

Dig *Italia*

Anno IV, Numero 2 - **2009**

Rivista del digitale nei beni culturali

ICCU-ROMA

Google Books e il patrimonio librario italiano

Antonia Ida Fontana – Pierantonio Metelli

Biblioteca nazionale centrale di Firenze

La proposta recentemente avanzata da Google per realizzare un accordo con le più importanti biblioteche europee, biblioteche nazionali ovviamente in testa, per la digitalizzazione del loro patrimonio librario e la sua successiva accessibilità attraverso il servizio Google Books, ha prodotto un largo dibattito, che, oltre alle sedi più propriamente istituzionali, ha investito anche le pagine dei grandi quotidiani nazionali. Questo è un segno della rilevanza del fatto. Vi sono due temi fondamentali che hanno portato al centro dell'attenzione dei media il mondo delle biblioteche, abitualmente ignorato o preso in considerazione solo per le problematiche connesse alle limitazioni della erogazione dei servizi all'utenza a causa della carenza di risorse disponibili. In primo luogo la considerazione che se tale progetto fosse concretamente realizzato si darebbe finalmente risposta, indipendentemente da chi e perché fornisce il servizio, a quella richiesta di immediata fruizione di contenuti digitali di cui l'era della rete e le politiche bibliotecarie stesse hanno suscitato l'aspettativa senza peraltro riuscire poi a riempirla di contenuti. In fondo un generico utente, se usa abitualmente Google per reperire ormai quasi ogni tipo di informazione in rete, non si porrebbe molte domande se lo stesso Google gli fornisse direttamente i contenuti che cerca, in questo caso la possibilità di vedere, istantaneamente e ovunque egli sia nel mondo, milioni di volumi direttamente sullo schermo del suo computer. In secondo luogo i commentatori più accorti hanno visto in questa operazione "democratica", oltre al rischio della creazione di una posizione dominante sul mercato, anche una specie di pericolo di crollo delle culture nazionali: il gigante di Mountain View andrebbe a costituire una sorta di biblioteca universale delle culture nazionali, espresse dai libri che tali culture hanno prodotto nel tempo, una specie di *mare magnum* in cui le distinzioni, la storia, la originalità e la natura delle diverse produzioni culturali si appiattirebbero tendendo a scomparire.

Infine le biblioteche, che non sono generici ipotetici utenti del servizio Google Books, si pongono giustamente tutta una serie di altre domande, che hanno a che fare con la loro funzione, o meglio con la funzione che hanno avuto fino a questo momento dopo l'avvento della rete, con i loro compiti istituzionali, con il fatto di voler difendere la parte scientifica della loro attività, con le più che legittime problematiche di conservazione e così via. Indipendentemente dalla posizio-

ne che si può avere sulla questione e dall'esito che la trattativa per l'accordo avrà, bisogna riconoscere senza mezzi termini che l'ingresso di Google in questo settore, con la sua capacità di investimenti, ha prodotto un'accelerazione impensabile soltanto pochi mesi fa sulla problematica del digitale, un'accelerazione a cui le Amministrazioni centrali nazionali ed europee della politica bibliotecaria pubblica sono costrette a rispondere o almeno devono lanciare delle controproposte operative e credibili, perché ciò che è in discussione non è un dibattito teorico o squisitamente professionale su come costruire la "biblioteca digitale", ma una opzione concreta che in pochi anni potrebbe realmente cambiare le carte in tavola e realizzare ciò che, certamente in modo diverso e con interlocutori diversi, per molti anni è stato al centro di dibattiti, progetti, prototipi, iniziative parziali. La sensazione, comunque si voglia affrontare la questione, è quindi che l'ingresso di Google in questo settore fornisca comunque, in un modo o in un altro, un'opportunità. Ora che un nuovo e più avanzato scenario è stato improvvisamente aperto, pur con tutte le problematiche che comporta, quello che sarebbe inaccettabile pensare è che in fondo si possa anche ignorarlo e andare avanti senza il progetto avanzato da Google o senza un altro progetto alternativo ma di portata simile. Si viveva anche, e può darsi meglio, con cinema e televisione in bianco e nero, o magari solo con la radio, senza gli *smartphone* e le reti satellitari, o con il treno a carbone e senza automobili e così via. Ma le cose avanzano ed una sfida che ha contenuti vincenti può essere sostituita da un'altra equivalente, ma non azzerata.

