

# Web connected information seeking

ANTONELLA IACONO

Università di Roma "La Sapienza"  
antonella.iacono@fastwebnet

---

## La ricerca nell'OPAC dall'informazione alla conoscenza

---

Oggi si parla sempre di più dei *Linked data* come tecnologia unificante per collegare i dati nel web e in grado di aprire un nuovo capitolo del percorso evolutivo del catalogo.

Com'è noto, il termine *Linked data*, nella definizione classica fornita dal fondatore del web Tim Berners Lee, si riferisce a una serie di buone pratiche per pubblicare e collegare dati strutturati in rete, favorendo la costruzione di una rete di dati globale, i cui contenuti possono essere scambiati e interpretati dalle macchine, formando la base per la costruzione del Web semantico.<sup>1</sup>

Adottando i *Linked data* come nuovo modello logico per esprimere l'informazione bibliografica, le biblioteche potranno non solo operare un cambiamento nei metadati con cui hanno rappresentato per decenni i loro dati bibliografici,<sup>2</sup> ma anche trovare nuove forme per riprogettare la ricerca negli OPAC, i quali rappresentano l'interfaccia tra l'utente e il catalogo.

Sappiamo bene, infatti, che i nostri cataloghi nell'ultimo decennio sono diventati più amichevoli, collaborativi e sociali, ma non sono mai stati pienamente integrati in rete. Sono "sul web", ma non fanno parte "del Web";<sup>3</sup> i cataloghi contengono una grande quantità d'informazione che non è accessibile alle applicazioni semantiche. Se, dunque, oggi vogliamo che le biblioteche continuino a essere il punto di riferimento per chi fa ricerca è necessario che i loro dati siano recuperabili, aperti, e integrati con gli altri dati presenti in rete.

Ciò è ancor più vero se pensiamo che anche nel web si cominciano a mettere a frutto le tecnologie del Web semantico (un noto esempio è Google Knowledge Graph)<sup>4</sup>

per rendere più "intelligenti" i motori di ricerca, ai quali ogni giorno milioni di utenti si rivolgono, dimostrando come il vero valore aggiunto sia quello di recuperare informazione organizzata e contestualizzata.<sup>5</sup>

Nel nuovo scenario del Web semantico, in cui il passaggio a una rete di dati corrisponde alla creazione di una rete di saperi e di conoscenze intrecciate tra loro, il catalogo può così diventare lo strumento con cui le biblioteche contemporanee possono colmare le lacune di un sapere oggi sempre più frammentato e agevolare i processi di trasformazione delle informazioni in sapere, collegando le conoscenze prodotte dalle biblioteche agli altri dati presenti nel web.

I *Linked open data* aprono nuove prospettive per realizzare questo importante passaggio evolutivo:

- 1) consentono di disporre di metadati bibliografici non "bibliocentrici" e più adatti a connettere le conoscenze per mezzo di una tecnologia comune al web; l'adozione dei *Linked data* è posta, oggi, come alternativa al formato MARC,<sup>6</sup> che da tempo viene additato come il principale ostacolo all'interoperabilità tecnologica dei dati con il web;<sup>7</sup>
- 2) permettono, in prospettiva, di costruire interfacce di ricerca migliorate, cioè in grado di sfruttare appieno i vantaggi provenienti dalla scomposizione dei record bibliografici in "dati collegati" ai fini di un recupero qualitativamente più ricco.

### La ricerca "connessa al web": dall'informazione alla conoscenza

In questo contributo si tenterà di delineare alcuni aspetti dell'interazione utente-catalogo che potrebbero beneficiare dell'applicazione dei *Linked data* alla struttura catalografica, seguendo un approccio costruttivista e utente-centrico, entrando così in quel campo di studi della Scienza dell'informazione denominato *Information*

---

\* Testo rivisto dell'intervento tenuto in occasione del convegno "La biblioteca connessa. Come cambiano le strategie di servizio al tempo dei social network" (Milano, Palazzo delle Stelline, 13-14 marzo 2014).

*seeking behavior*, dove la ricerca viene intesa come una serie di passaggi in cui l'utente manifesta determinati comportamenti.<sup>8</sup>

L'adozione di questa tecnologia apre, infatti, la possibilità di rivedere i modelli di ricerca attualmente in uso, ancora troppo fissati ai sistemi, in favore di un recupero dell'informazione cognitivo,<sup>9</sup> cioè in grado di favorire l'apprendimento durante il processo informativo, trasformando così il catalogo da sistema orientato al documento a sistema orientato alla conoscenza.<sup>10</sup>

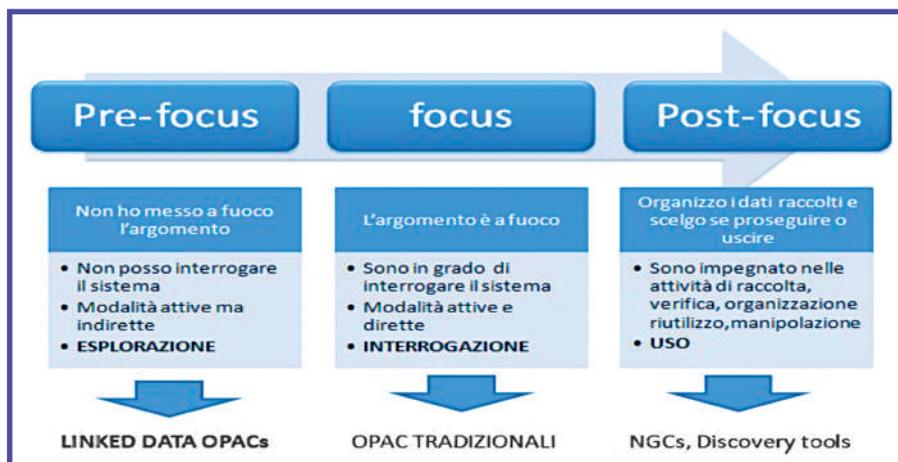
L'apprendimento durante un processo di ricerca è un fattore essenziale al buon successo del dialogo tra utente e catalogo,<sup>11</sup> ma sia al momento dell'attivazione della ricerca sia lungo il processo informativo l'utente incontra numerosi ostacoli di tipo *cognitivo*.<sup>12</sup> Questi consistono:

