

TEMI E ANALISI

Vecchi paradigmi e nuove interfacce: la ricerca di un equilibrato sviluppo degli strumenti di mediazione dell'informazione

SECONDA PARTE

di Roberto Raieli

Lo storico è un profeta dallo sguardo rivolto all'indietro.
Friedrich Schlegel

Principi biblioteconomici

La valutazione dei nuovi criteri

Prima di considerare con sincero ma immediato ottimismo il rapporto tra il documento esteso a ogni risorsa del nuovo Web e le metodologie e teorie bibliotecarie, occorre che le opinioni più favorevoli e possibiliste finora esaminate siano confrontate con altre opinioni orientate a porre la massima attenzione ai principi biblioteconomici, e tese valutare in modo più cauto e più compiuto le eventuali rivoluzioni di paradigma.

Tra i primi scritti che espongono una presa di coscienza della necessità di preparare una vera accoglienza teoretica ai nuovi metodi e strumenti bibliotecari, ridiscutendo criticamente i paradigmi da porre alla loro base, c'è un intervento di Alberto Petrucciani¹. Di là dai «luoghi comuni» e dalle pratiche 'alla moda' passivamente accettati, l'autore indica come non si possa essere succubi della rivoluzione del mondo dell'informazione nata sul Web, ma si debba invece guidarla nel profondo, traendo spunto per tale discussione dalla presentazione della pubblicazione della risposta di Tho-

ROBERTO RAIELI, Università degli studi Roma Tre, Biblioteca di area delle arti, piazza della Repubblica 10, 00185 Roma, e-mail roberto.raieli@uniroma3.it.

Ultima consultazione siti web: 5 dicembre 2014.

L'autore desidera ringraziare Andrea Marchitelli per i preziosi suggerimenti iniziali, Paola Castellucci per l'accuratezza delle prime letture del testo, Carlo Bianchini e Alberto Salarelli per l'attenta lettura del testo finale. Ogni responsabilità sulla versione definitiva del testo qui pubblicato resta, comunque, interamente dell'autore.

¹ Alberto Petrucciani, *La catalogazione, il mercato e la fiera dei luoghi comuni*, «Bollettino AIB», 46 (2006), n. 3, p. 177-185.



mas Mann² al celebre *report* per la Library of Congress di Karen Calhoun³. I luoghi comuni sulla necessità di cambiare la natura del catalogo, sull'inevitabilità di adeguarsi ai progressi tecnologici della rete, sull'importanza di considerare le nuove esigenze degli utilizzatori del Web, sono «continuamente ripetuti e quasi mai esaminati criticamente», e questo porta a convincersene senza capirne i fondamenti, che andrebbero invece analizzati con lucidità, senza pregiudizi critici dannosi quanto l'acriticità.

Molte delle decisioni prese da organismi 'determinati', come la stessa Library of Congress, sull'onda della semplificazione e dell'alleggerimento dei cataloghi, possono essere preoccupanti dal punto di vista della vera professionalità bibliotecaria, in quanto possono portare all'annullamento dell'identità specifica della biblioteca. Tale identità deve essere invece sempre ben distinta da quella di altri sistemi di aggregazione, gestione e diffusione di dati bibliografici, qualificandosi per la specifica attenzione alla valutazione e verifica dei dati stessi, alla loro correttezza, alla disambiguazione delle informazioni dubbie e in generale alla costruzione di un universo bibliografico chiaro, 'onesto' e rispettoso dell'identità e della natura degli oggetti culturali che rappresenta, oltre a essere 'dedicato' all'utente finale.

Tra i principali bersagli della critica di Mann, sottolinea Petrucciani, c'è il 'modello imprenditoriale', applicato spesso a sproposito alle problematiche delle biblioteche. Il rischio è la creazione di un modello di *management* «autoreferenziale», che imita quello di tipo economico ed è applicato con parecchie forzature, distante dalla realtà della biblioteca in sé. Da questa applicazione diffusa, derivano luoghi comuni come quello della preoccupazione per il poco interesse che il catalogo riceve sul 'mercato dell'informazione', dimenticando che nella sostanza il catalogo non ha mai voluto o dovuto essere una «merce di successo», quanto piuttosto uno strumento di organizzazione, mediazione, ricerca, conoscenza, più o meno difficile da usare, ma percepito proprio dalle persone come tale e non come un'applicazione semplicistica. L'interfaccia di un OPAC o di un discovery tool deve sicuramente essere ricca e amichevole, ma non ha necessità di «scimmiettare» Google, Amazon, o altre interfacce di presentazione di libri nate piuttosto con la preoccupazione di non far perdere 'quote di mercato' alle aziende che rappresentano. Le biblioteche non puntano a un profitto con i loro servizi, non hanno dunque dei veri concorrenti da superare, ma solamente una precisa identità culturale da valorizzare, ponendo attenzione quindi a non sparire dietro i servizi commerciali, ma senza fare compromessi sulla qualità e affidabilità della propria attività.

Inoltre, la fornitura di informazioni non è né l'unico né il principale servizio di qualità fornito dalle biblioteche, essendo prioritari anche i servizi per la lettura e lo studio. A tal proposito, dunque, il problema non è la fruizione a distanza di un'opera che risiede su un sito bibliotecario o di altro tipo, ma piuttosto il ruolo di orientamento e assistenza nella scelta che possono avere le biblioteche, e di conseguenza la costruzione e l'organizzazione delle loro raccolte non solo di posseduto. Sempre Mann nota che, nei casi più comuni, l'utilizzo dei più diffusi strumenti del Web è sufficiente per la ricerca e il recupero delle risorse utili a soddisfare le esigenze più semplici. Le biblioteche, allora, senza temere 'perdite' di utenza di questo genere, non devono 'ridursi' a tentare di fare le stesse cose dei servizi commerciali, devono anzi meglio

2 Thomas Mann, *Il catalogo e gli altri strumenti di ricerca: un punto di vista dalla Library of Congress*, «Bollettino AIB», 46 (2006), n.3, p. 186-206.

3 Karen Calhoun, *The changing nature of the catalog and its integration with other discovery tools: final report*. Washington: Library of Congress, 2006, <<http://www.loc.gov/catdir/calhoun-report-final.pdf>>.

predisporre i propri strumenti per le esigenze più complesse, per cui sono necessari servizi di orientamento specifici e affidabili. Non si devono negare mai le qualità tipiche del servizio bibliotecario, la precisione, la diligenza, in nome di generi di velocità e di semplificazione non appropriati ai suoi principi, richiesti da una visione più 'di mercato' del mondo dell'informazione e delle risorse della conoscenza⁴.

Se nella ricerca di nuove metodologie bibliotecarie, più adeguate alla contemporaneità, non si deve cadere nei luoghi comuni collegati allo spirito del mercato attuale, allo stesso modo si deve guardare criticamente anche ai modelli di tipo tecnologico-sociale dei nostri tempi. Ciò è sottolineato anche in un saggio di Fabio Metitieri, che mette in guardia dai luoghi comuni derivanti dallo spirito del così denominato Web 2.0⁵. Le biblioteche, infatti, entusiasmata dallo spirito della 'Google generation', puntano a volte a riprodurre i modelli *google-like* o quelli dei *social network* per l'organizzazione dei propri OPAC in maniera amichevole e sociale, ma senza affrontare criticamente il problema dei 'contenuti' di tali operazioni. Se anche si può pensare a un modello di OPAC 'collaborativo', tale progetto deve essere però accuratamente pianificato e valutato nelle sue implicazioni con i principi della catalogazione.

Metitieri punta il dito contro i semplicistici fraintendimenti di proposte come quelle di David Lankes e di David Weinberger, secondo le quali gli utenti potrebbero orientarsi nel Web 2.0 utilizzando strumenti accurati e 'conversando' tra loro⁶, o taggando e recensendo 'dal basso' le diverse risorse⁷. In tal maniera sembra superfluo tutto il lavoro bibliotecario di raccolta e conservazione delle risorse, di catalogazione e classificazione, come di informazione bibliografica. In realtà si tratta, invece, solo di scoprire nel profondo il nuovo valore delle attività identificative della qualificata azione delle biblioteche, alla luce dei tanti mutamenti verso il digitale e verso la rete che stanno trasformando il mondo dell'informazione e della conoscenza.

