La rete come biblioteca

Fabio Di Giammarco
digiammarc@tiscali.it

Una grande impresa che si misura con ostacoli di diversa natura

Con l'annuncio di "organizzare l'informazione del mondo per renderla accessibile a tutti", più di tre anni fa Google dava il via a una delle più grandi imprese tecnoculturali mai tentate: la digitalizzazione del sapere umano. Nel giro di qualche anno, prometteva l'azienda di Mountain View, i "primi" 16 milioni di libri, custoditi in alcune delle più importanti biblioteche del mondo, sarebbero diventati liberamente consultabili da qualsiasi computer collegato a Internet. Insomma, trasformare la rete in un'infinita biblioteca. Un'idea che al momento sembrò visionaria, ma che oggi lo è molto meno. E non solo perché le tecnologie necessarie (scanner superveloci e ad alta definizione, formati di visualizzazione, banda larga ecc.) sono diventate infrastrutture acquisite e in continuo perfezionamento, ma soprattutto perché ora risulta più evidente che si tratta di un'impresa la quale risponde a esigenze tecnologiche, culturali ed economiche sempre più pressanti, cioè trasformare il web immettendovi una tale massa critica di informazione qualificata da farlo emergere, attraverso il reperimento di informazioni selezionate mediante applicazioni avanzate tipo "booksearch", come strumento di conoscenza, di studio, di lavoro, ma anche come spazio per un nuovo modello di business basato sulla presentazione, la moltiplicazione e la distribuzione sia fisica che virtuale del contenuto "libro". In realtà, il grande interesse per la digitalizzazione dei patrimoni librari, la vera e propria svolta, ha inizio tra il 1999 e il 2000, e non riguarda soltanto Google. L'iniziativa di trasferire i testi dalla carta ai bit era in verità partita molto prima, rimanendo però sempre confinata – per mancanza a quel tempo di tecnologie in grado di innescare grandi prospettive – nell'ambiente del volontariato telematico. Un interesse che poi quasi all'improvviso si è trasformata in una sfida globale tra major di Internet, colossi dell'informatica e grandi organizzazioni non-profit. Con la banda larga, con lo sviluppo delle reti digitali e il conseguente aumento esponenziale degli utenti del web, comincia la grande caccia ad accaparrarsi la "polpa" futura di Internet: i contenuti. Tra questi ce n'è uno, fino a quel momento, per niente capitalizzato: i testi racchiusi nei milioni e milioni di libri "sepolti" nelle biblioteche di tutto il mondo. Da qui, il passo è breve: nasce l'idea della *mass digitization*, ovvero dell'acquisizione a tappeto dei magazzini librari in formato digitale.

I primi a gettarsi nell'impresa sono quelli del Million Book Project, nato dalla collaborazione tra le Carnegie Mellon University Libraries e l'Internet Archive. L'obiettivo è interessante: creare una biblioteca universale, se non altro dal punto di vista linguistico, con il coinvolgimento di biblioteche e istituzioni culturali asiatiche, in particolare cinesi e indiane. Poi, con l'ingresso in partnership anche della Bibliotheca Alexandrina, il decollo definitivo e la trasformazione in Universal Digital Library, con il risultato di più di un milione e mezzo di libri finora digitalizzati, appartenenti a circa venti lingue diverse. Lo *start* di Google arriva subito dopo: accordo con cinque grandi biblioteche americane e inglesi e 270 milioni di dollari come budget iniziale. La tecnologia messa in campo è di tutto rispetto: sistema di scanner superveloci da 15.000 libri a settimana. Ma l'aspetto ancora più interessante sono i due concetti di base che muovono il progetto: la conquista della leadership nella distribuzione e nel reperimento delle informazioni "culturalmente qualificate" in rete, e il perseguimento di un nuovo modello di business basato sui libri o meglio sulla promozione – mediante l'offerta dei testi on line – di nuovi titoli, autori e editori. L'iniziativa "pesante" di Google non poteva che scatenare reazioni. Si mobilita, infatti, a seguire una sorta di "Santa Alleanza" anti-Google: l'Open Content Alliance (OCA). Un'iniziativa con caratteristiche di eterogeneità, dove compaiono insieme l'Internet Archive, Yahoo!, Microsoft, una quarantina d'importanti istituzioni bibliotecarie, e come support-