- in una scarsa comprensione dell'oggetto della ricerca qualora l'utente si trovi in una fase iniziale del processo;
- in una difficoltà di espressione dell'oggetto di ricerca, derivante dall'inadeguata conoscenza dell'area disciplinare che si sta indagando;
- nella limitata conoscenza di un sistema e dei suoi meccanismi di recupero dell'informazione (barriere tecnologiche, semantiche, e concettuali) o delle modalità di rappresentazioni dei testi;
- nelle barriere linguistiche, cui si possono aggiungere anche quelle barriere più "sottili" provenienti dai modelli mentali che gli utenti hanno costruito nell'utilizzo di altri strumenti di ricerca della rete.<sup>13</sup>

Oggi, grazie agli studi che nell'ultimo ventennio si sono occupati di analizzare la ricerca dell'informazione, sappiamo che quest'attività è ben lontana dal modello classico e "ingenuo" di *information retrieval*, nel quale l'utente formula una domanda e il sistema fornisce la "risposta perfetta", modello che ha ispirato fin dall'inizio la progettazione delle funzionalità dei cataloghi elettronici, ma che è un processo dinamico, iterativo e interattivo.

Facendo riferimento a uno dei più fortunati modelli cognitivi di rappresentazione del processo di ricerca – il noto modello a fasi di Kulthau<sup>14</sup> – possiamo così immaginare che, nell'interazione con l'OPAC, l'utente attra-

Figura 1 – Fasi della ricerca e attività prevalenti dell'utente. L'approccio *Linked data* favorisce in particolare modo l'esplorazione dei contenuti del catalogo



versi varie fasi, corrispondenti a diversi gradi di "consapevolezza" della ricerca, in cui cambiano sia le strategie di ricerca adottate che le attività svolte.

È noto che nella prima fase (*pre-focus stage*), quando l'oggetto di una ricerca non è ancora "a fuoco", l'utente sceglie modalità di approccio al catalogo attive e indirette, ovvero esplorative; nella fase centrale (*in-focus stage*) prevalgono modalità attive che condurranno l'utente a formulare interrogazioni; nella fase conclusiva (*post-focus stage*) l'utente utilizzerà e manipolerà l'informazione recuperata.

Le tecnologie in uso nei primi OPAC sono state tradizionalmente rivolte a risolvere le problematiche relative alla sola interrogazione, dando per scontato che l'utente fosse in grado, in qualsiasi momento, di conoscere l'oggetto della sua ricerca e di esplicitarlo al sistema e limitando le opzioni di *browsing* allo scorrimento di liste di autori o soggetti. Con le interfacce dei primi *Next generation catalogs* si è avviato il percorso verso modelli di recupero dell'informazione più attenti ai comportamenti di ricerca dell'utente; queste, infatti, hanno introdotto l'emulazione delle interfacce del web "alla Google", il paradigma esplorativo delle faccette (*faceted navigation*) e varie funzionalità di riutilizzo e manipolazione dell'informazione, come la possibilità di interagire con le risorse catalografiche, sia per organizzarle (*tagging*) sia per riutilizzarle (esportazione di citazioni bibliografiche, interazione con i social network). Il paradigma dei *Linked data*, come si vedrà, consente di adottare nuovi e innovativi approcci alla ricerca esplorativa, migliorando nel complesso anche le ulteriori funzioni svolte dall'utente nelle altre fasi del processo informativo.

Il più grande vantaggio di un'esperienza di ricerca "web-

connessa” sarà quello di utilizzare il web dei dati come nuovo contesto, allargato ai fini della ricerca. Gli OPAC che utilizzeranno la tecnologia dei *Linked data* saranno profondamente integrati nel web e ciò consentirà all’utente di partire dalle informazioni bibliografiche contenute nei cataloghi, per scoprire senza soluzione di continuità altri dati presenti nello stesso spazio informativo, ovvero nel web dei dati, allargando così il contesto utile per la navigazione e migliorando complessivamente le possibilità di ricerca e di creazione di nuova conoscenza. Diviene così centrale il passaggio dalla navigazione nel catalogo all’esplorazione, intesa come l’azione di mettere in relazione le risorse per ricostruirne il contesto.<sup>15</sup>

### I *Linked data* e la ricerca esplorativa

I *Linked data* offrono la base tecnologica per costruire percorsi trasversali tra i contenuti provenienti da diverse fonti informative, migliorando la possibilità di esplorazione delle risorse del catalogo.

Ciò avvantaggia l’utente che si trova in una prima analisi del suo bisogno informativo e che è naturalmente predisposto ad attività come la navigazione e il *browsing*. Possiamo immaginare che in questa fase preliminare l’utente sia coinvolto – riutilizzando il concetto espresso da Dervin<sup>16</sup> – in una continua acquisizione di senso, rivolta a comprendere e interpretare l’universo bibliografico ivi rappresentato. L’approccio esplorativo delle faccette, introdotto da OPAC di nuova generazione e *discovery tools*, ha certamente contribuito a superare una visione rigida e statica della ricerca *query orientated*,<sup>17</sup> tuttavia questa tecnologia è basata sull’estrazione automatica di dati dai record bibliografici e non sull’esposizione della struttura delle relazioni bibliografiche tra le entità. Tale approccio si rivela così non del tutto adeguato a navigare nel catalogo, in quanto non consente di ricostruire il contesto in cui ogni entità è collocata.

Il cambio di formalismo dei *Linked data* e la conseguente scomposizione del record bibliografico nelle entità costitutive (cose, persone, luoghi, concetti) consente di ri-orientare la progettazione delle interfacce di ricerca sul modello entità-relazione, riorganizzando il catalogo intorno a tali entità e ricostruendo attorno a esse un contesto esplorabile dall’utente. Sarà così possibile costruire pagine dinamiche per ogni entità di interesse del catalogo, per esempio un’opera, un soggetto, un autore. In tal modo viene ampliata la possibilità di recupero di un’entità nel suo contesto, rappresentando la sua collocazione nell’universo bibliografico del catalogo e

usufruendo di collegamenti con gli altri dati del web. La decostruzione del record bibliografico in dati tra loro collegati e la possibilità di qualificare i collegamenti tra le informazioni, ricostruendo le relazioni previste dai modelli concettuali *Functional Requirements* dell’IFLA, consentono una navigazione gerarchica che dal raggruppamento di opere correlate mostri le singole opere, le espressioni e tutte le manifestazioni e gli item a esse collegati. In tal modo si compiono le funzioni bibliografiche che l’utente svolge nel catalogo e si realizza una navigazione “significativa” per l’utente e, al contempo, un processo di apprendimento.

