

Dig *Italia*

Anno V, Numero 1 - **2010**

Rivista del digitale nei beni culturali

ICCU-ROMA

La bancarella planetaria e la biblioteca digitale: il punto di vista della ricerca e una possibile agenda per l'Italia

Alberto Petrucciani

Università degli studi di Roma "La Sapienza"

Il progetto Google Books (Google Book Search) è oggetto di un vivace dibattito, sotto vari aspetti e da vari punti di vista, in tutto il mondo. L'articolo si propone innanzitutto di esaminare concretamente, anche tramite esempi, quanto oggi il sistema offra dal punto di vista delle esigenze degli studiosi e dei ricercatori, ponendolo a confronto con altri servizi che mettono gratuitamente a disposizione risorse digitali (in primo luogo digitalizzazioni integrali di libri e periodici, liberi da diritti d'autore). Nella valutazione del progetto vanno distinte tre componenti o funzioni: il motore di ricerca a testo completo, che costituisce ovviamente l'aspetto più innovativo; il "catalogo" (ossia i dati bibliografici e le funzionalità di ricerca su di essi) e la biblioteca digitale (ossia l'insieme dei documenti digitalizzati e visualizzabili). L'articolo esamina quindi le prospettive del futuro sistema di accesso a pagamento e i rischi che questo comporterà, soprattutto per le biblioteche. Si propongono, infine, alcune considerazioni sulle esigenze e priorità a cui ispirare una politica nazionale per potenziare, anche con risorse molto limitate, la fruibilità delle risorse digitali di cui già disponiamo in Italia.

Esigenze di ricerca e biblioteche digitali

Di Google Books e dintorni, così come di biblioteche digitali, si discute molto, nel mondo e in Italia, in questo periodo. In questo contributo mi propongo di esaminare la questione dal punto di vista di chi fa ricerca, usando continuamente e intensamente Google Books e altri strumenti di ricerca d'informazioni e di recupero di testi (per motivi di semplicità non considererò ricerche con finalità diverse), per ricavarne qualche elemento di valutazione concreto e realistico, che spero possa essere utile ad orizzontarsi in un quadro caratterizzato per lo più da informazioni molto sommarie o generiche, spesso con carattere direttamente o indirettamente promozionale¹. Cercherò quindi di ricavare, dagli elementi di valutazione

¹ Per un'ampia e dettagliata presentazione del progetto di Google e delle controversie in corso rimando all'articolo di Gino Roncaglia, *Google Book Search e le politiche di digitalizzazione libraria*, «DigItalia», IV (2009), n. 2, p. 17-35. Tra i molti contributi sul tema segnalato, anche per i riferimenti bibliografici che offrono, gli articoli di Anna Maria Tammaro, *Il caso di Google Book e il*

e di riflessione esposti, alcune indicazioni e proposte per una “agenda” possibile, per il nostro paese, in questo campo.

Ovviamente chi scrive non può pretendere di rappresentare se non le *proprie* esperienze di ricerca, mentre molto diversificati sono i settori in cui il mondo della ricerca si articola, le esigenze più sentite e le fonti d’informazioni più usate. D’altra parte, tenersi sul vago e sul generico è oggi l’ultima cosa da fare: se non si distingue, se non si entra nel concreto, ci si perde in una nebbia di luoghi comuni, di affermazioni generiche e indimostrate, di cifre di molto incerta attendibilità e di ancor meno chiaro significato.

Senza imbarcarci in minuziose e spesso poco utili rassegne sui comportamenti di chi studia e fa ricerca (s’intende, dal livello di uno studente che debba preparare una relazione o una tesi, basata su una certa quantità di ricerca personale, fino al docente o ricercatore di professione che pubblica lavori in un campo specialistico), possiamo dire che il tratto comune di quest’attività è l’aver bisogno di un *numero notevole di testi*, sia importanti e molto diffusi sia meno importanti, poco conosciuti, di difficile reperimento, talvolta per un esame approfondito ma spesso solo per un rapido controllo o “per sicurezza”, ossia per accertarsi, in negativo, che non sia sfuggito niente di utile.

Questa *pressione* su chi fa ricerca *tende a crescere*, per la maggiore pressione a pubblicare (secondo gli standard di valutazione correnti nelle università, ogni docente deve produrre almeno alcune pubblicazioni ogni anno), per il forte incremento della letteratura specializzata (monografica e periodica), per le maggiori esigenze di precisione oggi richieste nelle citazioni, e per altri fattori ancora. E questa pressione tende a cumularsi con una crescente pressione sul *tempo*: si ha sempre meno tempo per dedicarsi alla ricerca (e alla lettura), anche se non è qui il caso di esaminarne i motivi.

Di conseguenza, anche se si può avere la tentazione di ridurre questa pressione abbassando la qualità (facendo ricerche e letture più frettolose, superficiali, approssimative), la strategia fondamentale resta quella di sempre, in condizioni di maggiore pressione ma per fortuna con strumenti migliori: “scremare” progressivamente ciò di cui si ha bisogno. Ossia, riuscirci a procurare con il minimo sforzo e il minimo dispiego di tempo la maggior quota possibile di ciò che ci occorre, così da dedicare via via più tempo e più fatica a una minoranza sempre più ristretta del materiale, mantenendo l’impegno complessivo entro limiti praticabili.

futuro della biblioteca digitale, «Biblioteche oggi», vol. 27, n. 5, giugno 2009, p. 28-34, e di Antonella De Robbio, *2010 Odissea Google Books*, «Biblioteche oggi», vol. 28, n. 3, aprile 2010, p. 44-59, entrambi disponibili anche in rete. Data la complessità del progetto, il carattere molto intricato dell’accordo proposto da Google e della giurisprudenza americana coinvolta, e la indisponibilità di molte informazioni, qualsiasi affermazione sui contenuti del progetto GBS va comunque, allo stato attuale, riguardata con molta (moltissima) cautela, essendo facile imbattersi in informazioni in forte contrasto tra loro (p.es. nei contributi di carattere giuridico diffusi negli Stati Uniti, anche da parte di legali esperti di *copyright*) o semplicemente erronee.

È evidente, perciò, quale importanza rivestano oggi le biblioteche digitali, o più concretamente il materiale che si può avere a disposizione con un clic, in pochi istanti, sulla rete, dovunque ci troviamo (sempre meno in condizioni di recarci fisicamente in una biblioteca, o in luoghi fuori orari in cui non potremmo farlo).

Per completare un quadro realistico della situazione è bene ancora aggiungere che normalmente le ricerche non si fanno “da zero”, né tutte in una volta, ma vanno piuttosto concepite come un’attività che si innesta continuamente su conoscenze e informazioni di cui già si dispone, andando ovviamente a modificarle.

Nella ricerca possiamo distinguere due grandi ambiti: quello della *individuazione e selezione* del materiale e quello della sua *utilizzazione* (consultazione, lettura). Sappiamo bene che oggi la cosa che più chiaramente e unanimemente l’utente vuole è passare istantaneamente, con un solo clic, dalla prima alla seconda fase, ma questo non rende irrilevante la distinzione. Come non fa cadere l’importanza della distinzione il fatto che, per una parte delle esigenze, il primo passo sia sufficiente senza il secondo: per esempio se si vuole semplicemente censire la fortuna di un’opera, o se si vuole rimpolpare di riferimenti bibliografici un testo senza la fatica di leggerli (un comportamento non lodevole ma diffuso).

Non dimentichiamo una considerazione anch’essa banale, ma di fondamentale importanza: chi si occupa di un certo campo generalmente è già a conoscenza di molto di ciò che gli occorre (lo ha letto in passato, lo ha trovato segnalato in precedenti circostanze, ne ha avuto notizia da colleghi o amici, ecc.). Inoltre, come è noto, la letteratura scientifica (nel senso più ampio) è fittamente interconnessa, a ritroso, da una rete di citazioni, rete che ha un tasso elevatissimo di ridondanza (gli scritti su uno stesso argomento condividono in genere la grande maggioranza delle citazioni ai lavori precedenti più significativi), cosicché una grande quantità di materiale si individua velocemente e semplicemente nel corso del proprio lavoro. Per questa via può essere difficile individuare i contributi più oscuri, obsoleti, di scarsissimo rilievo, ma la massima parte delle ricerche non aspira alla completezza bibliografica.

Di conseguenza, gran parte delle esigenze di accesso di chi fa ricerca *non* consegue immediatamente da una ricerca di individuazione: il ricercatore vuole sì accedere con un clic a ciò di cui ha bisogno, ma in molti casi *sa già cos’è*, non l’ha trovato in quel momento con Google (o in altro modo).

Inoltre, per completare il quadro d’insieme, sarà bene ricordare anche che le ricerche di individuazione non sono tutte dello stesso genere, e che una ricerca a testo completo come quella offerta da Google Books (Google Book Search) può dare frutti preziosi, non sostituibili da altre modalità di ricerca, da aggiungere alle informazioni di cui già si dispone, ma è uno strumento inadeguato per altri scopi (per esempio per identificare i contributi di maggiore rilievo su un tema che disponga di una vasta letteratura). Grandi cataloghi, bibliografie e altre fonti (per esempio quelle di orientamento specialistico) mantengono tutta la loro utilità nell’econo-

mia complessiva di una ricerca di qualche consistenza scientifica.

Come ultima premessa, è bene avvertire che ci si riferirà a situazioni tipiche della ricerca condotta in ambito umanistico, in senso lato, quello in cui rientrano le mie esperienze. Tuttavia sono convinto che estendendo quest'analisi ad altri ambiti apparirebbe chiaro quanto sia superficiale l'abituale bipartizione tra area umanistica e area scientifica: per esempio anche nell'ambito delle scienze pure o sperimentali vi sono aree (nella matematica, nella chimica, nelle scienze della terra, ecc.) in cui è necessario l'accesso a testi di cinquanta o cento anni fa, spesso difficili da reperire. In altri ambiti, per esempio quello giuridico, quello statistico-economico o quello dell'architettura, vi sono pure forti esigenze di accesso a materiale non contemporaneo, oltre a un peso fondamentale della letteratura di carattere nazionale rispetto a quella internazionale (o che si pretende tale).