Bisogna anche obiettivamente riconoscere che nonostante i richiami ad Europea come al progetto antagonista a quello di Google, Europea è oggi sostanzialmente un aggregatore di metadati che si arricchisce solo grazie all'apporto di contenuti digitali provenienti dalle varie istituzioni europee (biblioteche, musei, ecc.) che vi partecipano. Poiché attualmente i progetti di digitalizzazione europei hanno subito, anche in paesi che vi hanno molto investito quali la Francia, forti riduzioni a causa della mancanza di risorse disponibili, sembra opportuno cogliere le occasioni offerte dai privati, seguendo anche le sollecitazioni più volte espresse dalla Commissione Europea per stabilire partnership pubblico-privato che diano modo di affrontare la vastità dell'operazione, della quale il solo settore pubblico non sembra in grado di farsi carico.

Il Ministero per i beni e le attività culturali, come peraltro analoghi organi centrali di altri stati europei, sta valutando la possibilità di partecipare all'iniziativa promossa da Google Books. La Biblioteca nazionale centrale di Firenze (BNCF), insieme alla nazionale di Roma (BNCR) ed all'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), è stata conseguentemente coinvolta nelle riflessioni intorno a questa ipotesi, in modo da verificare la fattibilità della partecipazione italiana al progetto.

La prima cosa che si deve rilevare è che l'ipotesi dell'eventuale accordo con

Google è spezzata in due parti: una relativa alle opere di pubblico dominio (che sono state fissate come quelle pubblicate in Italia tra il 1465 ed il 1868), l'altra relativa ad opere in commercio o comunque sottoposte a diritti. È bene chiarire che la discussione fino a questo momento ha riguardato concretamente solo il primo dei due punti, essendo il secondo molto più complesso per sua stessa natura e per il necessario intervento di altri soggetti, ovvero delle organizzazioni che rappresentano gli aventi diritto (Associazione italiana editori – AIE, Società italiana degli autori ed editori – SIAE, ecc.) e soprattutto dei singoli editori. In quest'ultima direzione la BNCF, insieme all'ICCU, collabora al progetto ARROW (Accessible Registries of Rights Information and Orphan Works towards Europea), del quale condivide finalità e metodologia, ed attraverso la Foundation Conference of European National Librarians (CENL) è in rapporto con la Federation of European Publishers (FEP) con la quale si raggiungono abitualmente posizioni comuni. Bisogna anche riconoscere però che ARROW, come sempre accade per i progetti Europei, è allo stato attuale sostanzialmente un prototipo e non una procedura operativa.

All'interno delle opere sottoposte a diritti, sarà necessario avviare una discussione a parte per quanto riguarda le opere fuori commercio e le opere orfane. La BNCF condivide gli obiettivi e promuove l'uso per il futuro del *Modello di accordo per una licenza di digitalizzazione di opere esaurite*, pubblicato sul numero 2 (dicembre 2007) di «*Digitalia*», così come sta promuovendo la formazione di una Commissione nazionale dedicata allo studio di questi due fondamentali temi.

Mentre per quanto riguarda le opere fuori commercio solo i detentori dei diritti devono poter decidere sulla loro digitalizzazione e relativa remunerazione, per le opere orfane, conformemente con quanto è allo studio del progetto ARROW, occorre un atteggiamento più flessibile che consenta la digitalizzazione delle opere a condizione che vengano fatte ricerche accurate ed in buona fede sui detentori dei diritti (in questo senso ad esempio la BNCF è disponibile ad effettuare le ricerche sui propri fondi, dare pubblicità all'iniziativa e pubblicare sul proprio sito Web gli elenchi delle opere orfane), salva sempre la possibilità di questi ultimi, ove comparissero, di chiedere la rimozione della copia digitale o la remunerazione economica.

