per riassumere gli aspetti salienti di queste ultime due esperienze.

- Cri\_daup: Progetto di ricerca finanziato dall'allora MURST. Significativo, e antesignano, il modello di indicizzazione di record bibliografici attraverso il set Dublin Core di record provenienti da database differenti, l'accesso al record sul database originale, l'harvester di Google di file html D.C. del record originale.
- Urbadisc, (ora Urbadoc): Esperienza europea di raccolta di database di spoglio provenienti da istituzioni diversi e commercializzazione del prodotto.

Urbadoc<sup>7</sup>, la sua evoluzione, ora in fase di rilascio, ha approfondito il problema dei database eterogenei. In particolare non ha voluto/potuto utilizzare una struttura essenziale comune (Dublin Core) ma ha dovuto/potuto solo cercare di trovare fra database eterogenei (sia come tracciati, sia come contenuto) alcuni elementi comuni per rendere il tutto accessibile in forma integrata, anche attraverso un database cumulativo.

Problemi non indifferenti si sono trovati anche nel dover trattare set multilingue che vanno oltre ISO-latin, UTF-8, ma anche per quanto riguarda i dati per come sono stati strutturati dei diversi database. Comunque anche fra database simili è stato difficile riportare a standard comuni.

Dover mantenere la struttura originale dei dati e dover gestire direttamente i database che si mantengono diversi ha forzatamente portato ad una soluzione poco innovativa.

Più interessante appare la soluzione adottata per la visibilità dei dati: viene prodotto un tracciato standard XML che viene interpretato via XSLT "Soblotron" e quindi tradotto attraverso un convertitore quale SMARTY.

L'aspetto principale è la totale separazione fra presentazione e contenuti dei singoli database, peraltro tra loro indipendenti.

---

<sup>7</sup> Accesso a Urbadoc: <<http://www.urbadoc.com>>

## **MAIA e considerazioni sui metamotori**

MAIA è una realizzazione del CNBA da tempo entrato nell'uso comune<sup>8</sup>: piuttosto vorrei fare alcune considerazioni su MAI, che ne è stato il "padre" dal punto di vista architetturale, perché apre la strada a ragionamenti successivi sui "meta search engine".

### **MAI**

Il funzionamento di MAI è molto semplice. Esiste un database che contiene le sintassi di ricerca di tutti i database interrogabili: vuole essere un Meta Opac, ovvero un "meta search engine" orientato agli OPAC, ma non necessariamente a risorse informative di tipo bibliografico tout-court; non pretende di fare un "merge" delle informazioni raccolte, ma si limita a presentarle così come generate dai sistemi remoti.

Il limite principale che riscontro è l'altissimo costo di gestione, a fronte del mutare degli OPAC: ma non potrebbe essere diversamente perché MAI si pone l'obiettivo di fornire un accesso globale a tutto il mondo degli OPAC italiani e non vuole dettare regole o creare discriminazioni, filtri etc.

### **Meta Search engines**

In questi ultimissimi mesi sono iniziate anche in Italia alcune applicazioni in particolare del più conosciuto, MetaLib. Che fanno questi oggetti?

Similmente a MAI hanno un database interno con la sintassi dei diversi sistemi da interrogare, ma in più permette di elaborare i dati recuperati fornendo una visualizzazione dei dati più organizzata. Inoltre è in grado di creare per i singoli record recuperati una OpenURL (se non preesistente) da utilizzarsi per ulteriori elaborazioni.

Non è l'unico prodotto del genere sul mercato: ne esistono altri: dal tedesco IPS (in realtà creato da una azienda legata ad IHS), al nuovo meta motore di Sebina. In generale si caratterizzano tutti per il basso costo del software, e l'alto costo del "knowledge database", ovvero delle diverse "sintassi".

Vi è quindi un apparente paradosso: ogni acquirente paga non il prodotto, ma ex-novo ogni volta la possibilità di interrogare tutte le risorse rese interrogabili dal meta motore (knowledge database), anche se già pagato da un altro, anche se l'interesse sarebbe focalizzato a solo una piccola parte di esso. Ma questo è il mercato, ovviamente.

Sebbene generalmente io sono considerato un pessimista, credo ancora alle utopie, ovvero alla possibilità di ribaltare il quadro: vorrei pertanto azzardare un'ipotesi, una volta condivisi alcuni presupposti.

## **Uno scenario possibile**

E' mia ferma convinzione che, indipendentemente dal "tipo" di informazione, sia possibile operare una ricerca con sufficiente probabilità di successo (salvo, per ora, se voglio ricercare una forma in una immagine...) utilizzando solo campi del set Dublin Core<sup>9</sup>. E forse già sono troppi: oramai Google ha imposto delle regole non scritte, delle abitudini consolidate alle quali sembra che non ci si possa sottrarre.

---

<sup>8</sup> Accesso dal sito CNBA, con introduzione: <[http://www.cnba.it/maia\\_intro.php](http://www.cnba.it/maia_intro.php)>

<sup>9</sup> Il testo completo è disponibile a: <<http://www.niso.org/standards/resources/Z39-85.pdf>>

L'utente chiede di poter ricercare un'informazione attraverso una stringa di parole, e se proprio vogliamo offrire la possibilità di organizzare fra campi, dovremo anche fornire degli strumenti di riconoscimento della "volontà" dell'utente.

Questo vale ovviamente sintanto che si parla di ricerca originata da informazioni tipo autore, titolo, etc. Diverso è quando si parte da rappresentazioni spaziali: ma anche in questo caso, il set di metadati Dublin Core permetterebbero di creare legami con altre notizie.

Vero è il problema della notizia strutturata, ed è proprio questo l'elemento qualificante: come riportare il link a tutta la struttura.