Si farà riferimento quindi, per lo più, ad esigenze di ricerca che non riguardano il materiale più recente ma anzi, soprattutto, quello meno recente, e in particolare quello che cade al di fuori dei limiti del *copyright* e del diritto d'autore (limiti che possono essere molto diversi, come si sa, in paesi differenti e tra opere monografiche e periodici o altre opere collettive). Ovviamente il peso relativo del materiale più recente e meno recente nella ricerca può variare secondo i settori e le esigenze. Tuttavia, il materiale più recente è in generale quello più facilmente reperibile, tramite vari canali (bibliotecari, commerciali e personali, tradizionali ed elettronici). In ogni caso, nell'obiettivo indicato all'inizio di ridurre tempo e fatica della ricerca procurandosi facilmente una larga quota del materiale che occorre, il materiale libero dal diritto d'autore è evidentemente quello che oggi è più semplice mettere a disposizione in forma digitale e per tutti.

Google, il motore di ricerca e la biblioteca digitale

Google è – mi si perdoni la banalità – un *motore di ricerca*, cioè un certo tipo di servizio, oggi importantissimo e usatissimo per accedere al materiale disponibile in Internet, basato su una tecnologia specifica e sostanzialmente differente dalla fornitura di contenuti. La società che l'ha sviluppato ha via via esteso la sua attività ad altri campi, più o meno contigui e collegati, e probabilmente accarezza progetti che la portino a imporsi sempre più lontano, anche in aree che hanno una connessione molto indiretta con quella da cui è partita, ma che possono esservi collegate, con un evidente vantaggio competitivo rispetto alle altre imprese, anche in partenza meglio collocate in un dato campo. Considerazione, questa, che dovrebbe essere da sola sufficiente ad armarsi di molta cautela, dato quello che ben sappiamo di concorrenza e monopoli, posizioni dominanti e relativi abusi. Comunque, anche se ci limitiamo a considerare soltanto il motore di ricerca, è evidente, non appena ci si ferma a rifletterci, che le sue utilizzazioni sono di genere molto vario e riguardano esigenze (e curiosità) anch'esse diversissime, che in gran parte non hanno alcuna sovrapposizione – o hanno relazioni del tutto marginali – con le esi-

genze della ricerca, come definite al principio, o, da un altro punto di vista, con le funzioni e le attività delle biblioteche².

Se torniamo al quadro fatto al principio, è utile ribadire che un *motore di ricerca* può fornirci molte informazioni utili e interessanti di cui non disponevamo prima (per esempio riguardo a notizie presenti soltanto in siti Web, o a citazioni testuali in un libro in cui non avremmo pensato di andare a guardare³), ma che un *grande catalogo* (o una grande bibliografia: la distinzione di principio ai nostri fini è irrilevante, oltre ad essere nella realtà di oggi bisognosa di revisione) ce ne fornisce altre, non meno preziose, che per la prima strada non troveremmo (perché non ci sono o perché non sono individuabili con un dispendio ragionevole di tempo e di abilità).

Una ricerca nell'Indice del Servizio bibliotecario nazionale (SBN) con il nome dell'autore «Praga, Emilio», per esempio, ci restituisce 125 notizie bibliografiche di edizioni di opere del protagonista della scapigliatura milanese. Eliminando due dozzine di duplicazioni certe o quasi certe arriviamo rapidamente a identificare un centinaio di edizioni diverse. Lanciando la ricerca in Google Books possiamo trovare, con parecchia fatica per l'ordine non utile dei risultati, solo una piccola minoranza di queste edizioni, con dati bibliografici molto sommari (ricavati evidentemente da qualche catalogo bibliotecario), quasi tutte senza alcuna visualizzazione né informazioni ulteriori. Una sola edizione risulta visualizzabile a testo completo (la prima di *Penombre*, 1864)⁴.

La ricerca si svolge un po' meglio se usiamo la maschera di Ricerca avanzata di Google Books che comprende finestrelle analoghe a quelle di un OPAC (Titolo, Autore, ecc.): possiamo quindi ottenere risultati non più nell'ordine delle centinaia o migliaia (in gran parte inesistenti o ir reperibili, come vedremo fra poco) ma limi-

² Cfr., per alcune puntualizzazioni di base, Riccardo Ridi, *Biblioteche vs Google? Una falsa contrapposizione*, «Biblioteche oggi», vol. 22, n. 6, luglio-agosto. 2004, p. 3-5.

³ Un esempio sicuramente notevole è quello presentato recentemente da Alberto Salarelli, *Ricerca in Google Books: un caso di studio sull'aceto balsamico tradizionale di Modena*, «Culture del testo e del documento», n. 30, settembre-dicembre 2009, p. 57-70. La ricerca a testo completo tramite Google Books ha permesso di reperire una notizia molto importante, per la sua precocità, "sepolta" in una rivista tecnica inglese dell'Ottocento: insomma un bellissimo ago nel pagliaio. Tuttavia aghi nei pagliai, come si sa, capita di trovarli anche in tanti altri modi (i più comuni sono probabilmente le bancarelle, le chiacchiere con colleghi di altri settori e il girellare tra gli scaffali di una grande biblioteca). Mi sembra però che sarebbe un grosso errore prospettico quello di considerare il fortunato ritrovamento di qualche ago di pagliaio, anche prezioso, come l'esigenza tipica e primaria della ricerca, o comunque come qualcosa di più di una fra le tante esigenze della ricerca.

⁴ Può essere utile avvertire che i risultati presentati all'utente a seguito di una ricerca cambiano in maniera sostanziale se la ricerca viene svolta a partire dall'*home page* generale di Google, selezionando l'opzione *Libri* (invece di *Web*, *Immagine*, ecc.), o se invece questa viene svolta direttamente in Google Books (<http://books.google.it>). Il 2 maggio 2010, alla stessa ora, la ricerca «Praga Emilio» svolta nella prima modalità segnalava 27.100 risultati, mentre nella seconda modalità segnalava 1.190 risultati (cifre che peraltro sono entrambe, come vedremo più avanti, del tutto inattendibili). Anche il contenuto e l'ordine dei primi risultati sono diversi. Inoltre, come è noto, i risultati presentano variazioni più o meno ampie, a distanza di tempo, anche con una stessa modalità di ricerca.

tati a una quarantina di voci, che si riducono a circa due dozzine scartando gli abbondanti duplicati (alcuni facilmente identificabili, altri occultati da date errate). In Google Books troviamo quindi *meno di un terzo* delle edizioni reperite in SBN, e le troviamo presentate con dati molto lacunosi e spesso errati; una sola edizione, come si è detto, è presente col testo completo.

Se, alla ricerca di altre copie digitali, ci rivolgiamo ad Internet Archive (<http://www.archive.org/>), immettendo la stessa richiesta nella finestrella che compare subito in alto a sinistra, le sorprese non sono poche. Vi troviamo infatti 10 “libri” riprodotti integralmente (oltre al testo digitale di un’opera, proveniente dal Project Gutenberg), tra i quali, oltre alle *Penombre* digitalizzate da Google, tre esemplari di *Fiabe e leggende*, pure digitalizzati da Google (University of Michigan, University of California e New York Public Library), che non erano mai comparsi nelle precedenti ricerche⁵. Tornando a Google Books e smanettando un po’ con la Ricerca avanzata, dato che queste tre digitalizzazioni “ci devono essere”, finalmente si riesce a farle comparire. Non si va però oltre la Visualizzazione frammento. Come mai? I misteri di Google Books sono sicuramente tanti, forse insondabili, e comunque a me ignoti. Quel che è certo è che il *copyright* non c’entra: il povero Praga morì, a quanto si dice di stravizi, il 26 dicembre 1875, quindi dalla sua morte sono trascorsi quasi 135 anni. Altra cosa certa è che non si tratta di un caso isolato: ho esaminato decine e decine di digitalizzazioni realizzate da Google e accessibili in Internet Archive ma *non* in Google Books.

Al termine di questo *tour de force* abbiamo raggranellato l’edizione originale di due dei quattro volumi di poesie del Praga (*Penombre* in Google Books e lo stesso, più *Trasparenze*, in Internet Archive, oltre a un’edizione successiva di un’altra opera). Per inciso, la prima edizione di *Trasparenze* (1878) e la seconda di *Fiabe e leggende* (1884) si trovano anche, in una buona digitalizzazione completa, su un sito italiano (Università di Torino), ma non è tanto facile arrivarci, con un motore di ricerca, anche sapendo della sua esistenza.

Ma se avessimo bisogno anche, o soltanto, di un’edizione moderna di riferimento, quella in cui, in un lavoro serio, dovremmo controllare una citazione? In un’infinità di fonti (dalle comuni storie letterarie al catalogo SBN, e armandoci di pazienza anche tramite Google) possiamo trovare l’indicazione che esiste un’edizione critica, uscita nel 1969 negli «Scrittori d’Italia» di Laterza. E quindi, come molti (ma certo non tutti) sanno, disponibile integralmente e gratuitamente, in forma digitale, non tramite Google Books, ma su un sito italiano (<http://www.bibliotecaitaliana.it/exist/Scrittoritalia/>), insieme a tutto il resto della collana, per generosità dell’editore. Per chi non lo sa già, è difficile arrivarci tramite una ricerca con

⁵ Avverto che le digitalizzazioni accessibili nella sezione *Texts* di Internet Archive sono solo in parte reperibili anche nella Open Library (Biblioteca Aperta, <http://openlibrary.org/>) basata appunto su Internet Archive. Non mi è mai capitato invece di notare il contrario, anche se non sono in grado di escluderlo.

Google (Web), non ci si arriva dalla scheda di SBN, mentre ci arriverebbe facilmente chi fosse andato a guardare la pagina dedicata al poeta da Wikipedia (nonostante, come si sa, sia una sede molto avara di collegamenti esterni).