Vi sono vari interessanti esempi di applicazione di questo paradigma alle interfacce per la ricerca. L’esempio più noto agli utenti del web è certamente Open library,<sup>18</sup> l’ambizioso progetto collaborativo di Internet Archive che si propone di creare in Internet una pagina per ogni libro pubblicato e che genera “al volo” pagine in RDF per autori, titoli e soggetti.

Un altro interessante modello di interfaccia di questo tipo è Data.bnf.fr,<sup>19</sup> la piattaforma di ricerca creata dalla Bibliothèque Nationale de France basata sui *Linked data*, che sfrutta i dati bibliografici e le descrizioni provenienti da OPAC, archivi e dalla biblioteca digitale Gallica, per produrre visualizzazioni basate su FRBR e dunque, nel nostro caso, un esempio interessante per la progettazione di nuove interfacce di ricerca.<sup>20</sup> La ricerca è costruita sulle tre principali modalità di accesso ai dati bibliografici: l’opera, l’autore, l’argomento; essa prevede la creazione di pagine distinte per le tre entità che sono dinamicamente aggiornate attraverso la registrazione delle notizie come *Linked data*. La piattaforma riunisce dati in svariati formati (MARC, EAD e DC) che vengono dotati di identificativi persistenti e possono essere raggruppati in base all’opera.

La pagina dell’opera è creata sulla base dei record di autorità per l’opera fornite dal catalogo della Bibliothèque Nationale de France e strutturata secondo i livelli previsti da FRBR. Essa presenta, per ogni opera, le informazioni relative al contenuto, le opere correlate e le relazioni intero-parte tra opere e tra manifestazioni. La pagina web di un autore, oltre a contenere i dati completi, mostra l’elenco delle sue opere e dei differenti ruoli dell’autore nelle opere correlate. La pagina di un soggetto ha forme varianti e collegate, collegamenti agli autori che hanno trattato quel soggetto e l’indicazione delle opere estratte dai cataloghi. Tutte le pagine sono arricchite dai dati provenienti da diversi *datasets* esterni, come DbPedia, il grande dataset al centro della nuvola dei *Linked open data*, dal

quale vengono estratte informazioni supplementari sulle opere, titoli e soggetti. Le pagine, pubblicate in HTML, vengono rese “semanticamente visibili” ai motori di ricerca estraendone i dati descrittivi più rilevanti e applicando le marcature previste dallo standard RDFa, tramite l’uso di vocabolari condivisi come Schema.org<sup>21</sup> utilizzato per codificare le pagine con i microdata, oggi riconosciuto dai principali motori di ricerca, e lo schema di marcatura OpenGraph Protocol (OG)<sup>22</sup> che consente di descrivere questi dati nei social network. Nella stessa direzione è stato riorganizzato l’OPAC della Biblio-

Figura 2 – La piattaforma data.bnf.fr è strutturata sui tre principali accessi al catalogo: opere, autori, soggetti



teca Nazionale svedese LIBRIS.<sup>23</sup> Dal 2008 la biblioteca pubblica i dati come *Linked data* utilizzando note ontologiche come Dublin Core, Bibliontology e SKOS, e si prepara ad adottare RDF (Resource Description Framework) come formato nativo delle registrazioni del catalogo.<sup>24</sup> Oggi l’OPAC, grazie alla trasformazione dei dati in RDF, è in grado di fornire in risposta a un’interrogazione un elenco ordinato in base al modello FRBR e a breve sarà possibile accedere a una pagina per ogni titolo, autore e soggetto presente nella base dati.

Si mettono così in atto nuovi modi per creare interfacce che sfruttano la presenza di dati provenienti dalle fonti più diverse, per realizzare visualizzazioni in grado di manifestare naturalmente la scansione delle entità previste nei modelli FRBR.

La ricerca esplorativa nel catalogo potrà beneficiare, poi, dei nuovi collegamenti tra diversi domini della conoscenza che la tecnologia dei *Linked data* rende possibili, offrendo all’utente la possibilità di attivare percorsi di esplorazione che, partendo da dati bibliografici, raggiungono informazioni di altra natura. La navigazione potrà avvenire non solamente nel contesto del catalogo, ma potrà estendersi al web, inglobando dinamicamente informazioni provenienti da fonti esterne.<sup>25</sup> Il collegamento a grandi *dataset* del patrimonio culturale (come Europea, Linked Heritage o il nuovo progetto Cultura Italia)<sup>26</sup> permetterà di ricostruire intorno a un libro il contesto storico, culturale o letterario nel quale esso si inserisce; sarà possibile così saltare da un concetto a un oggetto museale che lo rappresenta, come un quadro,

una scultura. Ciò comporta per le biblioteche la capacità di rappresentare nuove tipologie di aggregazioni e di gestire nell’OPAC nuovi contesti esplorabili che l’utente potrà percorrere, collegando dati bibliografici, oggetti digitali, contenuti multimediali, risorse educative e materiale prodotto dalle altre comunità del web.<sup>27</sup>

La ricerca semantica potrà invece avvantaggiarsi delle possibilità di *ontology matching* di schemi diversi rese possibili dall’utilizzo di linguaggi come SKOS e OWL, le ontologie del Semantic web che consentono di esprimere relazioni tra concetti per modellare soggetti, thesauri e schemi di classificazione. L’ambiente dinamico e interattivo dei *Linked open data*, nel quale “ogni comunità può dire qualsiasi cosa su un concetto”, nel mondo delle biblioteche attiva così enormi effetti di rete in quanto rende possibile collegare, confrontare e unire schemi tra loro diversi per tipologia e profondità, come per esempio un soggettario o una folksonomia.<sup>28</sup> La scomposizione dei soggetti e delle classi nei concetti che li compongono potrà unire sistemi di organizzazione delle conoscenze creati dalle biblioteche e sistemi creati da altre comunità del Web semantico, offrendo modalità avanzate di recupero dell’informazione. Vaste reti bibliotecarie che utilizzano sistemi di organizzazione delle conoscenze diversi potranno unirli, mappando i loro vocabolari per favorire la ricerca.<sup>29</sup>

