

Il progetto Google Libri: l'esperienza della Biblioteca Nazionale di Napoli

Gennaro Alifuoco

Biblioteca Nazionale di Napoli

È nota la storia, un po' aneddotica un po' leggendaria, alimentata dallo stesso Larry Page – fondatore, assieme a Sergey Brin, di Google – sulla nascita del progetto Google Books¹. Nell'ormai lontano 2002, Page si domandò quando tempo ci sarebbe voluto per digitalizzare un intero libro. Insieme alla storica collaboratrice Marissa Mayer² ci provò lui stesso, armato di scanner e metronomo per dare il giusto ritmo all'operazione. Ci vollero, per 300 pagine, 40 minuti. Si recò allora presso la Biblioteca dell'Università del Michigan, pioniera nei progetti di digitalizzazione, dove apprese che il tempo stimato per l'acquisizione di 7 milioni di volumi in loro possesso era di 1000 anni. «Beh! - disse Page ai dirigenti, immagino alquanto stupiti - noi possiamo farlo in sei anni, mese più mese meno».

Il progetto divenne ufficialmente operativo due anni dopo, nel 2004, nella doppia direzione di accordi con gli editori³ e con le biblioteche⁴. Una strategia aggressiva, che suscitò subito reazioni ostili innanzitutto nel mondo dell'editoria, cosa che di fatto ha limitato le smisurate ambizioni originali del progetto, ma anche nel mondo delle biblioteche, che mal digerivano l'irruzione di questo intruso nei propri territori. Più in generale gli organismi governativi e cul-

turali europei, e anche americani, furono stimolati ad elaborare strategie alternative e concorrenti, volte a contrastare il predominio di aziende private nel controllo delle risorse informative digitali. Alla base di progetti come DPLA (Digital Public Library of America) o Europea, tra gli altri, c'è il proposito di creare depositi istituzionali in formati aperti e strutturati su standard interoperabili, esattamente il contrario di quello che fa Google.

Dopo una serie di contatti e guardinghi avvicinamenti tra Google e il nostro Ministero, anche le istituzioni bibliotecarie italiane furono coinvolte. Nel novembre 2012 fu stipulato l'accordo per la digitalizzazione di ca. 350000 volumi (all'origine si pensò ottimisticamente di raggiungere la cifra di 1.000.000), conservati nelle Nazionali di Roma e Firenze, a cui poi si aggiunse quella di Napoli, con la partecipazione dell'Istituto Centrale per il Catalogo Unico e per le informazioni bibliografiche (ICCU) per la gestione degli aspetti tecnologici e degli standard operativi. Quando fu reso noto l'accordo, tutte le possibili obiezioni rispetto a questa per alcuni innaturale intesa ritornarono a galla. Perché, è ovvio, le biblioteche funzionano come biblioteche, e i bibliotecari ragionano da bibliotecari. Cioè tendono a costruire

¹ Il progetto all'origine fu battezzato Google Print. In quest'articolo lo chiameremo, seguendo la localizzazione italiana, Google Libri.

² La Mayer faceva parte del gruppo dei primi 20 impiegati di Google, e tra questi la prima donna ingegnere informatico. In seguito fu nominata amministratrice delegata di Yahoo.

³ Tra questi Blackwell, Cambridge University Press, McGraw-Hill, Penguin, Springer ecc.

⁴ Le prime furono la New York Public Library, Università di Oxford, Stanford, Harvard e ovviamente Michigan.

organismi funzionanti attraverso complessi apparati di regole, standard, protocolli condivisi, nel quadro di progetti istituzionali. E, in questo contesto, Google Libri poteva apparire come un corpo estraneo.

Ma – lo confesso – nutro un’istintiva simpatia per Google, che prevale sulle inquietudini per gli aspetti pervasivi delle sue attività nella società globale dell’informazione. Ne ammiro la continua spinta all’innovazione, attraverso progetti a volte visionari in ogni campo dello scibile, che stimolano le migliori energie di tanti giovani ricercatori. E a tutte le possibili ed assai ragionevoli obiezioni, sicuramente semplificando e banalizzando, ma ricordando il raccontino iniziale, mi risposi di primo acchito: in quanto tempo alla Nazionale di Napoli saremmo stati capaci di digitalizzare una tale mole di volumi? E saremmo stati altrimenti in grado di affrontare i costi di una tale impresa? La domanda è retorica, le risposte scontate. Ma è il caso di sottolineare alcune cose. Innanzitutto, per le nostre biblioteche, l’operazione non è stata certo a costo zero, come forse all’inizio si illudevano i promotori dell’impresa. Google si è accollata i costi della movimentazione dei volumi, delle operazioni di digitalizzazione e messa in rete (la sua), di assicurazione, e in generale della complessa organizzazione dell’impresa. Ma di certo non poteva accollarsi – e suppongo che non immaginava fossero necessarie – le spese inerenti all’organizzazione interna alle biblioteche. Secondo le stringenti specifiche dell’accordo i volumi da digitalizzare avrebbero dovuto necessariamente essere catalogati in SBN, provvisti di codice a barre, selezionati in base alla data di stampa, controllati nel loro stato fisico, il tutto ad un ritmo di ca. 4.000 unità al mese da lavorare. Di fatto le biblioteche partecipanti non erano di certo pronte a tutto questo quando fu stipulato l’accordo, soprattutto per quanto riguardava la presenza nei cataloghi SBN della quantità di documenti preventivata. Per consentire alle biblioteche di avviare i necessari lavori interni furono per-