Torniamo ai grandi siti americani e prendiamo un esempio d'altro genere. Se avessimo bisogno di qualche vecchio articolo della «Nuova rivista storica», in Internet Archive troveremmo 12 annate complete, appartenenti al primo periodo della rivista (ormai fuori *copyright*) e digitalizzate da Google. In Google Books però non ne troviamo nemmeno una, a testo completo: vi compaiono invece, immagino per errore, le digitalizzazioni integrali di tre fascicoli relativamente recenti (i n. 1 e 2 del 1995 e il n. 3 del 1999). Risultati analoghi, e a volte anche molto migliori per il ricercatore, si ottengono in Internet Archive – e *non* in Google Books – per numerosi altri periodici di cultura italiani tra Otto e Novecento.

Proviamo ancora a cercare, come campione, i quattro dizionari biografici dei contemporanei (con vari titoli e pubblicati due in italiano e due in francese) di Angelo De Gubernatis, una fonte tuttora molto utile per i personaggi minori e perché attingono spesso a notizie di prima mano non attestate altrove. In Google Books – perdendo al solito un po' di tempo a fare e rifare la ricerca in vario modo – non troviamo, alla fine, nessuna digitalizzazione disponibile. È comunque bene ripetere una ricerca in vari modi perché, per esempio, libri segnalati con Visualizzazione completa nella selezione «Anteprima limitata e visualizzazione completa» scompaiono se scegliamo «Solo visualizzazione completa» (provare per credere).

La cosa è di nuovo alquanto misteriosa perché su Internet Archive troviamo alcune digitalizzazioni integrali realizzate da Google (Michigan e Stanford), per i repertori del 1879 e 1891, con l'aggiunta del supplemento del 1906 al repertorio del 1905. Anche in questo caso, a scanso di dubbi, va ricordato che il *copyright* non c'entra: De Gubernatis è morto nel 1913 (e la cosa risulta, oltre che da cento altre fonti, anche dall'intestazione della Library of Congress). In Internet Archive troviamo anche altri esemplari, di migliore qualità, dei primi due repertori (per inciso, quello del 1891 si trova anche su Gallica). Inoltre, da poche settimane, in Internet Archive è disponibile anche il *Piccolo dizionario* del 1895, scansionato all'Università di Toronto (al di fuori del progetto Google). Speriamo che Toronto ci regali presto anche la scansione del repertorio del 1905, completando la "collezione" resa disponibile in Internet Archive (e da nessun'altra parte, Google Books compreso).

Andrà meglio con opere più classiche, e più lontane dai limiti del diritto d'autore? Proviamo per esempio a cercare gli *Scrittori d'Italia* del Mazzuchelli, un'opera che non manca in tutte le grandi biblioteche. In Google Books troviamo qualcosa, ma si tratta di due soli volumi, scompagnati, su sette. Purtroppo al momento nessun volume è disponibile in Internet Archive (che contiene però altri suoi scritti) né su Gallica.

Proviamo ancora a confrontare i risultati della ricerca di qualche autore italiano in Google Books e in Internet Archive, considerando solo i libri visualizzabili o scaricabili per intero (ed escludendo raccolte di più autori e traduzioni): per il

Poliziano, 10 volumi disponibili in Google Books e 14 in Internet Archive, per Pietro Bembo 16 volumi disponibili in Google Books e 23 in Internet Archive, per Verga 27 volumi disponibili in Internet Archive e nessuno in Google Books⁶.

Come ultimo esempio proviamo a cercare un capolavoro della letteratura inglese, la *Ballata del vecchio marinaio*. Digitando «Coleridge ancient mariner» in Google Books (il 24 aprile) vengono annunciati 7.118 risultati, dei quali 5.947 con Anteprima e 1.264 con Visualizzazione completa (si ottiene lo stesso numero, 1.264 risultati, anche selezionando «Solo di dominio pubblico» invece di «Solo visualizzazione completa», ma l'ordine, e forse anche il contenuto, è diverso). Se però andiamo a verificare questi numeri, spostandoci con qualche clic fino alla fine della lista, scopriamo che i 1.264 risultati promessi finiscono con «166 di 275» (*sic*), cioè sono soltanto 166 (né 1.264 né 275)⁷.

La stessa ricerca, partendo dalla pagina generale di Google e scegliendo l'opzione *Libri*, annuncia 449.000 risultati. Se si sceglie l'ordinamento per data, però, i risultati scendono a 119.000 (cioè solo il 26% di quanto dichiarato inizialmente). Se andiamo alla fine della lista, con 5 clic, i risultati diventano 453: circa l'1 per mille rispetto al dato iniziale, senza ordinamento per data, e lo 0,4% (scarso) rispetto a quanto dichiarato al principio della lista (119.000)⁸. In altre parole, il 99,6% dei risultati annunciati (promessi) o non esiste, o non viene mostrato. Non si sa quale delle due ipotesi sia da considerare la peggiore. Rispetto ai risultati della ricerca dichiarati nella prima schermata, lo 0,4% è costituito da dati esistenti e accessibili, il 99,6% è *bluff*, o dati inaccessibili. Provando altre ricerche i risultati possono variare un po', in meglio o anche in peggio, ma la sostanza rimane questa⁹.

⁶ Questi dati si riferiscono a ricerche compiute il 4 maggio e *non* depurate da eventuali duplicati (o da più esemplari della stessa edizione). Nell'ultimo caso, pur essendo Verga largamente fuori dai limiti del diritto d'autore sia in Italia (essendo morto nel 1922) che negli Stati Uniti, si può ipotizzare che Google, per non dedicare tempo ad accertare la condizione giuridica di opere di autori non americani vissuti tra Otto e Novecento, abbia escluso in blocco dalla visualizzazione larghe quantità di materiale non americano della seconda metà dell'Ottocento e del primo Novecento. Vengono visualizzate integralmente, per esempio, parecchie edizioni manzoniane della prima metà dell'Ottocento, mentre quelle un po' più recenti si trovano ampiamente rappresentate in Internet Archive ma non in Google Books.

⁷ Identica ricerca, rifatta il 2 maggio, dichiara 7.264 risultati, dei quali 6.515 con Anteprima e 1.266 con Visualizzazione completa. Recandosi alla fine della lista però si rivelano «322 di 331», ossia soltanto 322, con un miglioramento numerico del dato finale ma senza cambiamenti sostanziali per quanto riguarda l'inattendibilità delle informazioni fornite.

⁸ Identica ricerca, rifatta il 2 maggio, dichiara 460.000 risultati, che ordinati per data si riducono a 122.000, e recandosi alla fine della lista si rivelano 471. I dati sono quindi leggermente aumentati ma le percentuali sono rimaste all'incirca le stesse.

⁹ Nel caso della precedente ricerca con «Praga Emilio» (2 maggio) i 27.100 risultati di Google con opzione *Libri*, andando con alcuni clic all'ultima pagina, si rivelano essere invece 385, mentre i 1.190 risultati di Google Books si rivelano 387 (una cifra molto vicina all'altra, come si può notare). I dati effettivamente accessibili sono quindi l'1,4% di quelli dichiarati con la prima ricerca e il 32% di quelli dichiarati con la seconda. Una ricerca svolta il 4 maggio per «Benedetto Croce»

L'inattendibilità non riguarda solo le informazioni sui risultati delle ricerche, ma anche quelle sui singoli libri. Cercando, per esempio, edizioni relativamente recenti della ballata di Coleridge (SBN include una sessantina di edizioni del XX secolo, comprese quelle italiane), quella che ci viene segnalata per prima da Google Books, con la data del 1947, è invece un'edizione elettronica (non autorevole) del 2008; quella che compare subito dopo con la data del 1900 è una vera digitalizzazione di un'edizione a stampa, ma di quella illustrata newyorkese del 1866 (*sic*).

La grande biblioteca digitale, insomma, sembra piuttosto una gigantesca bancarella, di dimensioni planetarie, dove girellando qua e là si trova sempre qualcosa – come al mercato delle pulci – ma è molto difficile trovare qualcosa di completo e di organico, e quel che c'è trovarlo in ordine. Per di più, parrebbe che, come sulle bancarelle, da una volta all'altra qualcosa compaia e qualcosa scompaia (ma più misteriosamente che nelle bancarelle reali, dove possiamo supporre che quel che non c'è più sia stato comprato da qualcuno). Qui, diversamente dalle bancarelle, gli oggetti hanno delle etichette, ma purtroppo sono molto spesso sbagliate. Infine, molto più che al mercato delle pulci, l'utilizzatore è assordato da sparate autopromozionali che non reggono alla verifica di quattro o cinque clic.

Una notevole quantità di problemi e di inconvenienti sorge anche quando finalmente, dopo un po' di tempo perso e di materiale da scartare perché non è quello che viene dichiarato, arriviamo alla digitalizzazione del libro che ci interessa. La qualità, come è noto, è spesso bassa, non tanto dal punto di vista tecnico dell'immagine (immagini di bassa qualità sono di solito leggibili e sufficienti per la grande maggioranza delle esigenze), quanto dal punto di vista procedurale: pagine saltate, pagine tagliate o mosse e difetti analoghi.