Anzitutto, si deve sottolineare che la stessa generazione di Internet non ha un'adeguata conoscenza della rete, del Web, e delle possibilità degli strumenti e delle risorse disponibili⁸. A questo però, commenta Metitieri, sembra che le biblioteche non rispondano potenziando le proprie attività di *information literacy* – come ci si dovrebbe aspettare –, ma tentino piuttosto di eguagliare e forse superare i mezzi del Web, anziché approfondire e definire, per se stesse e per gli utenti, le differenze specifiche e qualificanti, con il risultato di confondere ancor più le idee delle persone e di confondere se stesse nell'universo della rete. Considerando il modello di Facebook, per esempio, è immediato comprendere che non è uno spazio dove gli utenti amano dialogare con i bibliotecari: i *social network* sono nati, e utilizzati, per ragio-

4 Cfr. Alberto Petrucciani, *La catalogazione, il mercato e la fiera dei luoghi comuni* cit., p. 180-184.

5 Fabio Metitieri, *L'OPAC collaborativo, tra folksonomia e socialità*, «Biblioteche oggi», 27 (2009), n. 2, p. 7-12.

6 V. R. David Lankes [et al.], *Participatory networks: the library as conversation*, «Information research», 12 (2007), n. 4, <<http://www.informationr.net/ir/12-4/colis/colis05.html>>.

7 V. David Weinberger, *Crunching the metadata. What Google print really tells us about the future of books*, «The Boston globe», 13 novembre 2005, <http://www.boston.com/ae/books/articles/2005/11/13/crunching_the_metadata>.

8 In proposito, Metitieri cita il *report*: University college London, *Information behaviour of the researcher of the future*. London: University college London, 2008, <http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf>.

ni personali, lontane dall'uso dei servizi bibliotecari, servizi che le persone stesse si aspettano impostati con altre logiche.

In tal senso, anche il 'catalogo folksonomico', e il *social cataloguing*, non possono essere più che una «leggenda». Le folksonomie e i commenti delle persone si sono sviluppati per ragioni diverse da quelle bibliografiche e catalografiche, i *tag* possono essere errati, confusi, difformi, e le recensioni possono essere poco acute o di parte. Per quanto utili dove non siano disponibili dati bibliografici, le informazioni create dagli utenti possono semmai affiancare i 'prodotti' della catalogazione e della classificazione, ma senza pretesa di sostituirli. Per non parlare del disinteresse delle persone riguardo al 'conversare' tra loro sui documenti dell'OPAC.

I nuovi OPAC e i nuovi sistemi di ricerca e scoperta dell'informazione, dunque, devono pur sempre restare fedeli all'esperienza e alla teoria che ha prodotto i cataloghi e i servizi classici, non per questo, comunque, devono essere lontani dalle nuove tecnologie e dalle nuove esigenze delle persone. Utile può essere l'esempio 'googleiano' per semplificare la schermata della ricerca base o di primo accesso all'OPAC, come il modello di Facebook per connettere eventuali pareri degli utenti e proporli in calce alla scheda catalografica. Una ragionata innovazione delle interfacce può, anzi, sviluppare al meglio alcune caratteristiche dei cataloghi, integrando, per esempio, le funzioni di ricerca con sistemi 'non invasivi' di suggerimento o di correzione degli errori. Per non parlare dell'utilità delle colonne di guida al raffinamento delle ricerche, che consentono di raggruppare i risultati per autore, formato, lingua, oppure tramite *tag cloud*, 'nuvole' di etichette costruite sulla base di soggetti, tesauri o altri KOS. Tali colonne 'a faccette' possono essere determinati qualificatori dell'interfaccia di ricerca bibliotecaria, per quanto non evidenti all'utente generico e non prive di criticità⁹.

Altra questione su cui è necessario andare a fondo, non fermandosi all'acritica accettazione di ciò che il mercato e la tecnologia favoriscono come efficace soluzione anche nell'ambito biblioteconomico, è quella dell'ordinamento dei risultati delle ricerche. Una questione non nuova, che nella logica dello sviluppo dei nuovi sistemi di ricerca torna ad avere un posto centrale¹⁰. Le interfacce dei nuovi OPAC e dei discovery tool, infatti, nella loro tendenza a mostrare un'unica e semplice lista di tutti i risultati raccolti attorno al 'tema' della ricerca, li presentano in un unico elenco, quasi senza distinzione, se non grafica, relativa alla tipologia della risorsa, alla provenienza locale o remota, al database di appartenenza o al Web. Un ordine serve comunque fornirlo alla lista spesso lunga e 'decontestualizzata' – alla Google –, e quando non è un ordine materiale per data, autore o titolo, non resta che quello per 'rilevanza'¹¹. La riunione di risorse di natura diversa in uno stesso elenco, anziché semplificare la ricerca dei meno o dei più esperti, può confondere le idee degli utenti, e un affidabile sistema di *relevance ranking* può essere la soluzione adeguata anche per i professionisti dell'informazione.

9 Cfr. Fabio Metitieri, *L'OPAC collaborativo, tra folksonomia e socialità* cit., p. 10.

10 Vedere le indicazioni contenute in: Holly Yu; Margo Young, *The impact of Web search engines on subject searching in OPAC*, «Information technology and libraries», 23 (2004), n. 4, p. 168-181; California digital library, *The Melvyl recommender project final report*. 2006, <http://www.cdlib.org/services/publishing/tools/xf/melvyl_recommender/report_docs/Mellon_final.pdf>; Karen Calhoun, *The changing nature of the catalog and its integration with other discovery tools* cit.

11 Riguardo le tecniche più coerenti con le necessità dei primi OPAC, e il semplice ordinamento alfabetico dei risultati, v. Jeffrey Beall, *The value of alphabetically-sorted browser displays in information discovery*, «Library collections, acquisitions & technical services», 31 (2007), n. 3/4, p. 184-194.

La possibilità e l'affidabilità di modalità di ordinamento che prevedano la disposizione dei risultati delle ricerche in base a una «presunta rilevanza» per l'utente sono discusse in un saggio di Maria Teresa Biagetti, che indica anzitutto come troppo semplicisticamente le tecniche basate sulla frequenza dei termini, utilizzate dai motori di ricerca e dalle librerie online, siano applicate anche agli OPAC¹². Il *relevance ranking*, presentazione dei risultati secondo una graduatoria di rilevanza presunta, e il *relevance feedback*, riformulazione semi-automatica della query in base a indicazioni esplicite di rilevanza fornite dall'utente, sono basati su tecniche di trattamento dei documenti *full-text* sviluppate nell'ambito dell'information retrieval, che si fondano sui rapporti tra la frequenza dei termini nel testo pieno di ciascun documento e la frequenza 'inversa' dei termini nell'intero database che li contiene¹³. L'applicazione all'OPAC di tali tecniche è di fatto meno efficace, dato che i campi dei record contengono pochi termini, per quanto essenziali, e comunque tecniche di questo tipo non riescono a tenere mai conto dei rapporti semantici tra i termini, di legami e relazioni significativi, considerandoli come a sé stanti.

Le graduatorie di rilevanza, calcolate in base al computo della frequenza o anche all'analisi della prossimità dei termini, non possono soddisfare realmente le esigenze di persone che sono interessate pure al contesto concettuale in cui le informazioni sono inserite, e che cercano argomenti con un senso preciso e aspetti specifici. Per questo Biagetti sottolinea che dopo l'analisi critica del *relevance ranking* serve sviluppare un'analisi teorica dei concetti di rilevanza e di pertinenza, in quanto legati allo «stato delle conoscenze» delle singole persone. In tal modo si potranno definire nuove funzionalità di ricerca semantica basate su strategie di indicizzazione che colgano la varietà, la ricchezza e la complessità di approfondimento degli argomenti, in grado di proporre risultati veramente coerenti con le necessità di conoscere¹⁴. In tale prospettiva, non è possibile non subire il fascino dei progetti di trasposizione dei dati bibliografici nel Web semantico, che consentirebbero la ricerca semantica negli OPAC secondo le modalità dei linked data. È necessario, però, adottare alcune cautele, riflettendo su quale sia il 'livello' dei dati semantici che possono già essere disponibili per l'utente, e progettando un nuovo livello di indicizzazione meglio rispondente alle caratteristiche di una 'riscoperta' pertinenza multidimensionale, soggettiva e dinamica¹⁵.

La problematica definizione di risorsa

Come conviene essere cauti dinanzi ai luoghi comuni sull'innovazione delle interfacce, si deve essere disincantati anche dinanzi alle varie proposte di andare 'oltre' il documento – inteso in senso più classicamente bibliografico-biblioteconomico.

12 Maria Teresa Biagetti, *Nuove funzionalità degli OPAC e relevance ranking*, «Bollettino AIB», 50 (2010), n. 4, p. 339-356.

13 Per il chiarimento delle tecniche di misurazione, cfr. *ivi*, p. 342-345. Vedere anche: Karen Spärck Jones, *A statistical interpretation of term specificity and its application in retrieval*, «Journal of documentation», 28 (1972), n. 1, p. 11-21.