La Biblioteca nazionale centrale di Firenze si è quindi impegnata in una riflessione costruttiva, ovvero fissare una serie di punti ritenuti significativi nell'ipotesi di andare verso un accordo con Google relativamente alle opere di pubblico dominio. I più importanti possono essere riassunti nei seguenti:

- le digitalizzazioni devono essere fruibili gratuitamente oltre attraverso Google Books anche tramite i siti Web delle biblioteche proprietarie degli originali (con copia delle digitalizzazioni residenti sui propri server o attraverso un'apposita interfaccia di programmazione – API, Application Programming

- Interface – che visualizzi le digitalizzazioni attraverso i siti Web delle biblioteche, pur essendo le copie digitali fisicamente residenti sui server di Google);
- le immagini devono contenere un elemento identificativo circa la loro provenienza, ad esempio un logo delle biblioteche;
 - all'interno del servizio Google Books dovrà essere definito in modo opportuno l'algoritmo di *ranking* dei risultati delle ricerche, in modo tale che il primo ed il secondo risultato conducano alla disponibilità in forma digitale dell'opera richiesta rispettivamente all'interno del servizio stesso e presso il sito Web della biblioteca proprietaria;
 - il servizio dovrà essere fornito da Google su base gratuita, ma qualora dovessero porsi in essere delle condizioni attraverso le quali Google richieda delle forme di compenso all'utente (ad es. il download) dovranno essere contrattate delle *royalty*;
 - Google non può utilizzare le digitalizzazioni per scopi diversi dalla messa in rete, ogni altro uso dovrà essere autorizzato;
 - si pensa ad un tempo massimo di sei anni di validità contrattuale;
 - le biblioteche mantengono tutti i diritti sulle opere di loro proprietà e potranno sfruttarle economicamente (riproduzioni, facsimili, merchandising, ecc.);
 - le biblioteche potranno utilizzare le immagini digitalizzate per attività e progetti legati al *free source* (Europeana, progetti di ricerca, ecc.), mentre si impegnano a non stabilire accordi della stessa portata di quello oggetto della trattativa con Google con società di fatto concorrenti sul mercato;
 - Google consegnerà, per scopi di conservazione, una copia in alta definizione di tutte le digitalizzazioni effettuate, una copia in bassa definizione da utilizzare per la messa in rete sui siti Web proprietari, e l'OCR (Optical Character Recognition) effettuato sulle opere digitalizzate;
 - al fine di non ipotizzare cessioni unilaterali e comunque per non porre una posizione dominante di Google, si ritiene utile che le biblioteche stabiliscano accordi non relativi a tutto il materiale posseduto. Le selezioni dei materiali verranno effettuate dalle biblioteche coinvolte nel progetto;
 - può essere utile una riflessione sull'esperienza francese condotta a Lione, dove Google si è aggiudicato il progetto di digitalizzazione in seguito ad una gara bandita dall'istituzione pubblica.

Due punti critici sono apparsi però subito chiari. Innanzi tutto nella maggioranza delle biblioteche italiane le opere di pubblico dominio (anteriori al 1868) sono solo in parte catalogate elettronicamente in SBN (Servizio bibliotecario nazionale) e di queste è possibile naturalmente avere dati numerici precisi e differenziati. Tutte le rimanenti sono ancora descritte solo attraverso le vecchie schede cartacee. Per quanto riguarda la BNCF ad esempio si stima che le opere a stampa anteriori al 1868 e possedute in un'unica copia possano raggiungere le 500.000

unità, di cui solo circa 120.000 catalogate in SBN.

È chiaro che per una piena fruizione della digitalizzazione, compatibile con i sistemi nazionali, è necessario che i volumi siano catalogati in SBN e dotati di metadati gestionali e amministrativi. Questa seconda operazione può essere in parte realizzata attraverso un software di estrazione automatica dei metadati, ma la condizione necessaria e indispensabile è la catalogazione. Da questo punto di vista le prospettive di Google e quelle delle biblioteche divergono: Google agisce come non può che agire una società che si muove in base alle leggi di mercato (il miglior risultato del suo prodotto nei tempi più rapidi possibili e con i minori costi possibili, ottimizzando tutte le operazioni necessarie), le biblioteche seguono, a ragione, anche altri criteri. Naturalmente Google accetta di buon grado catalogazioni esistenti, ma non ha alcuna intenzione di farsi carico dei costi di catalogazione secondo gli standard SBN, in quanto per le finalità del suo progetto, che mette a disposizione in linea l'OCR dei testi digitalizzati, sono sufficienti pochi elementi che ne consentano l'indicizzazione. Appare quindi evidente che per mantenere una congruenza del progetto è necessario un forte investimento da parte del Ministero sulla catalogazione (è stata fatta una stima approssimativa che porta ad indicare un cifra non inferiore ai due milioni di euro relativamente ad un insieme di circa un milione di volumi).