Ultimo aspetto che è il caso di segnalare è l'attività di conversione delle immagini in testi, su cui lavora il motore di ricerca. Sappiamo che i programmi di OCR (Optical Character Recognition) lasciano ancora molto a desiderare e che la verifica dei risultati, se fatta da un operatore umano, è molto lunga e costosa. Tuttavia è bene essere consapevoli che i risultati su cui Google lavora sono spesso di qualità piuttosto bassa, anche se è raro che l'utilizzatore se ne accorga. Se ne accorge, ovviamente, quando gli errori compaiono nelle poche parole che vengono visualizzate nei risultati della ricerca, ma i file di testo utilizzati da Google non sono per lo più visibili all'utente. Può essere istruttivo guardare i risultati che si ottengono digitando parole inesistenti, per esempio «Mcrope» (errore frequente per «Merope»)

produce in Google Web l'annuncio di 648.000 risultati (ma la lista termina invece con il risultato 768), se si seleziona l'opzione *Libri* si dichiarano 224.000 risultati (ma la lista termina al numero 177, lo 0,08% di quanto "promesso"), mentre se la ricerca viene svolta direttamente in Google Books i risultati "promessi" diventano 4.910 e sono poi di fatto, spostandosi alla fine della lista, solo 185 (3,7% di quanto "promesso"). Quando questo articolo era già completato Google ha modificato le maschere di interrogazione e la visualizzazione dei risultati delle ricerche, che si presenta ora in maniera leggermente diversa dal punto di vista formale: non si notano però differenze di rilievo rispetto a quanto osservato nel testo.

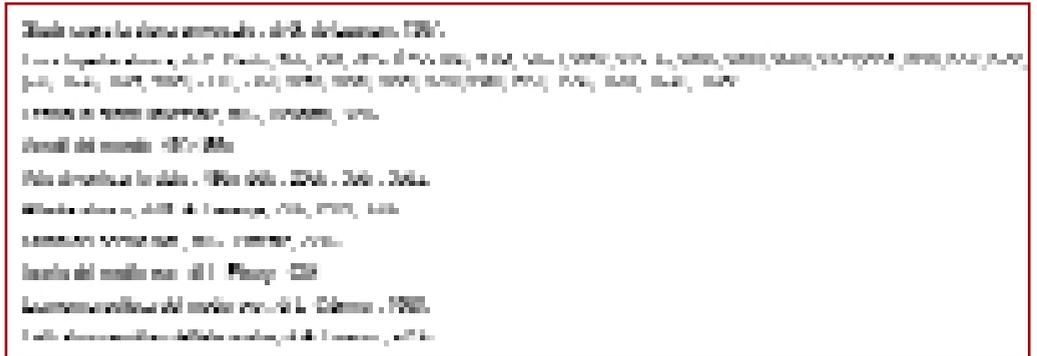


Figura 1. Estratto dalla *Bibliografia italiana* (Google Books)

o «morre» (per «morte» o «morire»). Dell'«Annuario per Panno scolastico» (*sic*) dell'Università di Roma risultano presenti ben dieci «Ulteriori edizioni».

È chiaro che questo problema può risultare spesso trascurabile quando si digitaliz-

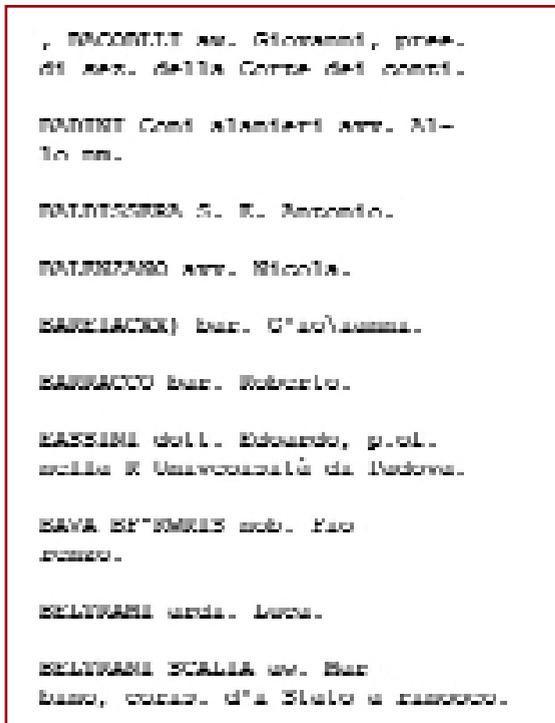


Figura 2. Estratto da *Chi è* (Internet Archive)

za un intero libro, in cui di solito le parole significative dal punto di vista del contenuto ricorrono parecchie volte, ma diventa molto rilevante per esempio nel caso delle opere di consultazione, in cui la notizia può andare completamente perduta. Come si vede dalle due figure (di qualità media, relative a parti di una pagina della *Bibliografia italiana* dell'editore Stella, da Google Books, e all'elenco dei senatori del Regno nel *Chi è* del 1908, da Internet Archive), una persona che cercasse il senatore Cibrario o il generale Bava Beccaris non li troverebbe mai¹⁰.

¹⁰ È un po' fastidioso anche che l'inizio dei *Sepolcri* diventi (nella prima edizione elettronica che Google Books propone nel formato epub): «11' ombra de' cipressi e dentro l'urne Confortate di pianto è forse il sonno Della morte nien duro?». Molto meglio leggerlo in formato immagine.

Dalla bancarella planetaria di oggi al mercato mondiale di domani

Dalle osservazioni precedenti non si conclude, naturalmente, che su Google Books non ci sia nulla, che non si trovi nulla, o che quel che si trova non serva a nulla. Ma si può concludere che si tratta di un servizio lontanissimo dal soddisfare tutte le esigenze di chi fa ricerca, e che molte di queste – come si è visto – sono soddisfatte non peggio, anzi meglio, da altri servizi o da strumenti diversi.

Si può prevedere che, col tempo, Google migliorerà i suoi servizi, accorgendosi di un po' di limiti e inconvenienti e cercando di ridurli, così che il suo uso risulti più fruttuoso e meno frustrante per chi fa ricerca. Ma difficilmente queste esigenze saranno prioritarie, dal punto di vista dell'azienda, perché riguardano un segmento di pubblico (cioè di potenziale mercato) quantitativamente ristretto, probabilmente di scarsa attrattiva dal punto di vista pubblicitario, e quindi molto meno rilevante dei milioni di utilizzatori non specialisti, che cercano un volo *low-cost*, un vecchio compagno di scuola, l'ultima intervista a Totti, una foto del cantante preferito, qualcosa da incollare in una ricerca scolastica, e così via. Cose degnissime, queste e tante altre che si potrebbero citare, sicuramente più importanti nel *core business* dell'azienda rispetto alle esigenze dei ricercatori. Queste degnissime cose, inversamente, sono sempre state, e sempre più saranno in futuro, del tutto marginali, o estranee, dal punto di vista della ricerca o delle biblioteche.

Allo stato attuale delle cose, comunque, per l'ambito che stiamo considerando, Google Books offre un interessante (e sicuramente unico) servizio come *motore di ricerca*, a testo completo, anche su una discreta quantità di materiale a stampa. Le sue funzionalità e i suoi risultati come *catalogo* (o *bibliografia*) – un secondo e diverso punto di vista – sono invece, come si è visto, molto limitate e carenti, sicuramente inferiori ai cataloghi delle grandi reti bibliotecarie.

E come *biblioteca digitale*? Da questo terzo punto di vista, o per questa terza funzione, bisogna pure concludere che allo stato attuale il servizio è molto limitato, salvo in alcuni ambiti particolari (per esempio, ovviamente, per la storia americana dell'Ottocento), oltre ad essere macchinoso e poco efficiente nelle funzionalità.

Ma da dove viene allora il mito, da cui siamo martellati, di una biblioteca digitale mastodontica, quasi universale, vicina a contenere tutto quel che si può desiderare? A parte la grancassa promozionale, che ovviamente Google non ha difficoltà a procurarsi dai mezzi d'informazione di massa (e non solo da quelli), il fatto che sia ben poco quanto è oggi visibile integralmente *non* significa che sia poco quanto Google ha già realizzato (anche se non lo rende accessibile, per ragioni legali dirette o indirette e probabilmente anche per motivi d'altro tipo, tecnici o commerciali). Le cifre assolute dichiarate (in milioni di volumi digitalizzati) hanno, peraltro, un'attendibilità molto limitata – anche se probabilmente sono un po' meno fantasiose del numero di risultati annunciati nelle ricerche – per vari motivi, che vanno dalle maniere anche molto differenti di conteggiare (come sappiamo da tanto tempo per le statistiche bibliotecarie) alla quantità di

duplicazioni (veri e propri file duplicati presentati con dati varianti, o riproduzioni dello stesso oggetto presentate o commercializzate come edizioni differenti, oltre alla notevole quantità di esemplari diversi di una stessa edizione). La digitalizzazione di più esemplari di una stessa edizione, naturalmente, può essere graditissima agli studiosi di bibliografia testuale, oltre a costituire una risorsa per ovviare alla frequenza di difetti delle digitalizzazioni stesse, ma l'ampiezza del fenomeno è probabilmente molto superiore al tasso di duplicazioni medio tipico di singole anche grandi biblioteche (per esempio, Google ha digitalizzato almeno 4 copie dell'ultima edizione delle Regole di catalogazione di Cutter, oltre a varie ristampe anastatiche della stessa edizione¹¹). Questo fenomeno, insieme agli altri fattori citati, aiuta a comprendere come mai la gigantesca collezione di Google Books si riveli in concreto, dal punto di vista dell'uso, dare un tasso di risultati positivi analogo a quelli di una biblioteca fisica di dimensioni molto più limitate (forse un quarto, o anche meno, della biblioteca digitale dichiarata da Google).

Pur con queste precisazioni, e con gli ampi margini d'incertezza che conseguono dai risultati molto aleatori dei sondaggi che un singolo utente può fare sul sistema, credo che si possa comunque dare per assodato che Google disponga oggi, dal materiale più antico sicuramente di dominio pubblico fino a quello recentissimo sicuramente soggetto al *copyright*, di una collezione di digitalizzazioni molto ricca, probabilmente nell'ordine di alcuni milioni di unità, fortemente concentrata sulla produzione americana ma con presenze quantitativamente rilevanti di materiale anche degli altri paesi. Della produzione europea probabilmente è presente soprattutto il materiale contemporaneo d'interesse accademico che viene acquisito dalle grandi biblioteche di ricerca statunitensi. Basta dire, per esempio, che tra i libri che Google ha digitalizzato è compresa gran parte di quelli editi negli ultimi anni dall'Associazione italiana biblioteche, che non è certo un grosso editore né particolarmente in evidenza per distribuzione o promozione.