14 Biagetti espone il valore e la 'relatività' della rilevanza e della pertinenza riferendosi principalmente alle teorie di Tefko Saracevic e Birger Hjørland: cfr. Maria Teresa Biagetti, *Nuove funzionalità degli OPAC e relevance ranking cit.*, p. 349-353.

15 In proposito agli studi per la costituzione di biblioteche digitali 'sematiche' tramite la realizzazione di metadati semanticamente arricchiti, v. Maria Teresa Biagetti, *Sviluppi e trasformazioni delle biblioteche digitali: dai repositories di testi alle semantic digital libraries*, «AIB studi», 54 (2014), n. 1, p. 11-34.

Molte di queste proposte, infatti, si riferiscono a un'idea generale di risorsa che ancora non è stata a pieno definita, oppure partono dall'ipotesi di una possibile disaggregabilità e riaggregabilità delle risorse quasi incondizionata e orientata a fini molto differenti.

Riguardo i concetti di identità e valore del documento, individuabili o meno in nuove tipologie di risorse dell'informazione meno o più 'ampie', egualmente o maggiormente 'variabili', sono da considerare alcune riflessioni di Paola Castellucci¹⁶. La constatazione da cui prende avvio la discussione è quella che in molti programmi avanzati di diffusione dell'informazione in rete, come quello dell'*open access*, l'idea di metodologie di ricerca «document centered», fondate sull'idea stessa di documento, possa sembrare retaggio del passato, contraria al progresso della conoscenza. Ad avere posizione sempre più centrale, infatti, è il concetto di dato, estraibile non solo dai documenti 'canonici', ma da tutta una serie di altre risorse che crescono di importanza, e devono essere oggetto della ricerca.

Per esempio, nel progetto Open annotation di Herbert Van de Sompel¹⁷, l'attenzione è posta sui «nanodocumenti», o «nanopublications», quali annotazioni, commenti, *post*, *tweet*, che possono contenere informazioni preziose, molti dei quali da sempre in uso nella pratica della comunicazione scientifica, ma marginalizzati rispetto ai documenti principali a stampa, e quindi ad alto rischio di sparizione. La tecnologia della stampa ha favorito i documenti aventi una data «massa», non considerando le annotazioni e i commenti: in tale logica, addirittura, può essere una colpa della stampa «privilegiare il segno autoriale piuttosto che la traccia di percorsi di lettura». Le nuove tecnologie digitali e di rete, invece, possono ristabilire questo rapporto. In tutto ciò, è un fatto che una serie di informazioni, contenute in annotazioni, in *post* o *tweet*, o in altri nanodocumenti, cresce continuamente di importanza, dando autorevolezza a nuove forme autoriali, che si esprimono pubblicando in rete nuove risorse quali linee di ricerca, progetti, riflessioni, immagini, filmati e dati di ogni genere. Il singolo dato può essere estratto da ognuna di tali risorse, per quanto 'nana', ed essere condiviso nella ricerca scientifica. Inoltre, un discorso speculare si può fare per i «megadocumenti», o documenti «enhanced», risorse potenziate come «hub» in un sistema integrato di dati e servizi.

Ovviamente, scrive Castellucci, l'eventuale questione di quale debba essere la massa di una risorsa per essere considerata valida – come il documento 'tradizionale' – non riguarda gli aspetti quantitativi, ma esclusivamente quelli qualitativi. Per esempio, non ogni elenco di dati bibliografici aggregati può essere una bibliografia, a caratterizzare una bibliografia come risorsa di valore scientifico sono la sua strutturazione o il suo indice¹⁸. In ogni caso, un discorso relativo alla massa per validare una risorsa deve riguardare un totale ripensamento dei generi: devono essere ridefiniti i modi di produzione, validazione e diffusione delle informazioni e delle conoscenze. Nel sistema della produzione scientifica deve essere ben valutato il coinvolgimento di soggetti creatori e prodotti della creazione prima non considerati. Oltre ciò, una ponderata visione della questione della validazione, nonché della conservazione, richiama un'altra categoria fisica oltre la massa: il tempo. In questo senso,

¹⁶ Paola Castellucci, *Tempo e massa: una nuova energia nella comunicazione scientifica*, «Bollettino ALB», 51 (2011), n. 3, p. 237-244.

¹⁷ <<http://www.openannotation.org>>.

¹⁸ In proposito alla possibilità di 'espandere' il concetto di bibliografia e la relativa struttura, v. David G. Hendry [et al.], *Collaborative bibliography*, «Information processing & management», 42 (2006), n. 3, p. 805-825.

l'altro progetto di Van de Sompel, Memento¹⁹, si propone di risolvere il problema della volatilità delle risorse. Aggiungendo una marca temporale anche a semplici segmenti di risorse in rete, il sistema consente di risalire alle loro versioni succedutesi nel tempo, di verificare la primogenitura nelle citazioni dei risultati di una ricerca, nonché di ricontestualizzare i nanodocumenti nei luoghi originari. In tal maniera, si può offrire la possibilità di avere una dignità scientifica anche a risorse volatili, come molte pagine web continuamente aggiornate.

Nelle conclusioni, dunque, di là dal facile entusiasmo per la libera e democratica comunicazione in rete, accettare il ruolo di una data quantità di nuove risorse significa attuare una vera rivoluzione culturale, di cui devono prendersi carico gli studiosi dell'area LIS, approfondendo e «rinegoziando» tutti i concetti collegati alla tradizionale idea di documento, collegati ai processi di produzione, conservazione, valutazione e disseminazione della ricerca, collegati alle possibilità di relazione tra autori e lettori, e collegati al riuso dei dati²⁰.

Oltre alla concezione di nuove tipologie di risorse informative, anche il concetto 'classico' di documento deve essere riesaminato attentamente, in quanto centro di tutti i discorsi sui sistemi di trattamento, diffusione e ricerca dell'informazione e della conoscenza. Così come si possono valutare diverse visioni comunque ottimiste, ma più caute e ponderate, riguardo alla strutturazione e l'utilizzo dei nuovi sistemi di ricerca e delle nuove risorse, allo stesso modo, riferendosi alle imprescindibili condizioni di integrità e autorevolezza dei documenti – di qualunque genere – che sono nelle biblioteche e nel Web, si può tenere in conto un'analisi di Alberto Salarelli che riguarda i linked data e il Web semantico²¹.

Se l'adozione dello schema RDF per individuare la natura e gli attributi degli elementi che costituiscono una risorsa, al fine di consentire la creazione di legami semantici tra i dati di diverse risorse, può implicare tramite la granularizzazione dei suoi dati la «destrutturazione» della risorsa originaria, il rischio è quello che l'evoluzione del concetto di documento porti a intenderlo come qualcosa di troppo differente da come è stato costituito in secoli di riflessione sui suoi principi. Gli utilizzatori dei contenuti della risorsa cui fanno riferimento i dati granularizzati potrebbero iniziare a vedere i documenti 'coerenti' originari, nati digitali o meno, quasi solo come 'contenitori' di contenuti da disaggregare, per creare intorno a questi percorsi semantici spesso non 'forti', per riaggregare ogni volta liberamente nuovi contenitori, senza un'opportuna considerazione dei vincoli semantici che reggono i contenuti originariamente creati fuori da tale logica di riutilizzo, o di «ri-significazione», creati in base ad altri schemi di strutturazione del 'senso'. L'attenzione verrebbe posta, in tal caso, solo sull'«informazione» intesa in generale e in assoluto, e non sulle 'fonti' della conoscenza, sulla rete dei collegamenti e non sui contenuti collegati, sui dati e non sulla loro interpretazione, sull'«intelligenza collettiva» e non sul lavoro intellettuale del singolo. In ogni modo, indica Salarelli, i linked data rappresentano una forma «proficua» di trattamento delle diverse risorse, proprio perché consentendo di correlare i contenuti conservati in differenti archivi permettono la realizzazione di strumenti informativi nuovi, anche se al momento è difficile «tentare di deno-

19 <<http://www.mementoweb.org>>.

20 In proposito, Castellucci cita anche le considerazioni di George Landow: v. George P. Landow, *Hypertext 3.0: critical theory and new media in an era of globalization*. Baltimore: The Johns Hopkins University, 2006.

21 Alberto Salarelli, *Sul perché, anche nel mondo dei Linked Data, non possiamo rinunciare al concetto di documento*, «AIB studi», 54 (2014), n. 2/3, p. 279-293.

minare con terminologie mutuete dal passato» nuove tipologie di risorse ancora da ‘valutare’ approfonditamente²².