Non mancano, infine, modalità inedite di recupero e visualizzazione dell’informazione che potranno essere offerte all’utente come modalità alternative di ricerca: nello scenario dei *Linked data* è possibile infatti collega-

re i concetti definiti negli schemi di classificazione alle cose reali descritte nel web.<sup>30</sup> Sono in atto alcune sperimentazioni per ottenere, attraverso la scomposizione delle notazioni o stringhe di soggetto e il collegamento delle faccette geografiche a *datasets* autorevoli come Geonames,<sup>31</sup> la visualizzazione di dati (soggetti e classi) a partire da una mappa geografica.

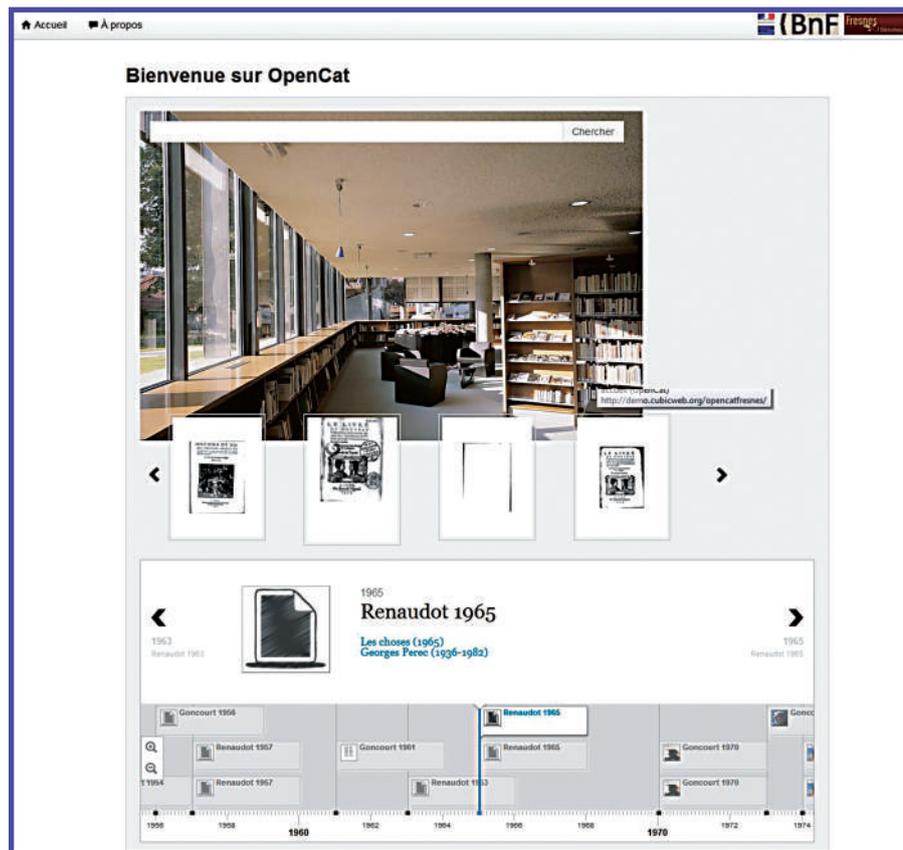
Allo stesso tempo si potranno realizzare sui dati collegati nuove forme di *visual retrieval* semantico, tramite applicativi che consentono di rappresentare visivamente le relazioni tra le entità recuperate e ottenere visualizzazioni complesse basate su dati e ontologie.<sup>32</sup>

### Interrogazione del catalogo e valutazione dei risultati

Non meno importanti sono i vantaggi che l'utente potrà sperimentare quando si troverà nella fase intermedia del processo informativo, cioè con un argomento di ricerca (*focus*) già formato. Sappiamo che in questa fase è necessario ricevere aiuto nel formulare l'interrogazione e avere più elementi utili per valutare i risulta-

ti. L'espressione dei dati bibliografici nella forma di dati collegati potrebbe facilitare questi compiti. Alla base di un recupero dell'informazione cognitivo e orientato all'utente vi è la possibilità di interrogare sistemi servendosi del linguaggio naturale e utilizzando la propria lingua. *Linked data* offre il supporto tecnologico che consente il superamento delle barriere linguistiche, ponendo le basi per creare piattaforme di ricerca multilingua; un URI, ovvero l'identificativo univoco e persistente associato al dato, infatti, è totalmente indipendente dal linguaggio usato per esprimere un concetto. Attraverso il collegamento a strumenti di *authority* condivisi come VIAF<sup>33</sup> tutte le forme varianti di un nome o di un titolo o di un soggetto possono dunque essere collegate alla forma accettata. L'utente potrà esprimere un concetto in una lingua diversa da quella per cui il catalogo è progettato e ottenere un risultato. Un ulteriore vantaggio per l'utente è quello di poter interrogare il catalogo per essere indirizzati ai termini più adatti per la ricerca, attraverso funzionalità di autocompletamento della *query* più intelligenti, che potranno basarsi su mappe semantiche o ontologie ricavate dal web dei dati.