¹¹ Un po' curiosamente, il 4 maggio nessuna di tutte queste digitalizzazioni risulta visibile a testo integrale in Google Books, mentre i quattro esemplari digitalizzati da Google e un quinto digitalizzato a spese della Microsoft sono visualizzabili e scaricabili in Internet Archive. Si può aggiungere che, lo stesso giorno, i primi 5 risultati forniti da Google Books cercando «Cutter» e «Rules» sono 5 recenti digitalizzazioni, prive di indicazione dell'edizione originale, prodotte da due imprese (BiblioBazaar e General Books) che "riciclano" in maniera molto discutibile e discussa, in formato immagine e/o in formato testo generato da OCR e non corretto, le solite edizioni originali, sulla base di scansioni la cui provenienza è pure dubbia e discussa. Al di là della poca chiarezza sull'attività di queste imprese (la prima è risultata la casa editrice americana che ha dichiarato nel 2009 il maggior numero di titoli prodotti, oltre 270.000, ed entrambe affollano i risultati di Google Books), il fenomeno aiuta a spiegare come una sola edizione possa moltiplicarsi, in Google Books, in una mezza dozzina (o anche più) di risultati distinti, con le conseguenze che questo comporta per il rapporto tra cifre totali dichiarate e contenuto effettivo, valutato in termini di edizioni diverse.

Se cerchiamo, in una situazione ancora fluida e difficile da prevedere, di identificare gli elementi più certi, quelli altamente probabili e quelli molto aleatori, possiamo iniziare da quattro punti fermi:

- Google ha messo insieme (non importa quanto lecitamente, questione sempre in discussione e su cui esistono vari procedimenti giudiziari aperti) una collezione di digitalizzazioni di materiale a stampa di dimensioni molto grandi, a scala nettamente superiore rispetto a qualsiasi altra iniziativa del genere;
- questa collezione, utilizzata dal motore di ricerca, è invece accessibile per l'utente, allo stato attuale, in una misura limitatissima, inferiore a quanto offrono altre biblioteche digitali di grandi dimensioni (soprattutto Internet Archive);
- Google sta mettendo a punto l'organizzazione e l'offerta di un servizio di accesso a pagamento, su larga scala, alla sua collezione digitale, come viene spiegato nel testo del *Settlement* e in altre fonti;
- per le funzionalità di ricerca (ossia per il motore di ricerca a testo completo), e anche per la massa di materiale che potrebbe venire offerto, questo prossimo servizio a pagamento si prospetta come sostanzialmente monopolistico.

Sul nuovo servizio a pagamento annunciato da Google non disponiamo, al momento, di informazioni affidabili, probabilmente anche perché non tutte le decisioni sono state già prese. Possiamo però ritenere altamente probabile che, come è avvenuto fin qui e come si promette anche per l'immediato futuro, il servizio di *ricerca* sui libri (cioè l'uso del motore di ricerca) rimanga un servizio *gratuito e liberamente accessibile*, con tutti i vantaggi che questo comporta per gli studiosi e il pubblico generale. Infatti è sulla libera accessibilità del servizio di base che Google fonda la sua crescita (e gran parte dei suoi ricavi attuali) e anche altre esperienze hanno mostrato chiaramente che la strategia commerciale vincente in rete (e non solo), per chi si rivolge al grande pubblico, è innestare fonti di ricavo (e servizi a pagamento) sopra un servizio gratuito di larghissimo uso.

Se è noto e scontato che saranno a pagamento i servizi di accesso al materiale digitalizzato che sia sotto *copyright*, ci sono ancora grossi margini di incertezza su altri elementi importanti per definirne il profilo:

- se sarà a pagamento anche l'accesso al materiale di cui non sia sicura la libera disponibilità e per il quale non siano identificati o identificabili i titolari di diritti (ma lo si può considerare praticamente certo, se questo materiale sarà reso accessibile e non escluso per motivi legali);
- se e in che misura sarà incluso nel servizio anche il materiale europeo (per la complessità dei problemi sia legali sia organizzativi), cosa che ovviamente ha

- molta rilevanza per chi fa ricerca in Europa o in ambiti in cui il materiale nazionale o europeo ha un'importanza fondamentale¹²;
- se (come è possibile ma tutt'altro che certo, sia nell'immediato che in prospettiva) continuerà a rimanere gratuita la visualizzazione integrale del materiale sicuramente fuori diritti;
 - quante e quali utilizzazioni libere o gratuite saranno possibili (per esempio anche riguardo alle anteprime), questione su cui circolano per ora semplici ipotesi, comunque molto restrittive e vivacemente discusse (per esempio che visualizzazioni integrali gratuite possano essere concesse da Google, con severe limitazioni, da postazioni interne ad alcune biblioteche americane);
 - *last not least*, le tariffe e le modalità contrattuali per l'accesso al nuovo servizio a pagamento.

Il quadro che si prospetta, quindi, è quello della nascita di un nuovo e sostanzioso mercato, inventato e gestito da Google (ovviamente con accordi con gli editori e le loro organizzazioni), che può avere qualche analogia – e forse in un futuro non immediato potrebbe avere anche una certa sovrapposizione o conflittualità – con quello, già sviluppato, dell'editoria elettronica di tipo scientifico o comunque specializzato (periodici elettronici e banche dati).

Il nuovo servizio ovviamente potrà rivolgersi all'utente finale, e soprattutto al pubblico generale, molto più di quanto non sia avvenuto per i periodici scientifici, ma è altrettanto ovvio che si rivolgerà in larga misura, in maniera finanziariamente determinante, allo stesso bacino, ovvero alle biblioteche e alle istituzioni di ricerca.

¹² L'*Amended Settlement Agreement*, com'è noto, esclude quasi completamente i libri pubblicati fuori dagli Stati Uniti (e da alcuni altri paesi di lingua inglese), per i quali dovranno essere fatti accordi particolari con gli aventi diritto. Ma questa limitazione, per quanto drastica, non risolve, come spesso si asserisce superficialmente ed erroneamente, i problemi legali riguardanti il materiale europeo. Infatti – com'è ovvio non appena ci si fermi a riflettere – tra i libri pubblicati negli Stati Uniti ve ne sono molti (per esempio, traduzioni) che coinvolgono diritti di autori europei, o di altri paesi. Un recentissimo *Memorandum* legale elaborato per la Open Book Alliance (<http://www.openbookalliance.org/wp-content/uploads/2010/05/Arato-Memo.pdf>) chiarisce per quali elementi l'accordo proposto da Google violi importanti clausole della Convenzione di Berna (a cui gli Stati Uniti hanno aderito nel 1988, oltre un secolo dopo la sua approvazione) e di altre convenzioni internazionali.

Può essere utile ricordare anche che, secondo le stime di Brian Lavoie e Lorcan Dempsey di Online Computer Library Center – OCLC (*Beyond 1923: characteristics of potentially in-copyright print books in library collections*, «D-Lib Magazine», vol. 15, n. 11/12, November-December 2009, <http://www.dlib.org/dlib/november09/lavoie/11lavoie.html>), i libri pubblicati negli Stati Uniti costituiscono solo il 18% dei libri a stampa attualmente presenti in WorldCat (15,5 milioni su 84,8 milioni). L'accordo proposto da Google quindi, nonostante i grossi rischi che comporta per la creazione di una posizione di monopolio, non è assolutamente in grado di portare al miraggio che viene sbandierato per convincere i renitenti che "il fine giustifica i mezzi", quello dell'accesso, sia pure a pagamento, a "tutti i libri". Costituirebbe invece un pesante precedente, un'ipoteca, sulle possibili strade per mettere a disposizione l'altro 82% dei libri a stampa, e i materiali diversi (a partire dai periodici) non compresi nell'accordo stesso.

Sembra molto improbabile che altri servizi connessi, come il *print-on-demand*, possano acquisire dimensioni consistenti in termini finanziari e si possono nutrire larghi e motivati dubbi, sulla base delle precedenti esperienze (appunto il *print-on-demand*, la vendita in rete di *ebooks*, la vendita di brani musicali), sulla consistenza del possibile mercato individuale e privato (e, correlativamente, del *pay-per-view*), rispetto a quello istituzionale (e quindi, correlativamente, della sottoscrizione in blocco o comunque tramite grossi contratti forfettari).

Una prospettiva da tenere in seria considerazione, quindi, è che si apra una vicenda complessa (e costosa) come quella che ha caratterizzato il decennio scorso riguardo ai periodici elettronici e alle relative licenze e contratti. Chi opera nel mondo delle biblioteche sa quante questioni problematiche e delicate e quali oneri per i bilanci sia stato necessario affrontare in questi anni, riguardo ai periodici elettronici, al di là dell'ovvio beneficio per l'utente di avere gli articoli che cerca a portata di un clic, sul suo computer.

Questa volta la vicenda potrebbe rivelarsi ancora più problematica, da un lato perché riguarda un unico fortissimo monopolio piuttosto che un certo numero di operatori, per quanto pochi e molto robusti (un oligopolio), dall'altro perché coinvolgerebbe un po' tutte le biblioteche, comprese quelle pubbliche, e non solo le università e le grandi istituzioni di ricerca, che nonostante le loro difficoltà e debolezze potevano contare su risorse finanziarie, tecnologiche e professionali più congrue per il compito. Il tutto in un contesto generale, economico e soprattutto del finanziamento pubblico alle biblioteche e alla ricerca, che è sicuramente molto peggiore di quanto non fosse dieci anni fa.

I rischi quindi sono notevoli¹³ e, come si sa, dai rischi si cerca di assicurarsi.