ciò stanziati dei fondi dalla Presidenza del Consiglio dei Ministri, a cui si sono poi aggiunti in corso d’opera quelli del MIBACT. Fondi destinati prevalentemente, oltre che agli adeguamenti hardware e software dei sistemi informatici, all’aggiornamento catalografico. A questo scopo la Biblioteca Nazionale di Napoli ha assunto a tempo determinato (12 mesi) 17 catalogatori selezionati attraverso il bando di una *short list* di collaboratori esterni, con specifici titoli per la catalogazione in SBN antico e moderno. Ovviamente, affinché questo impegno economico e organizzativo non servisse solo ad alimentare il serbatoio di Google Libri con il nostro patrimonio librario, sono stati fissati alcuni punti fermi, il primo dei quali è che non ci sarebbe stato nessuno scambio di diritti sui volumi digitalizzati, diritti che rimangono delle biblioteche che li posseggono, il cui unico impegno è quello di non concedere le risorse ad altre società esterne concorrenti con Google. All’interno delle istituzioni bibliotecarie quelle stesse risorse sarebbero state inserite e visibili nelle teche digitali locali, in Internet Culturale e negli altri portali istituzionali collegati, e i link negli OPAC di Indice e locali che puntano a Google Libri progressivamente sarebbero stati indirizzati alle piattaforme istituzionali.

Come abbiamo affrontato alla Biblioteca Nazionale di Napoli questa non facile impresa? Il progetto ci ha imposto, volenti o nolenti, un flusso di operazioni secondo ritmi cogenti e con metodiche per noi non usuali. Preliminarmente è stato necessario individuare i fondi da cui attingere i volumi da digitalizzare, verificando la possibilità di movimentazione nei locali, spesso angusti, in cui erano ubicati. Il nostro Istituto ha reso disponibili per il progetto volumi tratti da alcune delle sue raccolte più rappresentative: il Fondo Farnese, la Biblioteca Lucchesi Palli, il Fondo Aosta, la Provinciale, la Palatina, e altri ancora.

Elenco in sintesi le operazioni che abbiamo svolto, tenendo conto che si è trattato di una vera e propria catena di montaggio che ci ha tenuto impegnati mese per mese - per far fronte al rigido calendario pianificato delle spedizioni verso lo *scan center* che Google ha edificato per il progetto in una località laziale - dall'aprile 2013 all'ottobre 2015 (con una pausa tra ottobre 2014 e marzo 2015, e con un invio supplementare nel settembre 2016). Per la movimentazione dei volumi è stato necessario affidare il servizio ad una ditta esterna specializzata.

I. Fase preliminare

- estrazione dei record bibliografici delle edizioni presenti nel catalogo SBN;
- selezione dei volumi da digitalizzare;
- attraverso l'applicativo SBNweb, a cui è stato necessario apportare una serie di modifiche per le lavorazioni previste dal progetto, filtro dei volumi che rientravano nei limiti cronologici di stampa previsti (inizialmente dal 1600 al 1872, poi esteso dal 1500 al 1873), esclusione di volumi mutili, fuori formato, con tavole ripiegate;
- esame diretto dello stato di conservazione e della dimensione dei volumi, con creazione di liste dei libri esclusi e dei libri eleggibili, ma da sottoporre a piccolo restauro; quando necessario, operazioni di timbratura, sfogliatura e cartellinatura;
- registrazione inventariale del valore economico di ciascun volume ai fini assicurativi, per escludere edizioni di valore superiore a \$10.000,00;
- generazione della lista contenente i volumi scelti dalla biblioteca, e invio della stessa a Google, dopo aver convertito il formato dati UNIMARC a MARC21XML, come richiesto dalle specifiche del progetto.