Figura 3 – Il progetto sperimentale OpenCat sfrutta dati provenienti da dataset esterni per arricchire i dati catalografici dell'OPAC



In una ricerca connessa al web dei dati, il collegamento dei dati bibliografici ad altri *datasets* autorevoli rende possibile un grande arricchimento dei metadati bibliografici, utile nella valutazione dei risultati di un'interrogazione. Strutturare i record bibliografici come dati collegati consentirà di ampliare lo spettro delle informazioni che possono far parte di un reticolo bibliografico, aggiungendovi dati provenienti da fonti bibliografiche e non bibliografiche. Ciò risponde pienamente alle esigenze dell'utente: le indagini sui desiderata degli utenti dei cataloghi hanno infatti rivelato che quando l'utente richiede più "informazione di soggetto" non si riferisce alle forme controllate fornite dagli accessi per soggetti o per classi, ma a elementi per lo più esterni alle descrizioni catalografiche (toc, abstract, link a risorse esterne).<sup>34</sup> L'utente del catalogo vuole recuperare le risorse su

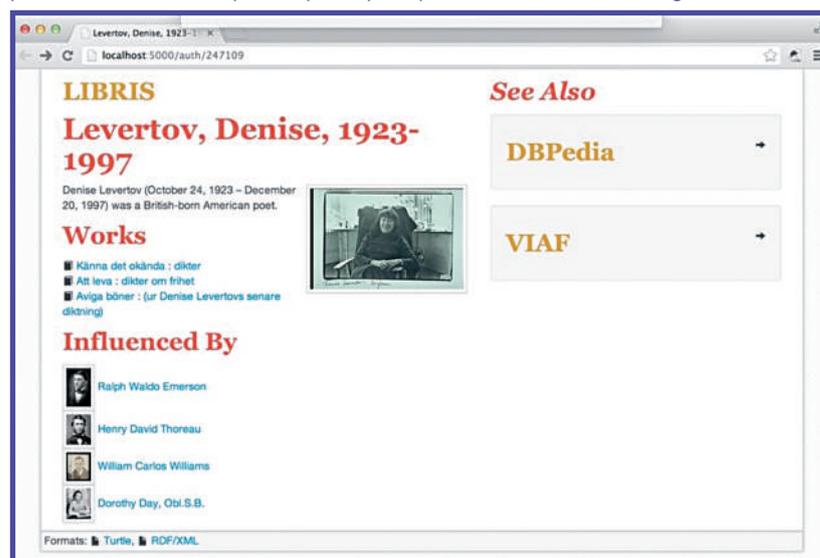
un soggetto (possedute o non possedute dalla biblioteca), visualizzare soggetti correlati, avere più informazioni di base su un argomento, che lo aiutino a stabilire la pertinenza e l'utilità dell'informazione ricevuta.

Le biblioteche possono sperimentare in questa direzione nuove forme di integrazione dei loro dati con informazione strutturata proveniente sia da altri cataloghi sia da prodotti di terze parti. Il nuovo progetto per convertire il catalogo SBN in *Linked open data*<sup>35</sup> mira a offrire alle biblioteche un'interfaccia per derivare entità (opere o espressioni) di loro interesse da utilizzare nei propri cataloghi. Un'esperienza simile, riguardante l'arricchimento del catalogo con dati provenienti da dataset del patrimonio culturale è il progetto OpenCat,<sup>36</sup> avviato nel 2012 della

Bibliothèque Nationale de France<sup>37</sup> allo scopo di riutilizzare dati provenienti dalla piattaforma data.bnf.fr, esposti come *data dumps* e semi FRBR-izzati, per creare un'interfaccia di ricerca innovativa per il catalogo della piccola biblioteca di Fresnes, sfruttando dati esposti come *Linked data* per ottenere un'interfaccia di nuova generazione.<sup>38</sup> Nel prototipo del progetto, la pagina di un autore contiene elementi informativi provenienti dai record di autorità prodotti dalla Bibliothèque Nationale de France (che vengono automaticamente combinati generando link alla pagina della piattaforma dedicata all'autore), note biografiche e altri collegamenti utili. Tra le funzionalità più interessanti ritroviamo la *timeline* che mostra tutte le opere dell'autore secondo la data di prima pubblicazione (funzione che oggi manca anche nelle interfacce degli OPAC più avanzati), ottenuta rielaborando i dati ricavati dai dataset della BNF.

Un'altra grande fonte per l'arricchimento dei cataloghi è Dbpedia,<sup>39</sup> il grande dataset che raccoglie informazione strutturata estratta da Wikipedia; ogni pagina mostra una serie di collegamenti a entità che normalmente non figurano nel catalogo di una biblioteca: per la pagina di un autore sono presenti informazioni aggiuntive come la nazionalità, il ritratto, la biografia, i personaggi da lui creati, il sito web, alcuni accessi per soggetto collegati, e molti *links* che possono aiutare l'utente a valutare più appropriatamente l'informazione. Un recentissimo esempio di arricchimento si ritrova nel nuovo OPAC sperimentale della Biblioteca Nazionale svedese,

Figura 4 – Un'anteprima del nuovo Libris XL, il cui rilascio è previsto per la fine del 2014. La pagina di un autore è arricchita di informazioni provenienti da Dbpedia e VIAF, ai quali sono presenti anche i rimandi per una più ampia esplorazione delle risorse collegate<sup>40</sup>



Libris XL, che sarà attivo entro la fine del 2014. La pagina di un autore è arricchita di informazioni provenienti da Dbpedia e VIAF, ai quali sono presenti anche i *links* per una più ampia esplorazione delle risorse collegate.

### Utilizzo dell'informazione recuperata

Infine, le tecnologie del Web semantico consentiranno all'utente di utilizzare in nuovi modi l'informazione recuperata in un catalogo. In un mondo in cui l'attenzione è scarsa e le risorse sono abbondanti diventa sempre più importante trovare il modo di utilizzare le risorse del catalogo in nuove modalità e formati, offrendo svariate possibilità di utilizzo dei dati, che vanno anche oltre la ricerca nel catalogo; le informazioni del catalogo dovranno poter essere recuperate dai motori di ricerca del web, dai social network e da altre applicazioni semantiche, ampliando le possibilità di accesso ai dati bibliografici e di riutilizzo degli stessi.<sup>41</sup> Nel quadro così delineato, i dati bibliografici non rimarranno chiusi negli OPAC, ma saranno disponibili per l'utilizzo in altre applicazioni del Web semantico, ovunque e in qualsiasi momento:

Library data about our resources is no longer just meant to be stored in catalog drawers as the inventory tool to access an individual library's collections. It can now be made available to anyone, anywhere in the world, at any time.<sup>42</sup>

La conservazione dei dati bibliografici in *triplestores* e la creazione di *endpoints*<sup>43</sup> interrogabili tramite i linguaggi del Web semantico, come SPARQL, rendono possibi-

le indirizzare al catalogo interrogazioni complesse che potranno essere usate sia dalle biblioteche sia dalle altre applicazioni del web per offrire all'utente nuovi servizi, che sfrutteranno dati sempre aggiornati.<sup>44</sup> Questa profonda connessione con il web dei dati renderà la ricerca nel catalogo più efficace, generando un ritorno di utenza verso il catalogo.