¹³ Nella vasta letteratura sul tema abbondano, come si poteva prevedere, i minimizzatori dei rischi che ci troviamo di fronte e i fautori della privatizzazione di qualsiasi cosa come unica strada praticabile e foriera di risultati positivi. I secondi potrebbero forse ricordarsi che l'innovazione più importante e utile degli ultimi vent'anni, Internet e il Web, si deve al mondo della ricerca e delle amministrazioni pubbliche (*non-profit*), mentre i colossi di allora, IBM e Microsoft, stavano perseguendo – per nostra fortuna senza successo – la strada alternativa di sistemi e protocolli proprietari e di reti commerciali chiuse e a pagamento. Anche per i primi non mancano recenti istruttive esperienze su cui riflettere, dalla TV commerciale e dal software di base per personal computer all'evoluzione dell'editoria tecnico-scientifica. In particolare, le argomentazioni anche di provenienza economica per sminuire i rischi a fronte dei vantaggi dell'operazione ignorano sistematicamente la distinzione tra il motore di ricerca offerto gratuitamente, che certo comporta molti benefici e che nessuno discute, e il futuro servizio a pagamento in regime monopolistico, che è l'oggetto in discussione. Per quest'ultimo, i benefici sono molto più limitati e i rischi evidenti sotto il profilo sia giuridico sia economico (sia per il pubblico che per i detentori di diritti), mentre esistono varie possibili strade alternative, sia per il materiale che fa ancora parte del circuito commerciale (e che quindi è spesso già accessibile a pagamento per altri canali, che si trovano facilmente – ironia della sorte – proprio con Google), sia per le varie fasce (sulla cui consistenza esistono stime molto divergenti) di materiale non più in commercio. La pretesa inesistenza di alternative è spesso, a ben guardare, il principale, se non unico, argomento a favore della proposta di Google.

Le assicurazioni non sono gratuite, ma il prezzo che si paga è molte volte inferiore al danno che si rischia, che può andare anche al di là delle proprie risorse.

Il sostanziale monopolio di Google per quanto riguarda la *ricerca a testo completo nel contenuto dei libri* non ha competitori e non sembra che possa averne nel futuro prossimo, ma il rischio che ciò comporta si può considerare relativo, allo stato attuale, perché è molto improbabile che questo servizio cessi di essere offerto gratuitamente da Google, almeno a breve termine. Sicuramente vi sono dei rischi riguardo alla qualità e soprattutto all'“imparzialità” dei risultati¹⁴, ma nel sistema dell'informazione commerciale si tratta di un rischio ordinario.

Molto diverso e maggiore è invece il pericolo di una situazione di monopolio per l'*accesso al materiale digitalizzato*, trattandosi di un servizio già annunciato come in massima parte a pagamento e sostanzialmente privo di altri offerenti in concorrenza. Un quadro sintetico, chiaro ed equilibrato delle prospettive e dei loro pericoli, insieme a sensate proposte per possibili azioni concrete da intraprendere, è stato recentemente esposto da Robert Darnton, che oltre ad essere uno dei più competenti storici e osservatori del mondo del libro ha vissuto in prima persona, come direttore della biblioteca universitaria di Harvard, l'esperienza di un accordo su larga scala tra Google e un grande istituto bibliotecario¹⁵.

In assenza di azioni positive, e quindi innanzitutto di una consapevolezza concreta e diffusa dei problemi e dei rischi che abbiamo di fronte, lo scenario che si prospetta è quello di vedere, in tempi brevi, le biblioteche e le istituzioni di ricerca costrette a tagliare dai propri magri e sempre più martoriati bilanci una fetta consistente, ogni anno, per l'accesso a pagamento a Google Books, fornitore in regime di sostanziale monopolio di un servizio a cui in pratica non si potrà rinunciare. Dato che i loro finanziamenti non aumentano, o aumentano a un tasso inferiore all'aumento di prezzo (e di offerta) dei materiali, quella fetta andrà tolta ad altri acquisti, già ora strozzati dalla pressione dei contratti per le risorse elettroniche e dalla spirale dei prezzi d'abbonamento ai periodici.

Per questi motivi, è sicuramente molto importante non chiudere o strangolare, ma anzi aprire ed allargare, ogni forma alternativa di accesso a contenuti digitali e digitalizzati, da parte delle istituzioni pubbliche ma anche, a mio avviso, delle altre figure coinvolte (editori, distributori, aggregatori, ecc.), e anche nelle modalità più semplici. Non solo, insomma, tramite motori di ricerca o con forme di analisi e indicizza-

¹⁴ È del mese scorso lo scandalo suscitato negli Stati Uniti dalla scoperta che la graduatoria dei libri più venduti era stata truccata, o per meglio dire censurata, da Amazon per far scomparire i libri in qualche modo (e in alcuni casi erroneamente) connessi al tema dell'omosessualità. La notizia (per la cui segnalazione ringrazio Gregorio Montanari) si può leggere a http://news.cnet.com/8301-1023_3-10217715-93.html.

¹⁵ Cfr. Robert Darnton, *Google and the new digital future*, «The New York review of books», vol. 56, n. 20, 17 December 2009, disponibile in rete a <http://www.nybooks.com/articles/23518>. Altri suoi interventi precedenti sono compresi, con una prefazione che ne aggiorna le conclusioni, nel recente volume *The case for books: past, present, and future*, New York: Public Affairs, 2009.

zione integrale del contenuto, ma anche con le modalità più semplici e più tipiche del servizio bibliotecario, quelle che hanno consentito fino ad oggi di accedere liberamente a libri e riviste del passato e di oggi: cataloghi (oggi più potenti e consultabili anche a distanza) e scaffali (oggi anche elettronici e accessibili in rete).

La libera disponibilità di contenuti digitali e digitalizzati, in più sedi e anche nelle forme più semplici, è la migliore assicurazione contro il pericolo di trovarsi, in un futuro molto prossimo, con le mani legate e nessuna possibilità di scelta.

Un'agenda possibile per l'Italia?

Se proviamo a ragionare sulle esigenze e le priorità per il nostro paese, in maniera molto concreta e realistica (il Web è disseminato di cadaveri di progetti di grandi ambizioni e scarsi o nulli risultati), non sembra impossibile arrivare a definire qualche punto di partenza, anche elementare, su cui si possa concordare.

Per il posto n. 1 della lista proporrei: *Far trovare quello che abbiamo già.*

Il materiale che è già oggi disponibile in rete ad opera di biblioteche e altre istituzioni italiane (e non di Google) non è poco. È difficile, naturalmente, sapere *quanto* sia, ma la cosa ha un'importanza relativa perché viviamo in un periodo di grandi sparate numeriche che sono poi di scarsa credibilità e soprattutto irrilevanti per i risultati concreti.

Quello che sicuramente conta, per l'utente, non è il numero sbandierato di documenti disponibili, ma il fatto di poter andare su *una* fonte di riferimento, un sito leader, e trovarvi *almeno un po'* del materiale di cui ha bisogno, così da ridurre significativamente l'onere della sua ricerca.

È chiaro che questo è un elemento di forza assolutamente primario di Google (che non ha nemmeno bisogno di farsi trovare, anzi è il posto da cui vengono trovati gli altri), ma tanti altri successi in rete, da Wikipedia ad Amazon o a YouTube, mostrano quanto l'utilizzatore di Internet tenda a concentrarsi su *una* fonte di riferimento (principale).

Ma il materiale di cui disponiamo dove lo si vede, o come lo si trova? I tanti siti o portali in vario modo connessi alle esperienze italiane di realizzazione e messa a disposizione di contenuti digitali (Biblioteca digitale italiana, Internet culturale, CulturalItalia, ecc., senza dimenticare l'Indice SBN e, su scala più vasta, Europea) non segnalano che frazioni minuscole di quanto è effettivamente disponibile, nell'ambito qui considerato. Anche tramite i motori di ricerca questo materiale emerge di rado, probabilmente per la maniera in cui sono organizzati i siti che lo contengono.

Se si eccettua forse l'*Emeroteca digitale* della Biblioteca Braidense, realizzazione di grandi dimensioni e quindi abbastanza nota (ma davvero a tutti?), spesso in evidenza nei risultati di Google e segnalata (anche se in posizione infelice e poco visibile) nell'Indice SBN, quasi tutto ciò di cui disponiamo è sparso, poco noto,

difficilmente individuabile. Anche collezioni importanti come quella dei *Periodici italiani digitalizzati* della Biblioteca di archeologia e storia dell'arte è difficile che appaiano in buona posizione in una ricerca fatta tramite un motore. Restando nell'ambito dei periodici, molto spesso il singolo istituto rende disponibile un piccolissimo numero di titoli (per esempio quelli dell'ex Istituto di patologia del libro e dell'Istituto centrale per il catalogo unico, nei loro siti, e periodici locali nei siti di numerose biblioteche, dai giornali della Biblioteca universitaria di Pisa all'Emeroteca del Sistema bibliotecario di Brindisi per le riviste pugliesi), e anzi abbondano siti dedicati a (o che comprendono) un solo periodico, per quanto importante (dalla «Rivoluzione liberale» alla «Rassegna storica del Risorgimento», da «Emporium» a «Capitolium» e al «Marzocco»).

Dal punto di vista della ricerca, non fa differenza che questi periodici siano messi a disposizione nell'ambito dei comuni progetti di digitalizzazione retrospettiva di una singola biblioteca, oppure dall'ente che produce il periodico stesso (molti enti pubblici, a partire dal Mibac e dai suoi istituti, ma anche enti di natura privata, associazioni, istituti di ricerca, ecc.), o come risultato di progetti di ricerca di ambito universitario (per esempio il sito dedicato dalla Scuola normale di Pisa a «Emporium», ora citato, ma per la sua ricchezza va ricordato soprattutto CIRCE, il Catalogo Informatico Riviste Culturali Europee, realizzato dall'Università di Trento e contenente la digitalizzazione integrale di una cinquantina delle più importanti riviste letterarie del Novecento), o per altre strade ancora. Inoltre, questo fenomeno non riguarda solo il periodo retrospettivo, il materiale libero dal diritto d'autore, ma anche molti periodici correnti, anche se spesso in maniera non completa o con l'esclusione di annate o numeri più recenti: si possono citare come esempi i siti di alcuni grandi quotidiani, quelli di «Biblioteche oggi», degli «Annali di storia delle università italiane» e di tante altre importanti riviste specializzate. È il caso di ricordare anche la grande quantità di pubblicazioni correnti di notevole interesse che vengono ormai abitualmente messe a disposizione gratuitamente in rete da enti pubblici: per esempio gran parte delle pubblicazioni dell'Istat, una parte di quelle della Camera e del Senato, molte di quelle degli archivi di Stato, senza dimenticare le recenti iniziative dell'Istituto della Enciclopedia italiana (a partire dalla digitalizzazione delle voci del *Dizionario biografico degli italiani*).