II. Preparazione delle spedizioni

- creazione e associazione del codice a barre (*barcode*) per ciascun volume;
- stampa dei singoli moduli forniti di *barco-*

de, da porre al posto del libro negli scaffali e all'interno di ogni volume;

- prelievo dei volumi dalle sezioni e dai magazzini, secondo un ordine topografico, con l'inserimento del *barcode*, e trasferimento nel punto di raccolta;
- composizione dei "carrelli" (ogni spedizione si componeva di 20 carrelli forniti da Google) sui quali dovevano essere collocati i volumi nell'ordine stabilito. I singoli carrelli e l'intera spedizione doveva essere corredata dai documenti di accompagnamento elaborati attraverso i codici a barre, che consentivano di tracciare a noi della biblioteca e agli operatori di Google tutte le operazioni di carico e scarico, di verificare eventuali anomalie, di registrare lo stato del volume nella gestione inventariale di SBN in modo da segnalarlo automaticamente come indisponibile alla consultazione, ecc. I carrelli venivano assemblati, fotografati e sigillati in modo tale da assicurare il trasporto dei volumi in tutta sicurezza sul camion diretto verso il centro di digitalizzazione nel Lazio;
- trasporto dei carrelli al punto di carico/scarico;
- carico dei carrelli sul camion della ditta di trasporto.

III. Restituzione del materiale

- ricevimento al punto di carico/scarico dei carrelli restituiti e trasporto nel punto di raccolta;
- esame e confronto dei materiali restituiti con quelli previsti nell'elenco della spedizione
- esame delle condizioni del materiale restituito;
- esame del rapporto redatto da Google
- trasporto dei carrelli ai vari piani del magazzino;
- ricollocazione dei volumi al proprio posto
- scarico dei volumi e attivazione della loro disponibilità all'utenza per i servizi SBN.

Gli addetti allo *scan center* di Google, dopo aver digitalizzato i volumi, indirizzavano i file

prodotti a centri di elaborazione in Europa e negli Stati Uniti per la postproduzione (ottimizzazione delle immagini secondo i protocolli del progetto, riconoscimento ottico dei caratteri, ecc.) e per l'assemblaggio del pacchetto finale (costituito per ogni volume dalle pagine digitalizzate in formato JPEG 2000, dai metadati in formato METS, dal testo risultante dall'OCR), messo poi in distribuzione, debitamente criptato, alle biblioteche attraverso un'apposita interfaccia. Completate le procedure Google provvedeva a pubblicare le risorse sul suo portale, e i relativi link venivano poi, a cura delle biblioteche, collegati alla notizia bibliografica in SBN per consentire la fruizione della copia digitalizzata da parte degli utenti che accedono agli OPAC di Indice e locali.

Tornando al discorso iniziale, mi chiedo, ci chiediamo: ne è valsa la pena? Il bilancio numerico parla chiaro: più di 100.000 documenti digitalizzati e disponibili in rete, ca. 50.000 nuove catalogazioni in SBN, più un generale nostro lavoro di revisione catalogografica di ca. 200.000 volumi. Abbiamo scaricato sui nostri server le copie di tutto il materiale digitalizzato da Google e i relativi metadati. Attraverso queste copie le risorse, attualmente visualizzabili in Google Libri, saranno caricate nelle teche digitali delle singole biblioteche, e visualizzate anche in Internet Culturale e negli altri portali, in modo che i collegamenti presenti in SBN non punteranno più a Google Libri ma ai nostri depositi istituzionali. Inoltre una copia di sicurezza di tutti i documenti verrà caricata sui Magazzini digitali, il sistema per la conservazione permanente dei documenti elettronici pubblicati in Italia e diffusi tramite rete informatica, in attuazione della normativa sul deposito legale

Ma, ovviamente, non si tratta solo di un discorso quantitativo. Vorrei fare qualche considerazione sull'impatto del progetto nella nostra struttura organizzativa. Non possiamo nascondere che questa iniziativa ci è praticamente caduta tra capo e collo, provocando in alcu-