È difficile, dunque, continuare a usare il termine ‘documento’ per intendere a pieno le nuove risorse dell’informazione, che si possono distinguere, forse, solo in base al ‘valore informativo’ del contenuto veicolato in contenitori sempre più variabili e dinamici. Oltre il problema di valutare oggettivamente l’affidabilità o la scientificità della risorsa, quindi, si presenta il problema della ‘fiducia’ accordata dallo stesso autore all’integrità della propria creazione: se cioè, nelle proprie intenzioni, il creatore l’abbia strutturata per mantenere una ‘non decostruibile’ identità, all’interno e con le ‘regole’ di un ambito di utilizzazione specifico, o sia stata progettata per la decostruzione e riutilizzazione dei contenuti, in qualunque ambito, con il rischio di una certa debolezza della sua stabilità semantica primaria. In quest’ottica di condivisione dei patrimoni di informazioni e conoscenze, ogni organizzazione culturale responsabile di un dato settore non deve perdere il ruolo e il compito della verifica di tali patrimoni. Le diverse organizzazioni della cultura dovranno «cambiare mentalità e procedure per operare in senso aperto», in modo da consentire alle diverse persone il libero ma consapevole riutilizzo degli *open data* tramite i LOD, vigilando che non si disperdano il contesto e il senso originario di ogni risorsa, il contesto e il senso dell’insostituibile «sostrato documentario», dei dati, dei metadati, favorendo valide e ‘senseate’ nuove aggregazioni, inserite in nuovi contesti²³.

Proseguendo l’analisi, Salarelli discute gli elementi critici del linked data rispetto alla generale teoria del documento, in considerazione dell’importanza di stabilire diversi gradi di autorevolezza e stabilità delle risorse – questione di orientamento epistemologico e ontologico il cui ‘rilancio’ è diventato imprescindibile nell’attuale realtà della comunicazione in rete. Nello specifico, l’aspetto su cui è posta l’attenzione è la logica dell’organizzazione dei dati riferiti ai contenuti di una risorsa in base ai fini della sua creazione. Quanto più i fini sono «dichiarativi», tanto più è determinate il ruolo dell’autorità che ha strutturato i contenuti ed espresso ‘dichiarazioni’, in forma discorsiva come in forma di tripla RDF, esprimendo dunque anche un’idea di valore dei contenuti e della forma di espressione. L’individuazione dell’autorità del *creator* non è un problema nei documenti ‘tradizionali’, costruiti seguendo regole e architetture che ne garantiscono la valutabilità. La questione nasce rispetto alla più libera logica di costituzione delle nuove tipologie di risorse che se, ovviamente, non possono giungere alla decostruzione delle risorse di partenza, le quali mantengono in sé il proprio grado di integrità e coerenza, possono però destrutturarle nella ‘percezione’, disaggregandone i contenuti attraverso i dati atomizzati che a essi si riferiscono e riaggregandoli in un nuovo insieme che ha una coerenza non ben definita, realizzabile da ‘chiunque’ anche senza caratteristiche specifiche di autorità. È spesso un problema di identità del creatore, dunque, tanto di chi definisce le ontologie quanto

22 Cfr. *ivi*, p. 280.

23 Cfr. *ivi*, p. 280-283. A proposito del ruolo delle ‘organizzazioni responsabili’, Salarelli cita il «contrasto» tra posizioni di tipo *top down* e posizioni *bottom up* collegate alle discussioni sulla transizione da MARC a BIBFRAME: v. Kim Talleraås, *From many records to one graph: heterogeneity conflicts in the linked data restructuring cycle*, «Information research», 18 (2013), n. 3, <<http://InformationR.net/ir/18-3/colis/paperC18.html>>. Riguardo il rapporto tra il ‘sostrato’ dei documenti coerenti e autonomamente dotati di senso e lo ‘strato’, «layer», delle relative connessioni aggiunte dai linked data, Salarelli cita: Christian Bizer; Tom Heath; Tim Berners-Lee, *Linked data: the story so far*, «International journal on semantic Web and information systems», 5 (2009), n. 3, p. 1-22.

di chi utilizza gli schemi, la cui autorità in rete sembra dissolversi eccessivamente rispetto a qualunque sistema scientifico, per quanto aperto.

Altri elementi critici, in tutto ciò, sono la «provenienza» e la «granularità» dei dati. Il problema della provenienza, e dell'identificabilità, dei dati dipende dalla possibilità di definire la responsabilità della loro creazione: «del dove, del quando e del perché essi sono stati creati»²⁴. Il problema della granularità, e dell'interpretabilità, dei dati riguarda il livello limite di atomizzazione cui può essere sottoposta una risorsa tramite l'atomizzazione della sua descrizione senza che si disperdano il suo senso e il suo significato: limite identificabile nel «concetto di 'molecola RDF', un'entità intermedia tra il documento originario e le singole triple». Senza la soluzione dei punti problematici il sistema dei linked data rischia di essere di limitata affidabilità, sia nel tentativo di stabilire relazioni articolate tra le risorse in base ai dati estratti da esse e tra tali dati e le risorse di provenienza, sia, ancor più, nel tentativo di costruire nuove risorse come dossier o «agglomerazioni» dei contenuti e dei dati estrapolati da altre diverse risorse e in vario modo correlati²⁵.

In conclusione, anche nei nuovi sistemi della rete e delle risorse digitali si deve sviluppare la ricerca di soluzioni per garantire ai dati e ai relativi costrutti una validazione di autorità, di datazione, di integrità, nonché di coerenza semantica, secondo principi riferibili al mondo dei documenti 'classici'. La necessità è quella di contenere la «liquidità dell'informazione», la decontestualizzazione e la reinterpretabilità perpetue, tramite sistemi dotati di quell'affidabilità caratteristica delle metodologie della tradizione bibliografica. La realtà sociale stessa si fonda sull'affidabilità di determinati documenti – in senso 'stretto' e archivistico – quali leggi, patti, trattati, e nondimeno quelli della conoscenza. Per le possibili categorizzazioni di risorse, di 'vecchio' o 'nuovo' tipo, il principio non cambia: la validità è collegata alla possibilità di trasformare la volatilità dell'informazione in «aggregazioni di dati sufficientemente stabili», utili come basi della conoscenza per il progresso intellettuale e sociale.

Limiti dell'information discovery

L'affidabilità dell'information retrieval

Sono necessarie, infine, alcune considerazioni sui limiti dell'«information discovery», come è possibile definire l'ambito generale di applicazione dei *web-scale discovery service* di cui sono stati discussi finora i vantaggi e gli svantaggi.

I sistemi basati sui LOD, le ipotesi di *next generation library management*, le varie *discovery solution*, e i più avanzati sistemi di discovery tool, non hanno il 'rigore' dell'information retrieval cui hanno abituato le banche dati, né quello degli OPAC e altri strumenti di riconosciuta affidabilità. I discovery tool possono essere definiti strumenti di ricerca 'di base', di avvio, le interfacce 'proprietarie' delle banche dati, di contro, sembrano essere lo strumento migliore con cui fare ricerca specia-

²⁴ Anche se la terza regola dei linked data – raccomandando «When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)» – tende a impedire che si diffondano dati che non siano qualificati con le specifiche necessarie: cfr. Tim Berners-Lee, *Linked data*. 2006, <<http://www.w3.org/DesignIssues/LinkedData.html>>.

²⁵ Cfr. Alberto Salarelli, *Sul perché, anche nel mondo dei Linked Data, non possiamo rinunciare al concetto di documento* cit., p. 286-288. Sul concetto di RDF molecule, v. Li Ding [et al.], *Tracking RDF graph provenance using RDF molecules (Technical report TR-CS-05-06, Computer science and electrical engineering, University of Maryland, Baltimore County, 30 April 2005)*. 2005, <ftp://www.ksl.stanford.edu/local/pub/KSL_Reports/KSL-05-06.pdf>.

listica, utilizzando metadati e campi dei record appositamente definiti, e soprattutto appositi linguaggi di indicizzazione. I discovery tool perdono molte funzioni di interrogazione avanzate, rendendo il percorso di ricerca 'omogeneizzato' al livello più 'basso', ma la soluzione migliore non è negare i nuovi strumenti e restare alle interfacce specializzate per i livelli più alti. Si tratta di riflettere sull'effettivo valore degli strumenti in rapporto alle persone che li devono usare, ma anche rispetto ai contenuti cui si applicano.