## Conclusioni

Nel percorso verso l'integrazione dei dati bibliografici nel Web semantico i cataloghi delle biblioteche hanno un ruolo di primo piano: se infatti la possibilità di accedere ai dati bibliografici da altre applicazioni del Web semantico rappresenta certamente un vantaggio per le biblioteche, non va dimenticato che il web dei dati è un ambiente fortemente eterogeneo; nei nostri OPAC, invece, questi dati possono mantenere la loro coerenza e completezza e le biblioteche potranno sfruttare i vantaggi dei *Linked data* in un ambiente controllato e autorevole e, allo stesso tempo, connesso al web. Per questi motivi il compito delle biblioteche e dei bibliotecari sarà sempre di più quello della scelta dei contenuti (*content curation*) da rappresentare nel catalogo, al fine di ricostruire i molteplici contesti collegati a ogni risorsa, che saranno così resi disponibili a vantaggio di chi fa ricerca. I *Linked data* offrono solo i presupposti tecnologici per connettere i dati al web, creando così un nuovo rapporto tra l'universo bibliografico e l'universo informativo tout court; la grande cooperazione che oggi è in atto tra le biblioteche e le istituzioni che operano nell'ambito del *cultural heritage*, della ricerca scientifica e dei servizi culturali costituisce la premessa per rendere la ricerca nei cataloghi sempre più intelligente e connessa.

"JLIS.it", 4 (2013), n. 1, p. 67-90, <<http://leo.cilea.it/index.php/jlis/article/view/6305>>.

<sup>3</sup> KAREN COYLE, *Linked Data: an Evolution*, "JLIS.it", 4 (2013), n. 1, p. 53-62, <<http://leo.cilea.it/index.php/jlis/article/view/5443/7889>>.

<sup>4</sup> <<http://www.google.com/insidesearch/features/search/knowledge.html>>. Il nuovo strumento di ricerca Google Knowledge Graph è stato inaugurato nel 2012 e dopo pochi mesi è divenuto operativo anche nella versione italiana di Google.

<sup>5</sup> Coyle sostiene che il principale motivo per cui gli utenti ignorano i cataloghi, rivolgendosi a strumenti come Wikipedia o Amazon, sia la possibilità di ottenere da essi "informazione organizzata".

<sup>6</sup> ANGELA KROEGER, *The road to BIBFRAME: the evolution of the idea of bibliographic transition into a post-MARC Future*, "Cataloging & Classification Quarterly", 51 (2013), n. 8, p. 1-18.

<sup>7</sup> KAREN COYLE, *Library data in a modern context*, "Library technology reports", 46 (2010), n. 1, p. 5-13.

<sup>8</sup> Nella letteratura si sono consolidati due orientamenti opposti: agli studi che considerano la ricerca come un meccanismo di comparazione tra i termini di un'interrogazione e i termini presenti in un documento, si sono opposti gli studi che la considerano come processo in cui l'utente si muove verso un risultato tramite una serie di passaggi o fasi. Nel primo caso (*system-centric*) tutta l'attenzione è focalizzata sulla creazione di un'interrogazione e sull'ottenimento, in risposta, di un risultato appropriato alle esigenze di ricerca personali. Nel secondo caso (*user-centric*) l'attenzione si sposta verso la ricerca come situazione di *problem solving* e la capacità di guidare l'utente verso un processo di ricerca soddisfacente. Per molto tempo le problematiche dei cataloghi sono state collegate alla capacità di rispondere a interrogazioni dell'utente, concentrandosi solo sull'analisi dei testi. Cfr. ANTONELLA IACONO, *Verso un nuovo modello di OPAC. Dal recupero dell'informazione alla creazione di conoscenza*, "JLIS.it", 4 (2013), n. 2 (<<http://leo.cilea.it/index.php/jlis/article/view/8903/8069>>), e EAD., *Dal record al dato. Il nuovo 'record bibliografico collegato' e la ricerca dell'informazione nell'OPAC*, ivi, 5 (2014), n. 1 (<<http://leo.cilea.it/index.php/jlis/article/view/9095>>).

<sup>9</sup> I modelli cognitivi o interattivi di *information retrieval* si situano in un'area di sovrapposizione tra le scienze cognitive, l'interazione uomo-macchina, *l'information retrieval* e l'informatica e derivano dalla necessità di riunire gli studi sull'*information retrieval* con il filone dell'*information behaviour*. Questi ultimi si basano su un approccio "centrato sull'utente" identificando modelli comportamentali e *process-based*, che indagano i modi in cui gli utenti cercano e usano l'informazione. A differenza dei modelli di *information retrieval* che rimangono sostanzialmente ancorati a paradigmi basati sul *matching*, questi modelli si sono orientati verso la scoperta dei processi che regolano l'accesso, lo scambio e l'uso dell'informazione. Questo approccio qualitativo alla ricerca dell'informazione è basato sulla necessità di comprendere i modi in cui l'informazione si manifesta in relazione al contesto e i processi di comprensione e apprendimento, ovvero la sua possibilità di trasformarsi in conoscenza. Cfr. *Theories of information behaviour*, a cura di Karen E. Fisher, Sanda Erdelez e Lynne Mckechnie

---

---

## NOTE

<sup>1</sup> TOM HEATH - CHRISTIAN BIZER, *Linked Data: Evolving the Web into a Global Data Space*, "Synthesis Lectures on the Semantic Web: Theory and Technology", 1 (2011), n. 1, p. 1-136, <<http://linkeddatatoolkit.com/editions/1.0/>>.

<sup>2</sup> Si vedano GETANEH ALEMU et al., *Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models*, 2012, in IFLA Council and General Conference (78<sup>th</sup>/2012, Helsinki), Helsinki, 11-17 August 2012, <<http://conference.ifla.org/sites/default/files/files/papers/wlic2012/92-alemu-en.pdf>> e MAURO GUERRINI - TIZIANA POSSEMATO, *Linked data: a new alphabet for the semantic web*,

Medford, published for the American Society for Information Science and Technology by Information Today, 2005; IRIS XIE, *Interactive information retrieval in digital environments*, New York, IGI Global, 2008; *Interactive information seeking, behaviour and retrieval*, a cura di Ian Ruthven e Diane Kelly, London, Facet, 2011.