Oltre al materiale liberamente disponibile, naturalmente, c'è una vasta offerta di periodici accessibili a pagamento, presso gli editori o distributori, qui non considerato, ma che costituisce comunque una risorsa importante presente in siti italiani e, almeno allo stato attuale, non altrove.

I siti pubblici del nostro paese, insomma, offrono giornali e periodici italiani digitalizzati in maggiore quantità rispetto a Google Books o a Internet Archive.

Ma, si può obiettare, questo vale per i periodici, non per i libri. È probabile, ma quella che si potrebbe battezzare la "prima legge della ricerca" (e che richiama la seconda e la quarta legge di Ranganathan), ossia "riduci il tempo e la fatica di

procurarsi il materiale che occorre”, ci dice che l’importante è ottenere facilmente una parte, il più possibile, di quanto serve, e non conta quale parte sia. Censire quanto oggi disponibile per le monografie, dalle cinquecentine segnalate in Edut16 ai servizi a pagamento di editori (come Il Mulino) o distributori (come Casalini), è certo più difficile che per i periodici, ma si tratta di un panorama che si va rapidamente arricchendo. D’altra parte, come si è visto, quanti e quali libri italiani potranno essere messi a disposizione nel prossimo futuro, anche a pagamento, da Google (a testo completo, al di là della semplice ricerca), è questione molto incerta e opinabile, non compresa nel *Settlement* e affidata ad accordi specifici tra la società e gli editori.

La prima, più semplice e più risolutiva risposta alla priorità proposta come n. 1 potrebbe quindi essere, a mio parere, *segnalare tutto il materiale disponibile* (o almeno quello liberamente accessibile) *nell’Indice SBN*.

Non credo che ci sia bisogno di dimostrare che l’Indice SBN è, di gran lunga, la fonte di riferimento più vasta e organica, nel nostro paese, per la ricerca di documenti pubblicati a stampa.

La segnalazione della disponibilità di risorse digitali, corrispondenti in misura più o meno fedele e in forma anche parziale a pubblicazioni a stampa, è largamente prevista, anche per situazioni differenti e con varie semplici modalità (nel record bibliografico della pubblicazione a stampa e non solo con record separati), dalle nuove *Regole italiane di catalogazione* (REICAT)¹⁶.

Naturalmente la disponibilità di risorse digitali connesse a pubblicazioni a stampa potrebbe essere segnalata anche in altre sedi (siti specifici dedicati solo al digitale, iniziative di cooperazione su scala più larga come Europeana, la stessa Biblioteca Aperta basata su Internet Archive, e ovviamente anche Google), ma questo non toglie che la segnalazione non dovrebbe mancare, in primo luogo, nel principale punto di riferimento italiano per il materiale bibliografico, che è l’Indice SBN. E magari, per quanto riguarda i periodici, anche in ACNP.

Questo sarebbe di grandissima utilità non solo per chi fa ricerca, ma per le biblioteche stesse. Segnalando ai loro utenti che certo materiale è disponibile in rete (e non importa se presso una biblioteca che partecipa al sistema, presso l’ente o l’editore che lo produce, in un sito accademico come CIRCE e magari anche in Internet Archive), non solo si facilita il lavoro all’utente, che sarà ben contento di poterselo scaricare e salvare quando e come vuole, ma si riducono oneri e inconvenienti per la biblioteca stessa. Ai tanti laureandi e dottorandi in letteratura italiana, storia, storia dell’arte e molti altri campi, si eviterebbe di dover spiegare che certo materiale (per esempio le principali riviste italiane dal Settecento al

¹⁶ *Regole italiane di catalogazione: REICAT*, a cura della Commissione permanente per la revisione delle regole italiane di catalogazione, Roma: ICCU, 2009. Le indicazioni generali sulla segnalazione di versioni digitali di una pubblicazione a stampa (o esistente anche a stampa) sono comprese nel par. 1.7.1.5, con esempi ripresi in maniera più specifica nel par. 4.7.8 B.

Novecento) è soggetto a particolari cautele o è stato immagazzinato altrove, di dover movimentare materiale ingombrante e fragile, o di ricorrere alla macchinosa consultazione e riproduzione di microfilm (che per sua natura porta via molto tempo agli utenti e al personale), per non parlare di costose e lente richieste di fornitura di riproduzioni o di prestito interbibliotecario.

Altri due aspetti di notevole importanza, a favore della valorizzazione di quanto in rete *già c'è*, mi sembrano i seguenti:

- un gran numero di produttori di pubblicazioni interessanti per la ricerca è costituito da enti pubblici (organi costituzionali, amministrazioni centrali, istituti nazionali come l'Istat e l'Istituto superiore di sanità, ecc.) e organizzazioni senza scopo di lucro (istituti culturali, associazioni scientifiche e professionali, ecc.), e tra questi molti sono già inseriti nella rete SBN tramite le loro biblioteche, o hanno rapporti organici con la Direzione generale per le biblioteche (per esempio nel caso degli istituti culturali) o con l'Istituto centrale per il catalogo unico (per esempio per la partecipazione a iniziative anche diverse da SBN, come il Censimento delle cinquecentine);
- la produzione di tutte queste organizzazioni non solo è quantitativamente molto rilevante (rappresenta probabilmente una quota approssimativamente pari a quella dell'editoria strettamente commerciale, ossia svolta da imprese, con fini di lucro, che hanno l'editoria come proprio ramo primario di attività), e spesso meno facilmente reperibile in forma cartacea, ma è evidentemente di grande interesse qualitativo, perché prodotta da organizzazioni che ne garantiscono la qualità scientifica, ed è quindi di grande utilità per la ricerca (mentre nella produzione editoriale commerciale vi sono larghi settori che hanno un interesse solo occasionale e marginale per la ricerca, per lo più come oggetto di studio invece che come letteratura primaria).

Se la prima priorità è indubbiamente quella di *un punto di riferimento principale*, una fonte a cui rivolgersi in prima battuta, come dimostra facilmente la fenomenologia dell'uso di Internet, l'elemento decisivo che viene immediatamente dopo è quello della *quantità*. Non della qualità, ma della quantità. Che ci piaccia o no, è così. E probabilmente ci sono buoni motivi perché sia così, se torniamo alle considerazioni fatte al principio.

La quantità è oggi, in moltissimi casi, gestita come un elemento promozionale, di scarsa affidabilità e rilevanza, come si è già notato. Al di là dei roboanti comunicati stampa sui milioni di questo e di quello, abbiamo visto che il materiale già disponibile e accessibile nel nostro paese non è poco, anzi è molto.

Ovviamente, però, questo materiale va incrementato, e questa è sicuramente la *seconda priorità* (dopo il rendere facilmente accessibile, in *un punto* – anche in più d'uno, se si vuole, ma almeno in uno –, quanto *già c'è*).

Anche sotto questo profilo, il quadro *non è così disperante* come spesso si crede. È notissimo – almeno a chi si occupa di queste cose – che se il materiale che può occasionalmente occorrere a qualcuno è sterminato, quello su cui si concentra una larga quota delle esigenze, e dell'uso, è molto ridotto. L'antica e famosa «regola 80/20» ci ricorda che, in generale e molto approssimativamente, l'80% delle esigenze si concentra sul 20% del materiale. Le recenti chiacchiere sulla «coda lunga», inconsistenti dal punto di vista statistico, non hanno cambiato affatto questo fenomeno, che è (relativamente) indipendente dalla scala: vale, per esempio, sia in una biblioteca di 10.000 volumi che in una di dieci milioni. Non sarebbe poi tanto difficile, quindi, andare a identificare delle priorità, individuare per esempio per settori e periodi quali sono i materiali (innanzitutto i periodici e le opere di consultazione) più frequentemente cercati e usati, così da conseguire notevoli benefici per la ricerca con investimenti molto modesti.

Con l'occhio alla *quantità* (ma più avanti si parlerà anche di qualità), ho l'impressione che si debbano riconsiderare un po' anche le scelte operative di digitalizzazione, con i relativi costi e tempi. È un campo in cui non ho competenza, ma è un fatto, per esempio, che la digitalizzazione veloce da microfilm, su larga scala e a basso costo, ha permesso di dar vita all'unico sito europeo di grande rilevanza in questo settore e di largo uso, Gallica, che probabilmente rimane tuttora la principale fonte di digitalizzazioni integrali di libri a stampa antichi per gli studiosi (in Google Books il materiale antico è almeno per ora poco rappresentato). Le digitalizzazioni di Gallica sono di solito di bassa qualità, certo, oltre che piene di righe dovute all'usura dei microfilm, ma sono comunque sufficienti per la maggior parte delle esigenze.

In poche parole, la *quantità* fa risparmiare al ricercatore molto più tempo della *qualità*: trovare molte delle cose di cui si ha bisogno (invece di poche) fa risparmiare tempo, la qualità anche bassa delle immagini è sufficiente per la massima parte delle esigenze, mentre quando le esigenze sono molto elevate (per esempio per certi lavori di bibliografia testuale, ma si tratta di casi molto rari) è probabile che anche immagini di alta qualità siano insufficienti per il ricercatore, che dovrà quindi andare a esaminare anche l'originale.