Se si parla di 'scoperta' dell'informazione, ci si riferisce in generale a un insieme di atteggiamenti, pratiche e strumenti che mirano a un tipo di ricerca 'esplorativa', in cui non è ben definito l'argomento della ricerca, e spesso non è chiaro l'eventuale linguaggio specifico, le parole chiave, o i termini di indicizzazione. In questo caso, l'aiuto di un sistema non complesso e ricco di funzioni guida è importante per instradare una persona, che si trova, appunto, a 'scoprire' dati, informazioni e risorse che non immaginava esistessero o fossero di suo interesse, di carattere generale o anche di tipo specifico. Nell'ambito di una ricerca 'consapevole', invece, si ha un chiaro obiettivo di informazione, o anche una risorsa precisa da 'reperire', così come una certa padronanza degli strumenti, dei termini e dei linguaggi di ricerca. Quello che serve, allora, è un sistema avanzato e preciso, che permetta uno sviluppo e una crescita di quello che già si sa o si possiede²⁶. In tutto questo, ovviamente, è sempre salvo il ruolo dell'operatore umano che può guidare l'utente interagendo con il sistema, inoltre non si possono definire con nettezza categorie di utenti e di comportamenti di ricerca, potendo ognuno ogni volta avere esigenze diverse e diversa padronanza dei temi e dei contenuti. In più, una persona che può inizialmente stare in una 'categoria' meno preparata alla ricerca e che usa sistemi di information discovery, può presto acquisire padronanza e informazioni tali da passare al 'livello' superiore dove ha bisogno di specifici strumenti di information retrieval²⁷.

Rispetto alla solida tradizione dell'information retrieval²⁸, l'ambito dell'information discovery è più sfuggente e più complesso da definire nella sua varietà e vaghezza²⁹. L'espressione generica information discovery – quando non indica comunemente anche l'information retrieval – è, però, sempre più usata per indicare l'i-

26 Cfr. Mike Sweet, *There's nothing wrong with discovery services that can't be fixed by the reference layer*, in *The Fiesole collection development retreat, April 12, 2012*, 2012, <http://www.casalini.it/retreat/2012_docs/sweet.pdf>.

27 Per tacere del percorso opposto, dove chi è o diventa esperto di un campo si ritrova allo stadio iniziale in un altro campo. Quindi, sia la persona che sale di 'livello', lo studioso che è 'novizio' in altro ambito, come anche lo stesso studioso nel proprio settore, possono trarre vantaggi dagli strumenti amichevoli. Per esempio, v. Tod A. Olson, *Utility of a faceted catalog for scholarly research*, «Library hi tech», 25 (2007), n. 4, p. 550-561.

28 Tra le varie trattazioni, v.: Kalervo Järvelin, *Information Retrieval (IR)*. In: *International encyclopedia of information and library science*, edited by John Feather, Paul Sturges. London: Routledge, 2003, p. 293-295; Christopher D. Manning; Prabhakar Raghavan; Hinrich Schütze, *Introduction to Information Retrieval*. Cambridge: Cambridge University press, 2008; Frederick W. Lancaster, *Indexing and abstracting in theory and practice*. Urbana Champaign: University of Illinois – Graduate school of LIS, 2003; Ricardo Baeza-Yates; Berthier Ribeiro-Neto, *Modern Information Retrieval*. New York: Longman, 2000; Robert R. Korfhage, *Information storage and retrieval*. New York: Wiley, 1997; Lauren Doyle; Joseph Becker, *Information Retrieval and processing*. Los Angeles: Melville, 1975.

dea delle nuove possibilità di scoperta dell'informazione e della conoscenza collegate ai *web-scale discovery service*, anche se la terminologia relativa ai nuovi strumenti e alle trasformazioni da essi portate è ancora in fase di definizione. La ricerca intorno ai sistemi di scoperta *web-scale*, comunque, è in fase di sviluppo, soprattutto negli ultimi due o tre anni, e si occupa di aspetti non solo applicativi³⁰.

Discutendo una serie di questioni in un'ottica mista tra l'applicativo e il programmatico, con qualche apertura più teorica, la letteratura degli ultimi anni ha investigato i limiti e l'efficacia dell'information discovery. In generale, i *discovery service*, oltre all'ampliamento dello 'spazio' di scoperta, sono in grado di «aggiungere valore» alla ricerca attraverso la facilità di utilizzo, il criterio dello *one-stop shop*³¹, i filtri e le faccette, e la ricchezza di informazioni collegate. Allo stesso modo, sono confermate alcune tipiche perplessità originarie, quali l'eccesso di risultati proposti, la scarsa rilevanza di molti di essi, la perdita di risultati utili, la mancanza della specificità e funzionalità tipica dei cataloghi e dei database³². Fermo restando che il processo di scoperta di informazioni utili riguarda complesse combinazioni dei rapporti tra ricercatore, strumenti di ricerca e risorse cercate, l'efficacia dei *discovery service* deve essere accuratamente testata, oltre e indipendentemente dal «comune elogio» della loro facilità di utilizzo. L'information discovery è «un'esplorazione in continuo sviluppo», e un processo iterativo, che precede e prepara l'accesso alle risorse o alle informazioni utili alla persona che si immette in tale percorso senza un'idea chiara delle proprie necessità. Inoltre, proprio da tale processo, sia le persone comuni sia gli studiosi possono apprendere ulteriori conoscenze relative al tema di ricerca, generale o specifico, familiare o nuovo che sia. Per questo, tutti i *web-scale discovery service* devono essere efficaci e soddisfacenti per ogni tipo di ricerca individualmente parametrizzata, offrendo strumenti sofisticati accanto a quelli semplici, filtri a faccette accanto ad ampi indici centrali³³.

Pur nella positività di una loro adeguata implementazione e di un opportuno impiego, i *web-scale discovery service* sembrano comunque restare limitati, almeno per il momento, all'essere strumenti di base per una prima univoca ricerca e scoperta delle risorse mediate da una biblioteca. Ogni biblioteca dovrà preoccuparsi di stabilire e 'insegnare' agli utenti le potenzialità e l'utilità dei nuovi strumenti in rapporto a quelli tradizionali, di spiegare loro le differenze che li fanno restare 'distinti' per

29 In proposito, v.: Hillary A. H. Richardson, *Revelations from the literature: how web-scale discovery has already changed us*, «Computers in libraries», 33 (2013), n. 4, p. 12-17; Priscilla Caplan, *On discovery tools, OPACs and the motion of library language*, «Library hi tech», 30 (2012), n. 1, p. 108-115; Henderik A. Proper; Peter D. Bruza, *What is information discovery about?* «Journal of the American society for information science», 50 (1999), n. 9, p. 737-750.

30 Come si può notare anche nei saggi contenuti in: *Planning and implementing resource discovery tools in academic libraries*, edited by Mary Pagliero Popp, Diane Dallis. Hershey: Information science reference, 2012.

31 Nicholas Joint, *The one stop shop search engine: a transformational library technology?* «Library review», 59 (2010), n. 4, p. 241-248.

32 Cfr. Beth Thomsett-Scott; Patricia E. Reese, *Academic libraries and discovery tools: a survey of the literature*, «College & undergraduate libraries», 19 (2012), n. 2/4, p. 123-143.

33 Nadine P. Ellero, *Integration or disintegration: where is discovery headed?*, «Journal of library metadata», 13 (2013), n. 4, p. 311-329. Nel saggio molte conclusioni sono tratte da una compiuta rassegna della letteratura sui *web-scale discovery service*.

distinti scopi, così come di spiegare le strategie di ricerca, la valutazione critica dei risultati, e i rischi dell'eccessiva disintermediazione³⁴.

L'obiettivo che i discovery tool rendano al meglio la propria funzione, che può essere organizzata come complementare a quella degli altri strumenti peculiari della biblioteca, dipende in ogni caso – come per quanto riguarda gli OPAC o i database – dalla qualità dei metadati e dei dati cui tali sistemi sono applicati. Per la strutturazione di guide realmente utili a orientarsi nell'ampio insieme dei riferimenti trovati e proposti dal sistema di ricerca, ovvero per la creazione di filtri a faccette in grado di raccogliere i riferimenti per soggetto, per classificazione, come per autore, per collana, per editore eccetera, sono necessari metadati ben strutturati, controllati e affidabili³⁵. I filtri a faccette, spesso proposti in colonne di orientamento ai fianchi del lungo elenco dei risultati di ricerca, hanno le potenzialità per essere uno degli strumenti più ricchi e utili dei *web-scale discovery service*, in grado di guidare le persone comuni come gli utenti più esperti. Questi strumenti seguono e potenziano la logica di consultazione degli OPAC e dei database, e si basano sulla disponibilità di ottimi dati catalografici e bibliografici, estratti e verificati con cura ed esperienza. La qualità della ricerca o navigazione a faccette di un discovery tool, quindi – pur collegata alle capacità degli algoritmi di trattamento dell'indice di ogni sistema – è strettamente dipendente dalla qualità del lavoro di metadattazione delle risorse cui si applica, dalla ricchezza, completezza e affidabilità degli elementi con cui sono poi indicizzate le risorse stesse³⁶.