Anche la movimentazione dei materiali diventa un altro punto critico. Google, da questo punto di vista, fatte salve le coperture delle spese e delle necessarie polizze assicurative, richiede la massima libertà di azione, arrivando a prefigurare dei centri di raccolta dei materiali e di esecuzione dei lavori al di fuori dei confini nazionali. È chiaro che le biblioteche hanno in proposito tutt'altra opinione e che ritengono che le operazioni di digitalizzazione dovrebbero avvenire all'interno dei singoli istituti, in considerazione della presenza di materiale raro e di pregio nelle proprie collezioni e comunque al fine di poter esercitare un maggiore controllo su tutte le fasi delle operazioni. Inoltre è evidente che anche soltanto il mettere in atto la semplice operazione della movimentazione interna (che implica il prelievo, la preparazione di elenchi, l'apposizione di codici a barre, ecc. e la ricollocazione dei volumi) comporta la necessità di risorse attualmente non disponibili a causa della drammatica carenza di personale. Movimentare centinaia di migliaia di volumi implica la disponibilità di una *task-force* e quindi di adeguate risorse finanziarie. Molte delle parti in causa hanno richiamato, come accennato prima, il rischio che il progetto avanzato da Google prefiguri la costituzione di una situazione di monopolio, o almeno di una situazione fortemente dominante. La preoccupazione è ragionevole, ma allo stesso tempo è vero che la proposta e l'investimento di Google restano di grande interesse. Quello che è necessario per evitare tale rischio è che le biblioteche mantengano saldamente nelle loro mani i diritti sulle opere eventualmente digitalizzate da soggetti esterni e che non si configuri una cessione in toto del patrimonio librario, ma si realizzino accordi differenziati con interlocutori

diversi per natura, specializzazione e finalità in relazione alle diverse tipologie di materiali da digitalizzare, costruendo quindi una struttura in cui l'Italia potrebbe cooperare con partner diversi affermando un modello forte ed autonomo di diffusione della cultura. In tal modo le biblioteche potrebbero guidare e gestire l'indirizzo generale della loro politica culturale e costituire inoltre un quadro di riferimento nei rapporti pubblico-privato.

Questa è la direzione verso cui sta cercando di muoversi la BNCF, ovvero puntare a realizzare accordi con partner diversi, ponendo una distinzione tra materiali pubblicati fino al 1700 e materiali pubblicati dal 1701 al 1868. I primi, che hanno elevati costi di catalogazione e necessitano di un trattamento di tipo specialistico, da affidare ad accordi specifici con società chiaramente indirizzate verso la realizzazione di prodotti scientifici rivolti in primo luogo ad un target istituzionale e di ricerca. I secondi, i materiali moderni, da affidare invece al servizio Google Books, a carattere divulgativo e rivolto invece alla generalità degli utenti.

In questa prospettiva la BNCF ha in corso da tempo, per un progetto pilota relativo ad alcune collezioni di materiali antichi, una trattativa, che sembra ora poter andare a buon fine, con ProQuest, società leader nel campo dell'informazione scientifica elettronica per il mondo accademico. Tale accordo presenterebbe dal nostro punto di vista diversi vantaggi ed in particolare garantirebbe:

- maggiori requisiti di scientificità e maggiori garanzie conservative rispetto al trattamento dei volumi rari;
- qualità delle digitalizzazioni superiore a quella offerta da Google (i master delle digitalizzazioni in altissima definizione restano di proprietà della BNCF sia per usi di conservazione che per fini commerciali e viene fornita una copia in microfilm come ulteriore supporto di conservazione). Google pone degli standard rigidi sui formati delle opere che possono essere digitalizzate: questo elemento, che sul moderno può apparire secondario, diventa invece rilevante sull'antico, dove si presenta un'ampia varietà di materiali (volumi in grandi formati, con tavole, cartelle, ecc.), spesso eccedenti le specifiche di Google ma di grande interesse per gli studiosi;
- costi di catalogazione e di produzione dei metadati totalmente a carico della società contraente;
- movimentazione delle opere a carico della società contraente con la supervisione ed il controllo esercitato dalla biblioteca;
- digitalizzazioni sempre effettuate all'interno delle strutture della biblioteca, il che darebbe soluzione a molte delle legittime preoccupazioni, in misura crescente con la rarità dei materiali coinvolti, relative alle istanze di conservazione;
- gratuità dell'accesso alle digitalizzazioni per tutte le istituzioni ed i privati che si collegano in rete dal territorio italiano, mentre il prodotto contenente le di-