Oltre alla qualità delle immagini un onere rilevante è sicuramente costituito dall'elaborazione dei necessari metadati. Anche in questo campo non ho particolare competenza, ma la mia impressione è che le esigenze veramente importanti siano molto limitate e che quindi sia opportuno destinare le risorse (scarse o scarsissime) ad *aumentare la quantità* di materiale disponibile, piuttosto che ad ampliare o perfezionare le informazioni su di esso. Mi riferisco a digitalizzazioni di materiale a stampa, non al materiale digitale nativo, per il quale ci sono esigenze di conservazione (in senso lato) riguardo alle quali i metadati rivestono un'importanza molto maggiore, ben diversa rispetto alle digitalizzazioni finalizzate alla fruizione di materiali per i quali si conserva l'originale (e si dispone già della sua descrizione).

Sarà bene anche chiarire – richiamando il quadro tracciato al principio – che la disponibilità di risorse digitali è utilissima *qualunque sia il livello delle informazioni fornite*, purché il materiale sia agevolmente rintracciabile. In moltissimi casi, come si è notato al principio, chi fa ricerca dispone già degli elementi d’identificazione di ciò di cui ha bisogno (per esempio, vuole vedere l’articolo pubblicato alle p. 481 e seguenti della «Nuova Antologia» del 16 dicembre 1937) e un elenco di testate e delle relative annate è perfettamente sufficiente per le sue esigenze (come un tradizionale catalogo dei periodici di una biblioteca). Analogamente, per chi deve vedere un particolare libro (per esempio, il quarto volume de *La letteratura della nuova Italia* di Croce o la prima edizione del *Robinson Crusoe*), la ricerca offerta da qualsiasi OPAC è perfettamente sufficiente. In Google Books e in altri grandi siti mancano invece buone funzionalità per svolgere questo semplicissimo compito: le annate dello stesso periodico o i volumi della stessa pubblicazione non sono raggruppati tutti insieme, separati da pubblicazioni diverse, né presentati in ordine, evidenziando i loro elementi distintivi. Spesso l’elemento più necessario (l’annata o il volume) manca del tutto nella prima visualizzazione dei risultati, o non è attendibile. Siti che invece offrono semplicemente un ordinato “scaffale elettronico”, come DigiZeitschriften per un buon numero di riviste tedesche, sono già perfettamente sufficienti per molte esigenze e comodi di usare (anche se poco conosciuti, non avendo un interesse commerciale e non beneficiando di una massiccia pubblicità diretta o indiretta).

Quando è possibile offrire un livello superiore d’informazioni (per esempio la ricerca su autori e titoli degli articoli contenuti in un periodico) meglio, ovviamente, ma non è indispensabile. Se poi dalla digitalizzazione si è ricavato anche il testo integrale, e lo si è reso ricercabile, meglio ancora, ma questo è ancor meno indispensabile, e si tratta del terreno meno rilevante per le biblioteche, per il quale sono meno attrezzate e per il quale invece – ma solo per questo – sono molto meglio attrezzati altri, e in primo luogo Google.

Un’ultima osservazione può essere utile fare riguardo alle modalità di utilizzazione. Quello che colpisce a prima vista chi utilizza materiale digitalizzato nei siti italiani e nei grandi siti americani è la differenza fondamentale di approccio al materiale stesso. Quasi sempre, nel secondo caso, viene immediatamente presentata la digitalizzazione integrale di un’unità fisica (il volume), con la possibilità di procedere dalla prima pagina (si cerca di solito di posizionarsi automaticamente alla pagina iniziale più significativa, ossia al frontespizio e non alla copertina o a una pagina bianca preliminare) in avanti, quanto si vuole, di saltare più avanti, e di scaricare o salvare l’intera digitalizzazione. Allo stesso modo è impostato anche Gallica, per la sostanza, con l’utile e chiara opzione del salvataggio di una sequenza di pagine definita dall’utente. Con l’attuale disponibilità di linee veloci e la capienza dei dischi dei pc questa è chiaramente la soluzione più comoda; l’opzione offerta da Gallica può interessare, come ulteriore alternativa

allo scarico integrale, non tanto per i vantaggi “economici”, a mio parere, quanto per la comodità e rapidità d’uso del file scaricato da parte dell’utente, se interessato per esempio solo ad un articolo.

Nei siti italiani, invece, non accade quasi mai che sia scaricabile un intero volume o annata, e spesso nemmeno un intero articolo di una rivista, ma soltanto singole pagine o immagini. Alla visualizzazione desiderata si arriva per lo più dopo una serie defatigante di schermate “gerarchiche” (a volte aggravate da una lenta progressione della chiamata delle pagine una ad una, quando l’utente non sa già esattamente il numero). A questo inconveniente si aggiungono spesso modalità di visualizzazione poco ergonomiche, un deliberato peggioramento della qualità dell’immagine (definizione, contrasto, sovrapposizione di filigrane e marchi) tale da rendere a volte difficile anche la semplice lettura, e procedure ostiche e macchinose per lo scarico o il salvataggio, fino a veri e propri blocchi, non sempre facili da eludere, perfino per la copia di singole pagine o brani di testo (e stiamo parlando, naturalmente, di materiale fuori diritti d’autore). Non citerò esempi perché non vorrei assolutamente demotivare chi si è sobbarcato i tanti oneri di mettere a disposizione in rete del materiale: gli esempi abbondano e chiunque, se non li ha già presenti, può trovarli anche tra quelli citati in questo stesso articolo. In questo fenomeno probabilmente gioca un’assurda, irrazionale paura che la disponibilità di materiale liberamente scaricabile in rete comporti chissà quali (inesistenti) pericoli o danni. È una paura che può essere comprensibile per materiale *venduto* al singolo utente, ma che non ha ragion d’essere per materiale liberamente accessibile. Un’avvertenza che proibisse qualsiasi utilizzazione commerciale del materiale offerto sarebbe più che sufficiente, oltre che più appropriata dei tanti macchinosi ostacoli posti a chi lo utilizza per fini di studio. Comunque, le interfacce attuali nei siti italiani sono nel complesso pochissimo idonee a svolgere in maniera rapida e comoda le attività tipiche di chi utilizzi quei materiali per fini di ricerca.

A parte la qualità delle interfacce, si è parlato fin qui essenzialmente di *quantità*: quantità di materiale disponibile, quantità di materiale rintracciabile con un’unica ricerca in un’unica fonte, quantità di materiale che si riesce a recuperare e utilizzare in un certo tempo. E la *qualità*?

Anche di qualità abbiamo bisogno, ovviamente, ma soprattutto per gli aspetti che abbiamo visto essere prioritari. La qualità è fondamentale innanzitutto nella *selezione* del materiale da digitalizzare e mettere a disposizione. I punti di vista personali, ovviamente, possono essere differenti, ma non sarebbe tanto difficile individuare il materiale più importante, più spesso cercato, di cui manchiamo, mentre purtroppo troviamo tante duplicazioni, spezzoni praticamente inutilizzabili, scelte francamente marginali che non costituiscono un buon impiego di risorse scarse. Insieme alla selezione, elemento fondamentale di qualità è la *completezza* (dei volumi, delle annate, delle collezioni, ecc.), tanto tra-

scurata, per motivi diversi, sia nelle grandi imprese come Google Books sia nelle piccole iniziative locali¹⁷.

Si può capire che la selezione e la completezza, come l'esattezza dei dati bibliografici, siano trascurate da chi digitalizza libri a qualche centinaio di migliaia alla volta, con risorse finanziarie che dal nostro punto di vista sono fantascientifiche e con un approccio "generalista" molto diverso da quello della ricerca. Qualità della selezione e attenzione alla completezza, oltre che controllo e precisione nei dati, sono invece da sempre punti di forza delle biblioteche (almeno di quelle ben gestite).

Conclusioni? Per prima cosa, non credere a tutto ciò che ci viene raccontato, perché è molto spesso inesatto (anche interamente, o per il 99,6%) o superficiale, ma andare almeno ogni tanto a verificare con i propri occhi, e non visitare solo il luogo che si fa (o a cui viene fatta) più pubblicità, ma anche qualche altro, che si fa meno pubblicità, o non se ne fa affatto, perché non ha scopi commerciali (o perché è un *business* meno potente).

Seconda cosa, non pensare che qualsiasi buon servizio richieda una barca di soldi e mirabolanti tecnologie, perché spesso con l'una e l'altra cosa si sono prodotti risultati praticamente trascurabili, mentre ci sono ottime cose che si possono fare in maniera semplice e con risorse modeste, se si hanno obiettivi chiari e che possano stimolare una larga cooperazione.

The Google Book Search project is at the centre of a lively debate at the world level - a debate tackling different aspects of the issue and looking at it from various perspectives. The present article tries to discuss in practice - and through a number of examples - what the system offers from the point of view of scholars and researchers, compared with services providing digital resources for free (and especially services offering integral digitised versions of books and periodicals free of copyright). In assessing the Google Book Search project, we have to take into account three different components or functions: the search engine for the complete text, undoubtedly representing the most innovative feature; the "catalogue" (i.e. bibliographic data and their related search functionalities); and the digital library (i.e. the complete body of digitised documents that can be visualised). The article thus examines future perspectives for what concerns the pay per view service, and the risks it could entail. On a final note, the paper advances some considerations on the needs and priorities around which a national policy to enhance the accessibility of Italy's current digital resources should be developed - a policy that could be implemented even with limited resources.

¹⁷ Per esempio, utilizzando con notevole pazienza (per duplicati, errori, record privi di indicazione dell'annata, ecc.) Internet Archive possiamo trovarvi la digitalizzazione integrale dei primi 305 volumi della «Nuova Antologia», ma con 8 volumi mancanti, e dei primi 76 volumi del «Giornale storico della letteratura italiana», ma con 4 volumi mancanti. In Google Books sono attualmente visibili solo pochi dei primissimi volumi della prima rivista. Nessuna delle due, che io sappia, è disponibile in siti italiani.