<sup>10</sup> RAHMATOLLAH FATTAHI, *From Information to Knowledge: SuperWorks and the Challenges in the Organization and Representation of the Bibliographic Universe = Dall'informazione alla conoscenza: le super-opere e le sfide dell'organizzazione e rappresentazione dell'universo bibliografico*, Lectio magistralis in Biblioteconomia, Università degli studi di Firenze, 2010; Firenze, Casalini Libri, 2010.

<sup>11</sup> Cfr. CHRISTINE L. BORGMAN, *Why are online catalogs still hard to use?*, "Journal of the American Society for Information Science", 47 (1996), n. 7, p. 493-505.

<sup>12</sup> Cfr. THOMAS D. WILSON, *Models in information behaviour research*, "Journal of Documentation", 55 (1999), n. 3, p. 249-270.

<sup>13</sup> HOLLY YU - MARGO YOUNG, *The Impact of Web Search Engines on Subject Searching in OPAC*, "Information Technology And Libraries", 23 (2004), n. 4, p. 168-180, anche in <<http://www.ala.org/ala/lita/litapublications/ital/volume23a/number4a/yl.pdf>>

<sup>14</sup> Il noto modello "a fasi" Kuhlthau si propone in questo contributo in una versione semplificata, già adottata da altri studiosi (es. Vakkari), che riunisce le prime tre fasi (*task initiation, topic selection, prefocus exploration*) e le ultime due (*information collection, search closure*), cfr. CAROL COLLIER KUHALTHAU, *Seeking meaning: a process approach to library and information services*, Westport, Libraries Unlimited, 2004.

<sup>15</sup> Il concetto di contesto è connotato nella definizione stessa di conoscenza, dato che ogni conoscenza è creata e collocata in un particolare contesto: una progettazione efficace del catalogo dovrebbe creare un contesto nel quale l'utente può collocare ogni singola informazione per mezzo della relazione tra le entità costitutive. In tal modo il catalogo si muove dall'essere un sistema di recupero dell'informazione verso uno strumento teso a facilitare la creazione di nuove conoscenze. Recentemente l'FRBR Review Group, il gruppo di studio in vista dell'armonizzazione dei modelli IFLA nella nuova versione consolidata, ha rivisto le funzioni svolte dall'utente che utilizza il catalogo (*funzioni-utente*) sottolineando come l'uso delle relazioni tra le risorse sia finalizzato a porle in un contesto: "Explore. To use the relationships between one resource and another to put them in a context". Cfr. PAT RIVA, *FRBR Review Group initiatives and the world of linked data*, "JLIS.it", 4 (2013), n. 1, p. 105-118, <<http://leo.cilea.it/index.php/jlis/article/view/5482/7893>>.

<sup>16</sup> BRENDA DERVIN, *From the mind's eye of the user: the sense-making qualitative-quantitative methodology*, in JACK D. GLAZIER - RONALD R. POWELL, *Qualitative research in information management*, Englewood, Libraries Unlimited, 1992.

<sup>17</sup> CHARLES R. HILDRETH, *Advancing toward the E3 OPAC: the imperative path*, in NOELLE VAN PULIS, *Think Tank on the Present and Future of the Online Catalog: Proceedings*, Chicago, American Library Association, 1991, p. 39-48 e Id., *Online catalog design models: are we moving in the right direction? A report submitted to*

*the Council on library resources*, August, 1995, <<http://eprints.rclis.org/archive/00000198/>>, oppure <<http://myweb.cwpost.liu.edu/childret/clr-opac.html>>.

<sup>18</sup> <<https://openlibrary.org/>>.

<sup>19</sup> <<http://data.bnf.fr/>>. Il *data model* è visibile in <<http://data.bnf.fr/semanticweb-en>>.

<sup>20</sup> ROMAIN WENZ, *Linked open data for new library services: the example of data.bnf.fr*, "JLIS.it", 4 (2013), n. 1, p. 403-416, <<http://leo.cilea.it/index.php/jlis/article/view/5509/7919>>.

<sup>21</sup> Schema.org (<<http://schema.org/>>) prevede l'utilizzo di dati strutturati nel web tramite i microformati. Il progetto è stato siglato il 2 giugno 2012 dai principali motori di ricerca Google, Bing, Yahoo! e nel novembre dello stesso anno dal motore di ricerca russo Yandex!; il suo scopo è quello di migliorare il recupero semantico da parte dei motori di ricerca tradizionali.

<sup>22</sup> <<http://ogp.me/>>.

<sup>23</sup> <<http://libris.kb.se/>>.

<sup>24</sup> L'introduzione della nuova versione del catalogo Libris XL espressa nel nuovo formato di dati collegati riguarda sia i record nuovi che il pregresso, allo scopo di sostituire MARC21 con un formato di dati più vicino al *data model* delineato in Bibframe. Cfr. HELENA FORSSBLAD, *Uppdaterad information om Libris XL och nya katalogiseringsverktyget*, "Librisbloggen.kb.se". I dati sono pubblicati in JSON-LD (Json for Linking data, <<http://json-ld.org/>>).

<sup>25</sup> Un interessante esempio è offerto dal progetto pilota di integrazione dei dati bibliografici della Library of the University of Amsterdam con un database teatrale del Theatre Institute of The Netherlands. A partire da entità bibliografiche (es. un'opera teatrale) è possibile per l'utente accedere alle informazioni relative alle rappresentazioni teatrali che si sono svolte, recuperando dati di attori, registi e sceneggiatori. Cfr. LUKAS KOSTER, *Linking library and theatre data*, <<http://igelu.org/wp-content/uploads/2011/09/Linking-library-to-theatre-Koster.pdf>>.

<sup>26</sup> Dati.culturaItalia.it è il progetto pilota di CulturaItalia dedicato ai *Linked open data* che raccoglie più di 2,4 milioni di metadati di musei, biblioteche, archivi, gallerie, mostre, monumenti, risorse educative ecc., rilasciati sotto licenza CC0.