ni di noi entusiasmo, o almeno qualche curiosità, ma in molti altri dubbi e perplessità, sia sui contenuti dell'operazione – fare da serbatoio ad una società privata – sia soprattutto sulla fattibilità tecnica con le nostre sole forze. Detto fuori dai denti, in un periodo di grande difficoltà per le biblioteche a gestire anche solo i servizi essenziali, noi bibliotecari abbiamo sentito risuonare dall'alto lo slogan: «Armiamoci e partite!». Facendo però di necessità virtù, abbiamo sperimentato nuove metodologie di lavoro, come, ad esempio, l'automazione della gestione inventariale attraverso i codici a barre (certo, niente di straordinario che in altre latitudini non si faccia da molti anni). Ma, soprattutto, ci siamo misurati con l'organizzazione di un lavoro di squadra pianificato nei minimi particolari, con diversi team che oltre ad avere funzioni specifiche (dal *project manager*, ai responsabili della catalogazione moderna e antica, della logistica, della conservazione, della gestione tecnologica, informatica e amministrativa, fino ai responsabili delle singole spedizioni) avevano la necessità di attenersi ad un cronoprogramma preciso e vincolante, ad un flusso di lavoro in cui un minimo intoppo in qualche settore (ritardi nella catalogazione, negli adeguamenti tecnici delle procedure, nella movimentazione) si ripercuoteva su tutto il processo. Un lavoro di squadra caratterizzato anche da periodiche riunioni in audioconferenza, per fare il punto della situazione con le altre biblioteche, l'ICCU e i coordinatori americani. Da questo punto di vista – è il mio personale parere, non so se condiviso da tutti i colleghi coinvolti – l'esperienza è stata più che positiva.

Certo non mancano alcune criticità. Noi, ad esempio, siamo abituati a standard di digitalizzazione e di metadatozione sicuramente più accurati. Ma queste differenze sono la conseguenza naturale di un accordo tra due entità assai diverse. Da una parte i bibliotecari si propongono storicamente come proprio obiettivo quello di tutelare, descrivere e valorizzare il proprio patrimonio bibliografico co-

me risorsa culturale intorno alla quale organizzare una serie di servizi attraverso la messa in opera di strumenti e procedure adeguati. La biblioteca digitale non è altro che questo, la prosecuzione della nostra secolare *mission*, rimodellata nei nuovi contesti tecnologici (ed infatti io non parlerei più di biblioteca digitale, ma di biblioteca *tout court*, dove le risorse analogiche e digitali, locali e remote, convivono e interagiscono all'interno di un'unica articolata struttura). Ma a Google la biblioteca digitale, da questo punto di vista, non credo interessi più di tanto; e, forse mi sbaglio, non mi pare che da nessuna parte Google Libri si definisca come una *digital library*. Alla società californiana interessa valorizzare il proprio *core business* che è il servizio di ricerca, e quindi incrementare il bacino di parole che possono essere ricercate attraverso il proprio motore. Perciò a Google poco importa dell'accuratezza della metadazione e della descrizione bibliografica dei testi.

In conclusione mi sento di affermare che per noi bibliotecari il principale problema non è quello di cedere i nostri spazi d'azione a *competitors* privati, quanto piuttosto quello di affermare con forza una propria rinnovata identità. Stiamo celebrando il trentennale di SBN in un momento difficile e contraddittorio. Assistiamo ad una straordinaria accelerazione nello sviluppo di strumenti e metodologie, e già la novità della semplice presenza di collegamenti nei nostri cataloghi alle risorse digitalizzate (siano state prodotte da noi o da Google poco importa) sta riscuotendo, come era prevedibile, un notevole gradimento da parte degli utenti. SBN si sta indirizzando ver-

so nuovi modelli catalografici che, attraverso l'applicazione dei *linked data* e l'apertura agli orizzonti del web semantico, prospettano nuove funzioni alla descrizione catalografica, in cui le risorse, analogiche e digitali, si collegano strutturalmente alle altre risorse presenti nel web, nella direzione – obbligata se vogliamo sopravvivere – di un nuovo modello di biblioteca come agenzia per la gestione e distribuzione dell'informazione, in ogni formato si presenti. Nuovi modelli, orizzonti, prospettive, che ahimè si scontrano con l'attuale situazione economica del nostro Paese. Lo stato di asfissia nella gestione dei beni culturali, con tagli ai finanziamenti e la progressiva riduzione del personale, sta portando al collasso le biblioteche. In questo contesto la collaborazione con organizzazioni esterne non può, a mio avviso, che essere positiva, ma potrebbe purtroppo contribuire ad innescare quell'assurdo corto circuito per cui anche coloro che dovrebbero avere una certa contezza del ruolo delle biblioteche – parlo di persone che hanno responsabilità politiche e decisionali – pensano, e qualche volta dicono, e purtroppo decidono di conseguenza, che delle biblioteche non c'è più tanto bisogno perché c'è Google, perché è tutto su Internet, come se noi dovessimo rinunciare al nostro ruolo fondativo solo perché stanno cambiando gli strumenti di lavoro e i contesti in cui andremo ad operare. L'esperienza di Google dimostra invece che può essere proficuo, e forse addirittura necessario, lavorare insieme, ma ciascuno nei propri ambiti di competenze e secondo i propri scopi. Complementari ma non sostituibili né sovrapponibili.