I protocolli consolidati, di cui più avanti altri parleranno ampiamente, sono un numero molto limitato:

- Dublin Core per la descrizione semantica;
- Z39.50<sup>10</sup>, Zing<sup>11</sup> e harvester MPH o XML per la comunicazione.

Di fatto due sono gli approcci, o si condividono protocolli per interrogare database diversi o si esportano i dati in formati accessibili a motori di ricerca, possibilmente in forma strutturata.

## Una nuova ipotesi

E allora ribaltiamo il quadro: perché, in un ambito dai confini definiti, non ipotizziamo che siano i fornitori delle informazioni, i proprietari delle banche dati, che le rendono accessibili in base a protocolli condivisi?

Certo, alcuni lo fanno già, ma sono pochi<sup>12</sup>.

Il caso dell'uso di Z39.50 è indicativo: anche chi rende i dati disponibili sotto questo protocollo lo fa in forma molto limitata (quali sono i sistemi che forniscono i dati in formato strutturato, XML in particolare, e che offrono anche le funzioni di "browsing"?)

Certo, è utopia pensare che le biblioteche di architettura spingano perché i propri sistemi, generalmente di ateneo, adottino o cambino i software per essere interrogati: tuttavia perché, almeno per nuovi sistemi, almeno nel caso di database non strettamente bibliografici, non ipotizziamo se non lo Z39.50 che i sistemi forniscano un output XML strutturato, tale da essere raggiunto da *harvester* tipo quelli che si stanno diffondendo per gli E-prints o semplicemente da motori di ricerca sia pubblici (Google) sia propri, magari in grado di recuperare l'informazione in base a chiavi di ricerca (meta AG), p. es. Swish-e?

Molti sistemi si stanno adeguando per rendere "accessibili" il proprio Web e i propri cataloghi (W3C). Non è possibile impegnarsi anche a garantire l'"accessibilità" anche delle informazioni ai diversi sistemi di indicizzazione? Similmente a quanto avviene per Open Access, rendendo disponibili le pubblicazioni scientifiche "anche" attraverso protocolli condivisi?

Inoltre, perché non ipotizzare un progetto Open Source per la messa a punto di un *meta search engine* che recuperi questi dati?

Una prima analisi evidenzia che il livello di sviluppo software non è estremamente complesso: molto è già disponibile a livello di librerie, in ambiente Perl, PHP, Python, ma non esiste un prodotto completo in questo senso e difficilmente una simile nicchia potrà essere occupata dal semplice volontarismo di qualcuno.

---

<sup>10</sup> Oltre che nel sito NISO: <<http://www.niso.org>> questo standard, insieme ad ampia documentazione di vari progetti evolutivi, si trova presso la Library of Congress, ufficiale Maintenance Agency dello standard stesso: <<http://www.loc.gov/z3950/agency/>>

<sup>11</sup> Una panoramica dettagliata a: <<http://www.loc.gov/z3950/agency/zing/zing-home.html>>

<sup>12</sup> Basta vedere l'elenco degli cataloghi accessibili via Z39.50 presso il client Z39.50 dell'ICCU per gli indici nazionali e, a mia conoscenza, le università di Venezia, Genova, Siena, Pisa.

Però riprendiamo l'esperienza che si sta sviluppando nel mondo E-prints e realizziamo qualcosa in tal senso: io investirò risorse nel progetto (se riesco a trovarle attraverso la partecipazione a progetti nazionali o europei), e chiedo ad altri di contribuire.

Un progetto *Open Source* non nasce per partenogenesi e/o senza finanziamenti. E' un'operazione complessa che vede partecipi diversi attori, dagli sviluppatori, al team di progetto, alle istituzioni pubbliche, ai privati.

Ognuno può trovare un proprio spazio, ma comunque è sempre un progetto con costi di sviluppo. Però i risparmi si hanno

- perché lo sviluppo del “*core*” è iniziativa di singoli e non di aziende, specie per quanto riguarda il team di progetto (*Leader*);
- perché ci sono sovente supporti *free* e gratuiti;
- perché il prodotto evolve grazie al contributo di molti e non di pochi,

ma comunque serve un finanziamento iniziale, quindi una idea condivisa che diventi una linea di azione concreta di più di una struttura.

Forse qualcuno dirà che già è stato fatto molto in proposito, che si sono già prodotti di questo tipo, primo fra tutti quello dell'università La Sapienza di Roma, MetaBIDS<sup>13</sup>. Tuttavia non è un prodotto *Open Source*, è un progetto sostanzialmente interno, è limitato all'uso interno, ripropone il modello di “meta” interrogazione *http get*, non permette una gestione della navigazione “avanzata”. E' tuttavia un ottimo punto di partenza che potrebbe trovare sviluppi ulteriori nell'ambito di un progetto *Open Source*.

Gli aspetti caratterizzanti di questa proposta dovrebbero quindi essere:

- rispetto della filosofia *Open Source*
- gestione di notizie strutturate a più livelli (problema neppure sfiorato dalla proposta RES-BNCF)
- navigazione avanzata tra le informazioni
- correlazione a “servizi aggiunti” di tipo nativo.

Condividere per cooperare è una vecchia filosofia del CNBA e di Archinet: proprio da queste esperienze viene la coscienza che se è importante “condividere i dati”, lo è soprattutto “condividere i protocolli”, per una gestione efficace del lavoro dei singoli e per la moltiplicazione dell'usabilità dei dati a vantaggio dell'intera comunità.

---

<sup>13</sup> Informazioni alla pagina: <<http://w3.uniroma1.it/cobai/pagine/cataloghi.aspx>>