<sup>27</sup> Oltre ad Europeana un altro esempio di utilizzo di contenuti di diverso tipo e provenienza nel settore del patrimonio culturale è Mis Museos (<[mismuseos.net](http://mismuseos.net/)>). La piattaforma mostra per ogni artista link a risorse esterne e molti contenuti educativi.

<sup>28</sup> G. ALEMU et al., *Linked data for libraries: benefits of a conceptual shift from library-specific record structures to RDF-based data models*, cit.

<sup>29</sup> MAGNUS PFEFFER, *Automatic creation of mappings between classification systems for bibliographic data*, paper presentato al convegno Semantic Web in libraries, 2013, Amburgo, 25-27 novembre 2013, <[http://swib.org/swib13/slides/pfeffer\\_swib13\\_121.pdf](http://swib.org/swib13/slides/pfeffer_swib13_121.pdf)>.

<sup>30</sup> Per esempio se Dewey e LCSH collegano un elemento geografico della notazione o della stringa all'URI creato da Geonames, i tre *datasets* verranno automaticamente collegati con la possibilità in fase di ricerca di passare da uno schema

all'altro ampliando le possibilità di ricerca e fornendo nuove modalità di visualizzazione. Il collegamento si affida alla proprietà foaf:focus dell'ontologia FOAF. Questa possibilità è tuttora al vaglio del gruppo che in OCLC si sta occupando dello sviluppo dello schema completo in Dewey.info. Cfr. JOAN S. MITCHEL - MICHAEL PANZER, *Dewey linked data: Making connections with old friends and new acquaintances*, "JLIS.it", 4 (2013), n. 1, p. 177-200, <<http://leo.cilea.it/index.php/jlis/article/view/5467>>. Altre applicazioni riguardano Faceted Application of Subject Terminology (FAST) per il Web semantico. Cfr. FAST Linked Data, <<http://experimental.worldcat.org/fast/>>.

<sup>31</sup> <[www.geonames.org/](http://www.geonames.org/)>.

<sup>32</sup> Un esempio di applicativo per realizzare interfacce visuali è Relfinder, <<http://www.visualdataweb.org/relfinder.php>>, che consente di visualizzare tramite l'uso dell'ontologia FOAF le relazioni tra persone. Hanno predisposto interfacce visive anche le biblioteche nazionali di Francia e Spagna.

<sup>33</sup> Virtual International Authority File, <<http://viaf.org/>>.

<sup>34</sup> *Online Computer Library Center (OCLC), Online catalogs. What users and librarians want: an OCLC report*, (a cura di) Karen Calhoun e Diane Cellentani, Dublin (Ohio), OCLC, 2009, <<http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf>>.

<sup>35</sup> Il progetto è stato presentato al recente convegno LOD 2014 il 20-21 febbraio presso l'Archivio Centrale dello Stato.

<sup>36</sup> PATRICK LE BŒUF, *Customized OPACs on the semantic web: the OpenCat prototype*, 2013, p. 1-15, <<http://files.dnb.de/svensson/UILLD2013/UILLD-submission-3-formatted-final.pdf>>.

<sup>37</sup> Il progetto è stato avviato in partnership con la società in-

formatica Logilab e la biblioteca pubblica di Fresnes, una piccola cittadina nei sobborghi di Parigi.

<sup>38</sup> Una demo del catalogo è visibile all'indirizzo <<http://demo.cubicweb.org/opencatfresnes/>>.

<sup>39</sup> <<http://dbpedia.org/>>.

<sup>40</sup> Si ringrazia per la collaborazione Martin Malmsten che ha gentilmente fornito questa anteprima. Cfr. ID, *Decentralisation, Distribution, Disintegration - towards Linked Data as a First Class Citizen in Libraryland*, contributo presentato al Convegno SWIB 2013.

<sup>41</sup> GIANFRANCO CRUPI, *Beyond the Pillars of Hercules: Linked data and Cultural heritage*, "JLIS.it", 4 (2013), n. 1, p. 25-49, <<http://leo.cilea.it/index.php/jlis/article/view/8587/7887>>.

<sup>42</sup> BARBARA B. TILLET, *RDA and the Semantic Web, Linked Data Environment*, "JLIS.it", 4 (2013), n. 1, p. 139-146, <<http://leo.cilea.it/index.php/jlis/article/view/6303/7876>>. Citazione a p. 140.

<sup>43</sup> Per farsi un'idea è possibile provare alcune interrogazioni preimpostate nell'endpoint di una grande *knowledge base* come Dbpedia (<<http://dbpedia.org/sparql>>).

<sup>44</sup> Un esempio interessante proviene dalla Oslo Public Library, che invia all'utente via RSS suggerimenti di lettura estratti dal catalogo e arricchiti dinamicamente tramite il collegamento al web dei dati. Cfr.: ASGEIR REKKAVIK - BENJAMIN ROKSETH, *The Library Catalogue as Linked Open Data. How to Do It and What to Do with It*, paper presentato al convegno SWIB12. Semantic web in libraries, Colonia, 26-28 novembre 2012, <[http://swib.org/swib12/slides/Rekkavik\\_SWIB12\\_110.pdf](http://swib.org/swib12/slides/Rekkavik_SWIB12_110.pdf)>.

DOI: 10.3302/0392-8586-201408-034-1

## ABSTRACT

The paper examines the potential for the application of Linked data to the online library catalogs.

The adoption of the Linked data as a new logical model for expressing bibliographic information, enabling libraries to operate not only a significant change in the metadata with which for decades have described their bibliographic data, but also to find new ways to redesign the search in the OPAC, which represent the interface between the user and the catalog.

Taking a cue from some concrete examples of implementation of innovative search interfaces, the discussion focuses on some aspects of user-catalog that could benefit from the application of linked data to the catalog structure, following a constructivist and user-centric approach and thus entering into the field of study of Information science named Information-seeking behavior, where the search in a catalog is intended as a series of steps in which the user expresses certain behaviors.

The adoption of this technology opens, in fact, the opportunity to review the search models currently in use, too anchored to the systems, in favor of a cognitive information retrieval, that can encourage learning during the information search process, thus transforming the catalog from a document retrieval-oriented system to a system oriented to the creation of new knowledge.