Il grado di accuratezza del *faceted browsing*, inoltre, e la sua affidabilità, sono collegati anche alla possibilità di 'condividere' tra le diverse risorse indicizzate schemi e tesauri per una metadattazione il più possibile 'omogenea' oltre che di qualità. Non solo è importante quanto è alta la percentuale di risorse che hanno una metadattazione completa, soprattutto relativamente ai soggetti o alle classi, ma anche quanto un sistema di classificazione, per esempio, possa essere usato concordemente per collezioni diverse di risorse, o quanto possa essere adottato un tesoro generale nato dalla 'fusione' di altri più specifici³⁷. La mancanza di schemi comuni per la creazione di metadati impedisce la messa a punto di valide funzioni di *faceted search*, e se i discovery tool che indicizzano metadati completi e accurati possono permettere un'efficace combinazione di richiamo e precisione nella strategia di ricerca, funzionali filtri e raggruppamenti dei risultati, un affidabile *dynamic browsing* e un'efficace *serendipity* nell'information discovery, di contro il 'mescolamento' nell'indice di differenti schemi di classificazione o soggettazione può restituire risultati di ricerca

34 Sulla necessità di questa rinnovata *information literacy*, v.: Jody Condit Fagan, *Discovery tools and information literacy*, «Journal of web librarianship», 5 (2011), n. 3, p. 171-178; Nancy Fawley; Nikki Krysak, *Information literacy opportunities within the discovery tool environment*, «College & undergraduate libraries», 19 (2012), n. 2/4, p. 207-214.

35 Vedere per esempio: Susan C. Wynne; Martha J. Hanscom, *The effect of next-generation catalogs on catalogers and cataloging functions in academic libraries*, «Cataloging & classification quarterly», 49 (2011), n. 3, p. 179-207; Michael Kelley, *Coming into focus: web-scale discovery services face growing need for best practices*, «Library journal», 137 (2012), n. 17, p. 34-40.

36 Il limite delle guide a faccette, dunque, è il genere di risorsa a cui sono applicate: se pure il filtro si applica a tutte le risorse trattate nell'indice unico del discovery tool, molte di esse possono non presentare tutti i dati utili e corretti per un efficace raggruppamento e *faceted browsing*.

37 V. Jody Condit Fagan, *Discovery tools and information literacy* cit.

confusi, liste di risorse decontestualizzate dall'ambito disciplinare, e disorientare il ricercatore anziché guidare la sua scoperta³⁸.

Oltre la facilità d'uso e l'univocità 'superficiali' dei *web-scale discovery service*, dunque, è necessaria una nuova strutturazione in profondità, che renda realmente efficace e affidabile una delle più importanti funzioni dei discovery tool: la ricerca guidata e filtrata tramite faccette. È necessario uno sforzo comune tra tutti coloro che sono coinvolti nei processi di produzione, raccolta, trattamento, indicizzazione e ricerca delle risorse, per comprendere nel profondo le ragioni e le funzioni della creazione e dell'uso di metadati delle risorse che consentano di rappresentarle anche semanticamente. Tra gli sforzi tesi a creare infrastrutture condivise per supportare ampi e ricchi sistemi di ricerca semantici, emergono quelli per la messa a punto dell'architettura dei linked data. Tale sistema punta infatti alla creazione di un complesso condiviso, potenziato e trasparente di collegamenti tra i dati, che rendano la propria semantica automaticamente esplicita, senza la mediazione di metadati atti a catturarne i significati o gli ambiti. Si tratta dello stesso sforzo attraverso cui anche le biblioteche mirano a integrare i dati dei cataloghi nel Web, condividendo dati e 'significati' ricchi e autorevoli che possono fare da base per un'affidabile sistema di information discovery³⁹.

Fino a quando non sarà possibile raggiungere un dato livello di ricchezza e interoperabilità dei metadati, oppure la diffusione di una solida e condivisa architettura di linked data⁴⁰, i discovery tool tenderanno a qualificarsi come uno strumento di primo accesso all'insieme delle informazioni e alle risorse, e l'information discovery resterà limitato a essere una ricognizione preliminare, da precisare poi con l'uso di strumenti di ricerca specifici, tipici dell'information retrieval. Come l'esplorazione complessiva di un insieme potenzialmente infinito di risorse, anche la ricerca univoca in un ambito molto ampio è difficile da realizzare fino a quando strumenti distinti, pur di elevata qualità, si applicheranno a dati diversamente trattati, pur con metodi molto affidabili. Per quanto sia un traguardo ancora lontano, con i connotati dell'utopia, un ambiente di dati e metadati «comune, standardizzato, aperto e integrato», allargato al Web, può rappresentare il giusto spazio, 'universale', per l'applicazione di sistemi di ricerca e scoperta unici e 'onnicomprensivi', adeguati e utili per tutte le persone, per tutte le esigenze di informazione e conoscenza⁴¹.

La creatività della scoperta

La condivisione, l'interoperabilità e la riutilizzazione dei dati appaiono, dunque, come

38 William Breitbach, *Web-scale discovery: a library of Babel?* In: *Planning and implementing resource discovery tools in academic libraries* cit., p. 637-645.

39 Tra i vari scritti che affrontano la questione: Marshall Breeding, *Linked data: the next big wave or another tech fad?* «Computers in libraries», 33 (2013), n. 3, p. 20-22; Lorcan Dempsey, *Thirteen ways of looking at libraries, discovery, and the catalog: scale, workflow, attention*, «Educause review online», 10 dicembre 2012, <<http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention>>; Library of Congress, *Bibliographic framework as a web of data: linked data model and supporting services*. Washington: Library of Congress, 2012, <<http://www.loc.gov/bibframe/pdf/marcl-d-report-11-21-2012.pdf>>; Karen Coyle, *Library data in a modern context*, «Library technology reports», 46 (2010), n. 1, p. 5-13.

40 Sulla questione, v. Erik Mitchell, *Leveraging metadata to create better web services*, «Journal of Web librarianship», 6 (2012), n. 2, p. 136-139.

41 In tal senso le conclusioni della rassegna di Nadine P. Ellero, *Integration or disintegration: where is discovery headed?* cit., p. 321-322.

una meta a volte prossima, a volte più difficile da raggiungere. Se i linked data che riguardano le tante differenti risorse disponibili e utili potranno popolare realmente il Web e altrettanto gli indici dei database bibliografici, allora il Web potrebbe rappresentare lo spazio in cui i motori di ricerca sono liberi di effettuare la loro indagine 'sconfinata', e gli spazi delle biblioteche rappresenterebbero quelli in cui i discovery tool possono sviluppare la ricerca entro confini disegnati nel disegnare una data biblioteca e la sua specifica *mission*, per quanto non facili da definire o teorizzare, per quanto ineffabili nell'essenza logica e tecnica. I sistemi di ricerca non generalisti, dunque, 'lavorerebbero' in un universo di risorse più controllato e affidabile, ma dovrebbero comunque avere la stessa 'capacità' e versatilità di un grande motore di ricerca di rete, insieme agli strumenti specifici per la ricerca avanzata. Tali caratteristiche consentirebbero una ricerca e una scoperta ampie ma 'sicure', oltre che strutturate. Allargare i confini di questo universo, poi, sarebbe semplice, e nulla di nuovo: i discovery tool dovrebbero applicarsi ai database di biblioteche 'alleanze', e non solo tra loro, ma anche con altre istituzioni culturali, riguardo a un dato ambito disciplinare o appartenenti a una tipologia funzionale, a livello nazionale o internazionale.

Senza costruire un universo di dati e risorse parallelo a quello del Web, le biblioteche possono offrire a tutte le persone – studiosi, professionisti, cittadini in genere – la possibilità di scelta di uno specifico 'punto di vista', di una linea da cui partire per addentarsi nell'informazione e nella conoscenza. Il punto di vista offerto dalle biblioteche, inoltre, non deve essere obbligatoriamente quello della ricerca più 'compiuta', più sicura dei propri passi o più assistita, ma possono anch'esse aprire un percorso più o meno avventuroso, o 'serendipitoso', per la scoperta delle risorse, che si svolge tra intuizioni e imprevedibilità, pieno di creatività, conducendo a risultati tanto inaspettati quanto preziosi⁴².