digitalizzazioni realizzato dalla società contraente potrà essere venduto all'estero. Entro un periodo fissato (si pensa a non più di dieci anni) le digitalizzazioni dovranno essere rese disponibili in un modello di libero accesso su base mondiale;

- una volta recuperato l'investimento da parte della società contraente, corresponsione di una *royalty* (pari al 15%) sugli ulteriori ricavi delle vendite;
- gli studiosi italiani potrebbero fruire gratuitamente oltre che di testi digitalizzati in alta qualità, rispondenti ai loro interessi scientifici, anche di una base dati di riferimento con catalogazioni e notizie bibliografiche di alto valore complessivo.

In questo modo si otterrebbe un risultato che identifica con forza e precisione la storia della cultura italiana, mentre nel servizio che Google sta cercando di allestire è inevitabile che tutto venga in qualche modo fatalmente mescolato e la nostra cultura avrebbe necessariamente minore evidenza e rilevanza. Inoltre con questa strategia si eviterebbe l'accusa di favorire la creazione di una posizione dominante e si affermerebbe il principio della realizzazione di accordi differenziati che consentono di tenere in mano la barra della gestione della propria politica culturale.

Un punto debole di questa prospettiva potrebbe essere individuato in un'ipotetica disparità di trattamento riservato ai cittadini dei vari Stati membri dell'Unione Europea. Ovvero l'accesso gratuito offerto ai cittadini degli stati partner relativamente ai materiali provenienti dalle istituzioni di quegli stessi stati e l'accesso a pagamento per gli altri. Tale questione, presa in considerazione anche dalla Commissione Europea, suggerisce però alcune riflessioni:

- dal momento che la Commissione stessa incoraggia la formazione di *joint-venture* pubblico-privato riconoscendo la impossibilità del solo versante pubblico di varare progetti di digitalizzazione di enormi patrimoni nazionali, appare strano che non vengano conseguentemente riconosciute le ovvie esigenze di mercato di una proposta commerciale. Siamo inoltre convinti che sia possibile arrivare ad un accordo che trovi un punto di incontro con i mandati fondamentali del versante pubblico, realizzando condizioni di reciproco vantaggio. In sostanza pensiamo che si possa ritenere che le dichiarazioni rilasciate in tal senso dalla Commissione suggeriscano più che altro un richiamo ad una posizione di principio;
- questa idea della equità di trattamento, astrattamente ineccepibile, è però appunto molto astratta e non tiene conto che in Europa, in campi ben più significativi, non vi sono affatto attualmente condizioni di parità tra i diversi Stati membri;
- non sembra poi trascurabile il risultato che otterrebbero comunque tutti i cit-

tadini, le biblioteche e le istituzioni culturali italiane dall'accesso gratuito a questi materiali di studio organizzati in modo scientifico. È a tutti noto quanto sia difficile e costoso rendere disponibili e curare la manutenzione in rete di contenuti digitali;

- forse non si tiene nella giusta considerazione il fatto che in ogni caso al termine della limitata validità contrattuale i contenuti digitali dovranno comunque essere resi disponibili in un modello di libero accesso su base mondiale e quindi in modello free source paragonabile a quello attualmente offerto da Google.

Quello che appare chiaramente oggi, in un quadro generale rapidamente evolutosi sulla problematica del digitale e del copyright, è che l'impegno principale dei singoli paesi dovrebbe essere la tutela degli autori, degli editori e delle opere conformemente al diritto nazionale, mentre a livello europeo dovrebbe essere raggiunto l'obiettivo di un'unica e più avanzata legislazione sul copyright che tenga conto dell'innovazione tecnologica e delle nuove possibilità di fruizione dei contenuti digitali da parte di tutti i cittadini, raggiungendo un equilibrio tra i diritti degli autori, i diritti degli editori e quelli degli utenti.