Gli ambiti delle biblioteche, definiti anche dal proprio punto di vista, possono realmente essere, dunque, una partizione più 'sicura' per non perdere il senso del contesto e per orientarsi tra le risorse dell'informazione e della conoscenza in un universo di dati potenzialmente infinito. In questa prospettiva, l'ipotetico 'contrasto' tra information retrieval e information discovery sarebbe, in ogni caso, di poco tenore. Questa 'attenuazione', ovviamente, non nel senso che i *web-scale discovery service* applicati all'ambito bibliotecario possano sostituire definitivamente gli strumenti di ricerca avanzati, restando i discovery tool e simili – almeno come concepiti adesso – strumenti di 'primo ingresso' al patrimonio delle informazioni e delle risorse mediate dalle biblioteche, come dagli archivi, dai musei, dalle gallerie eccetera.

La soluzione della forse solo apparente opposizione tra information retrieval e information discovery è più ampia, probabilmente, e risiede nella rivoluzione tecnologico-culturale che si può aspettare da parte dei *web-scale discovery service*, i linked data, il Web semantico, e anche l'*open access*. Si tratta, poi, di bilanciare accuratamente, da parte dei 'professionisti dell'informazione', il tipo di strumenti di ricerca e di recupero dei dati e delle risorse – libere o non libere – da fornire o da consigliare alle diverse tipologie di persone: a quelle che hanno bisogno di una veloce ricerca di informazioni generali o a quelle cui serve una guida per avviare uno studio, a quel-

⁴² Tra gli altri, v. Tammera M. Race, *Resource discovery tools: supporting serendipity*. In: *Planning and implementing resource discovery tools in academic libraries cit.*, p. 139-152. Inoltre, il comportamento e il percorso di ricerca fortuiti, basati sulla *serendipity*, possono essere anch'essi modellizzati, come gli altri *user behaviour*: v. Stephann Makri; Ann Blandford, *Coming across information serendipitously*, «Journal of documentation», 68 (2012), n. 5, p. 684-724.

le che richiedono un'assistenza avanzata per la ricerca scientifica o a quelle che possiedono un elevato grado di abilità nel Web senza il necessario approccio critico alla valutazione delle risorse.

Una delle conquiste dello sviluppo delle risorse della conoscenza stesse potrà essere l' 'indipendenza' dell'informazione dallo strumento di ricerca, la possibilità di trovare e valutare informazioni di qualità con qualunque strumento. Se cataloghi, banche dati, *repository* istituzionali, archivi *open access*, portali, si 'esprimeranno' tramite i linked data, se 'esporranno' le risorse della conoscenza da loro possedute o mediate tramite il nuovo metodo del *semantic Web*, se tutte le risorse valide saranno raggiungibili con il nuovo sistema, gli strumenti di ricerca, scoperta, recupero e condivisione, potranno essere più o meno specialistici, più o meno professionali, più o meno attendibili, ma nessuno avrà l'esclusiva dell'informazione di qualità, e di conseguenza dell'accesso alle risorse di qualità. Le risorse, dal punto di vista della loro raggiungibilità e valutabilità, saranno tutte sullo stesso piano: un documento eccellente potrà essere raggiunto dall'OPAC o dal Web, l'eccellenza sarà garantita dall'OPAC ma non dal motore di ricerca, chi si rivolge all'OPAC potrà avere una guida affidabile e chi usa un motore di ricerca dovrà orientarsi da solo, chi non è interessato alla garanzia di qualità, o sa bene come inquadrare quello che sta cercando, potrà preferire il motore all'OPAC...

Se cataloghi, banche dati, *repository* eccetera, devono aprire i propri contenuti, o meglio, devono trattare contenuti aperti al Web e ai *search engine*, non significa che devono anche trasformarsi in sistemi di ricerca generici: si perderebbe il senso e il valore di una distinzione tra strumenti specializzati e strumenti più semplici, strumenti realizzati da un'istituzione definita per i propri utenti specifici e altri utili a tutti anche se meno affidabili. A parte tale distinzione di 'specializzazione', però, basata sullo spazio di applicazione selezionato e sugli strumenti che è possibile applicare, non ci sono motivi per le biblioteche di mantenere una sorta di 'gelosia' per la qualità di un proprio 'posseduto', che già non può più essere veramente tale. Il 'valore aggiunto' da parte delle biblioteche, come da parte dei produttori di database specialistici, non cambierà nella sostanza e nei principi, e sarà sempre la capacità di valutare, selezionare e mediare tra le risorse disponibili, sarà la competenza per conoscerle e descriverle in modo appropriato all'utenza di riferimento.

In proposito, molto si deve investire sulle attività e i principi dell'*information literacy*, che deve ancor più essere approfondita come metodo di educazione all'uso e alla valutazione di tutti gli strumenti di information retrieval e discovery, e soprattutto per l'educazione alla valutazione delle risorse da essi raggiunte⁴³. Se apprendere il facile uso dei nuovi strumenti è immediato, cioè 'disintermediabile', rimane più tempo a disposizione per educare al 'pensiero critico' riguardo ai risultati ottenuti spesso con troppa facilità. Da parte dei bibliotecari e degli altri professionisti dell'informazione è più che mai necessario un ulteriore aiuto alle persone, native digitali o meno, per apprendere come scegliere e valutare criticamente sia le informazioni e le risorse, sia gli strumenti di ricerca e scoperta, soprattutto quando informazioni, risorse e strumenti appartengono ad ambiti altri rispetto alle biblioteche e alle istituzioni culturali, soprattutto quando lo spazio di ricerca e i risultati ottenuti si ampliano potenzialmente all'infinito⁴⁴.

43 Su questo punto, v.: Jody Condit Fagan, *Discovery tools and information literacy* cit.; Nancy Fawley; Nikki Krysak, *Information literacy opportunities within the discovery tool environment* cit.

Come corollario a tutto ciò, in uno spazio dove i contenuti sono egualmente raggiungibili indipendentemente dal sistema di ricerca, non ci sarà più spazio per i produttori di sistemi 'chiusi' – come per i database realizzati in modo da essere efficienti solo con i discovery tool distribuiti dallo stesso produttore. La concorrenza sarà il 'democratico' Web, prodigo e temibile come una fiumana popolare, e per confrontarsi e mettere a punto sistemi di ricerca, e interfacce proprietarie, più elevati di quelli liberamente disponibili si dovrà investire realmente sulla qualità delle operazioni informative possibili, sviluppandole tecnologicamente, ma nella direzione della migliore ricerca realizzabile, a diversi livelli, di dati e contenuti non più monopolizzabili⁴⁵. Quello che i produttori faranno pagare, dunque, sarà solo il frutto di tale lavoro altamente specializzato di raccolta, selezione, descrizione e mediazione – ciò che le biblioteche fanno gratis, o meglio, pagate insieme con altri servizi essenziali dalla comunità dei cittadini, fin dove è possibile.

Conclusioni

Non sarà possibile avere *web-scale discovery service* veramente rivoluzionari fino a che non ci saranno profondi cambiamenti nel trattamento dei dati cui essi devono essere applicati. Fino a che i cataloghi, i database, i *repository* e gli altri strumenti di mediazione e ricerca avranno un proprio schema per la creazione di descrizioni e metadati, i discovery tool dovranno creare un indice, per quanto unico, nato a partire da elementi differenti. Il risultato sarà sempre la 'semplificazione' dell'indice di ricerca, omogeneizzato al livello più 'basso' degli elementi che è possibile trovare in comune, e questo, mischiato a lacune di vario genere e a un'inevitabile confusione dei dati estratti da schemi diversi – per non parlare della decontestualizzazione – renderà l'affidabilità e la scientificità degli strumenti di ricerca scarsa, così come quella di molti strumenti del Web. È essenziale, dunque, pur mantenendo il rigore dell'analisi e della descrizione delle risorse, e anzi introducendoli per quegli strumenti che non ne sono molto dotati, che alla fine del processo si possano esportare, ovvero pubblicare, ovvero comunicare, secondo uno schema semplice ma rigoroso, condiviso ma autorevole, i dati reali delle risorse trattate. Secondo il metodo proposto dai linked data, nel Web semantico si ritroveranno tutti i dati delle risorse dell'informazione e della conoscenza, saranno collegati in modo da identificare le risorse e i loro 'significati', e potranno essere raccolti, trattati e proposti da strumenti di ricerca e scoperta di ogni genere: a partire da generici *search engine* fino a raffinate interfacce OPAC.

L'eccessiva disponibilità di informazioni e risorse, oltre a confondere le persone, fa già che queste si rivolgano all'insieme infinito del Web considerandolo come una fonte unitaria e complessiva, senza porsi il problema di individuare i confini tra diversi sistemi di ricerca, e nemmeno quelli delle risorse che nel Web non sono ricercabili: nell'abbondanza dell'offerta si trova spesso qualcosa, anche 'di qualità', che sod-

⁴⁴ Anche se orientati all'*information literacy* nell'ambito universitario, è utile vedere: Lisa M. Rose-Wiles; Melissa A. Hofmann, *Still desperately seeking citations: undergraduate research in the age of web-scale discovery*, «Journal of library administration», 53 (2013), n. 2/3, p. 147-166; Nina Exner, *Research information literacy: addressing original researchers' needs*, «The journal of academic librarianship», 40 (2014), p. 460-466.

⁴⁵ Riguardo le intenzioni di alcuni produttori commerciali, v. Axel Kaschte, *Linked open data on its way into next generation library management and discovery solutions*, «JLIS.it», 4 (2013), n. 1, p. 313-323. Inoltre, v. John J. Regazzi, *The battle for mindshare: a battle beyond access and retrieval*, «Information services & use», 24 (2004), n. 2, p. 83-92.

disfa le aspettative. I motori di ricerca, dal canto loro, tendono a confermare questa situazione, conferendo apparente omogeneità a risultati provenienti da fonti assolutamente eterogenee, non solo strutturalmente, ma anche per origini e fini. L'universo sconfinato, dunque, è reso univoco in modo 'forzato', anche se apparentemente semplice, mentre l'universo realmente univoco progettato dai linked data – realmente 'universo' – può essere meglio diviso in partizioni confinate, ma contestualizzate dato che restano in solidi rapporti con il 'resto', nelle quali i *discovery service* delle biblioteche e delle altre organizzazioni culturali possono condurre le proprie ricerche. Entro i propri confini – ineffabili, ma presenti, mutevoli, dinamici e vivi come la *mission* della biblioteca, ma sicuri – lo spazio di ricerca e di scoperta delle biblioteche può dimostrarsi sicuro anche delle proprie 'eterogeneità', governabili da strumenti di 'mappatura' che consentono un unico sguardo senza nascondere i contesti di provenienza delle diverse risorse.

Per quanto ne possano avere la tendenza, il Web semantico, i dati e gli strumenti di ricerca non si 'fanno' da soli. Necessario per le persone, tutto l'insieme deve essere creato e messo a punto dalle persone: nello specifico delle biblioteche, dai variamente identificabili professionisti dell'informazione. Senza l'aggiornamento dei modelli teorici e delle metodologie pratiche, senza lo sviluppo delle professioni e la volontà dei professionisti di 'svilupparsi', è inutile anche il semplice ipotizzare adeguamenti alla realtà, rinnovamenti, miglioramenti, crescite, rivoluzioni. I nuovi bibliotecari saranno quelli che, traendo sempre insegnamento dalla propria storia, puntellata di crisi e rivoluzioni, sapranno come inserirsi in un sistema centrato su nuovi paradigmi sviluppati a partire dai precedenti, in un nuovo universo bibliografico non più 'librocentrico' già da tanto. Saranno professionisti in grado di trattare e organizzare l'informazione e le risorse dovunque siano e qualunque aspetto abbiano, facendo la propria, sempre essenziale, parte tra i dati universalmente aperti e collegati⁴⁶.

Le conclusioni non possono essere che provvisorie e problematiche. Non si sta andando, comunque, verso la creazione di un sistema parallelo dove le macchine organizzeranno il mondo e le cose per noi. L'idea di «una rete di cose del mondo, descritta dai dati nel Web»⁴⁷, potrebbe essere giusto l'ennesima sovrastruttura culturale che farà vedere il mondo secondo una delle prospettive tramite le quali possiamo immaginarlo. L'importante è non perdere una visione ponderata e critica di tutto questo sistema, risalire sempre alla base culturale e umana di questa rappresentazione, e non farsi accecare, banalmente, dalla 'sottocultura' che spesso governa i luoghi comuni della rete e della visione acritica del suo potenziale e delle sue qualità. Lo spirito del progresso spinge da sempre a innovare tutto quanto circonda l'essere umano, a trovare la linea più adatta alla 'sopravvivenza' di tutto quanto è prodotto dalla sua azione, si deve semplicemente stare attenti a non perdere se stessi.

Articolo proposto il 5 dicembre 2014 e accettato il 27 gennaio 2015.

⁴⁶ Sul dibattito da sempre in corso sulle professioni, v. Mary L. Chute, *A core for flexibility*, «Information services & use», 32 (2012), n. 3/4, p. 143-147. Due altri spunti, tra le tante discussioni in proposito, sono rappresentati da: Maria Cassella [et al.], *Le professioni per le biblioteche accademiche di ricerca*, «AIB studi», 53 (2013), n. 1, p. 63-100; Antonio Sgobba, *I bibliotecari in crisi e la rivoluzione (attesa) degli open data*, «Corriere della sera», 6 giugno 2013, <<http://nuvola.corriere.it/2013/06/06/i-bibliotecari-in-crisi-e-la-rivoluzione-attesa-degli-open-data>>.

⁴⁷ Christian Bizer; Tom Heath; Tim Berners-Lee, *Linked data: the story so far* cit., p. 2.

ABSTRACT **AIB studi**, vol. 55 n. 2 (maggio/agosto 2015), p. 197-214. DOI 10.2426/aibstudi-11200

ROBERTO RAIELI, Università degli studi Roma Tre, Biblioteca di area delle arti, piazza della Repubblica 10, 00185 Roma, e-mail roberto.raieli@uniroma3.it.

Vecchi paradigmi e nuove interfacce: la ricerca di un equilibrato sviluppo degli strumenti di mediazione dell'informazione (SECONDA PARTE)

Scopo dell'articolo è proporre – con la forma della rassegna critica delle fonti, o *literary review* – una serie di temi ancora aperti, per indicare come sia effettivamente necessaria un'innovazione dei sistemi di gestione e ricerca dell'informazione, ma che sia profonda, equilibrata e scientificamente consapevole dei principi biblioteconomici. In particolare, è discusso il ruolo dei *web-scale discovery service* e dei discovery tool, inquadrando criticamente il loro accoglimento tra gli strumenti di mediazione della biblioteca.

Sono riassunte, anzitutto, varie considerazioni a favore dello sviluppo delle interfacce OPAC secondo alcuni principi riferibili al Web 2.0 e al *semantic Web*: estrema amichevolezza e usabilità, organizzazione interattiva e collaborativa, interoperabilità, granularità dei dati. È indagata, poi, la prospettiva dei discovery tool, proposti come nuovi strumenti rappresentanti il sistema 'unico' per la ricerca di tutte le risorse mediate da una biblioteca. Tale innovazione deve, però, puntare non solo alla trasformazione delle interfacce di interrogazione delle basi dati, ma anche a un nuovo modo di creare e organizzare i dati nelle basi stesse: secondo FRBR, RDF, linked data.

Sono, quindi, confrontate altre posizioni maggiormente orientate al rispetto dei principi biblioteconomici. È sottolineata l'importanza di mantenere la coerenza originaria delle risorse e l'autorevolezza della provenienza dei dati. Infine, sono proposte alcune riflessioni sui limiti dell'"information discovery", in quanto i nuovi sistemi non hanno, al momento, il rigore dell'information retrieval cui ci hanno abituato le banche dati, gli OPAC e altri strumenti di eguale affidabilità.

(Questo articolo è la seconda parte di un contributo più esteso. La prima parte è stata pubblicata sul numero 1/2015 della rivista)

Old paradigms and new interfaces: searching a balanced development of information mediation tools (PART TWO)

The paper presents a literary review of several pending issues that show the necessity of an innovation in the field of information management and search systems – still, this innovation must be deep, balanced and scientifically aware of LIS principles. In particular, it examines the role of web-scale discovery services and discovery tools, and their possible inclusion among the library mediation tools.

First, the article reviews a number of arguments supporting the development of OPAC interfaces according to the principles of Web 2.0 and semantic Web: extreme user-friendliness and usability, interactive and collaborative organization, interoperability, data granularity. Then, it investigates the prospect of discovery tools as the new 'single' systems to search for all the resources mediated by a library. However, such an innovation should aim not only to the transformation of the query interfaces of databases, but also to the development of a new way of creating and organizing data according to FRBR, RDF, linked data.

These arguments are then compared with other approaches more focused on LIS principles. In any case, it is essential to maintain the resources' original coherence, and the authority of the data origin. Finally, the paper remarks the limits of 'information discovery': at present, new systems do not have the same precision of databases, OPACs and other trustworthy tools.

(NB: this is the second part of a more extensive study. The first part has been published in *AIB studi* 1/2015)