er alcune major del calibro della software house Adobe Systems e della Xerox Corporation. Un massiccio dispiegamento che ha come scopo primario quello di esercitare una forza centripeta sia sui possessori dei contenuti (le biblioteche) sia sui potenziali fruitori, prospettando ai primi una *mass digitization* più attenta e di qualità, e ai secondi una biblioteca digitale più aperta, più amichevole, più innovativa e nello stesso tempo più professionale di quella enunciata da Google. A suggello di queste "buone pratiche", l'Internet Archive – *core* dell'OCA – lancia sul web la Open Library. Un prototipo proposto attraverso due siti: uno con funzioni di vetrina, dove sfogliare "virtualmente" i volumi già digitalizzati, l'altro invece per entrare nel progetto e testarne i requisiti d'apertura rispetto agli standard e di collaborazione con gli utenti nella gestione del catalogo on line, insomma un tentativo di rivoluzionare il profilo della biblioteca digitale con l'immissione di una robusta dose di "library 2.0". Nel frattempo, un altro colosso nel mondo delle *digital libraries*, la Library of Congress, non rimane certo con le mani in mano. E non solo per il continuo incremento dell'American Memory Project, archivio digitale di grande successo, nato nel lontanissimo 1990, con oltre 11 milioni di documenti storici messi finora on line, ma ora anche per l'avvio, dopo qualche anno di gestazione, di un progetto ambizioso: la World Digital Library. Accesso mondiale, libero e ubiquo: questo in estrema sintesi il programma pianificato insieme all'Unesco e con l'aiuto di Google, qui però nella sola veste di finanziatore e fornitore generico di supporto tecnologico. Progetto impostato sull'adesione di cinque grandi biblioteche e, come ulteriori prestatori di consulenza tecnica, anche su Apple e Intel; prevede, entro il 2009, la realizzazione di una biblioteca digitale multilingua e multiculturale, in grado di fornire tramite Internet un servizio per la distribuzione della conoscenza a livello mondiale.

Ma la *mass digitization*, differentemente da quello che gli inizi e i successivi sviluppi potevano far pensare, non rimane solo terreno di scontro tra major, mega-istituzioni bibliotecarie e grandi fondazioni non-profit a stelle a strisce. Infatti, dopo il famoso allarme lanciato dalla Bibliotèque nationale de France nel 2005 sui rischi di una digitalizzazione dei patrimoni librari lasciata ai soli Google ed epigoni, prende il via il progetto Biblioteca digitale europea (BDE). Nei piani della Commissione europea, la futura BDE, oltre a rappresentare una risposta alla colonizzazione anglofona della rete e alla commercializzazione della cultura, nel giro di pochi anni dovrebbe consentire la digitalizzazione e l'accesso on line a milioni di contenuti digitalizzati provenienti da biblioteche, archivi, musei ecc. dell'Unione, ponendosi come portale delle tradizioni culturali e scientifiche dell'intero continente europeo. Staremo a vedere, anche perché tra annunci così ambiziosi e concrete realizzazioni potrebbero mettersi di traverso vari fattori, a cominciare dalla burocrazia e ripensamenti vari di Bruxelles. In generale, rimane che la realizzazione di un grande progetto di biblioteca digitale si presenta – malgrado il superamento delle principali barriere tecnologiche – come un'impresa tuttora dagli esiti incerti. Nonostante gli annunci, il dispiegamento di mezzi, gli investimenti spesso cospicui, la ricerca e la messa a punto di nuovi modelli di business centrati sul "contenuto libro", la riuscita dei vari progetti, a dir la verità, non è al momento per niente assicurata. Pesano, infatti, problemi e ostacoli di vario genere: alcuni insiti nei progetti stessi, qualcuno fattosi immediatamente presente, e altri che magari cominciano appena a profilarsi all'orizzonte. Tutti però portatori degli stessi effetti: la possibilità di "sabotare" i piani con il rischio, alla peggio, di far fallire il progetto, o nel migliore dei casi, di snaturarne l'immagine o comprometterne gli obiettivi fondamentali. Tra questi, il copyright può essere considerato la "questione delle questioni". Dall'avvio della *mass digitization* si sta tentando di trovare una soluzione al fatto che i patrimoni delle biblioteche sono divisi tra libri di pubblico dominio e libri soggetti a diritti. I primi, pubblicati grosso modo prima del 1930, sono liberamente disponibili. I secondi invece, rispetto alle esigenze di copia digitale, totalmente indisponibili. Da qui, per questi ultimi, l'inizio di un'infinita ricerca nel tentativo di stabilire un nuovo equilibrio tra due esigenze a tutt'oggi inconciliabili: quella dei detentori dei diritti sempre più arroccati sulle loro posizioni e quella dei fautori del "diritto collettivo", dell'accesso libero e gratuito al sapere, raccolti sotto le insegne della "rivoluzione digitale". Google ad esempio, che ha puntato sul *fair use* e sull'opt-out, si è tirato addosso una bufera di critiche e denunce da parte di autori e editori. L'Open Content Alliance, che invece ha scelto per ora un'altra strategia, dichiara di voler trattare solo materiale "free", ovvero di pubblico domino. E nel frattempo, l'Europa studia. Per preparare la strada alla BDE, la Commissione ha deciso di riesaminare la normativa sul diritto d'autore, anche alla luce di una serie di consultazioni on line effettuate tra istituzioni culturali e detentori dei diritti, istituendo un gruppo di approfondimento ad hoc. Incombe poi – con il progressivo abbandono del supporto cartaceo – il problema "fragilità" del digitale. Rispetto alla secolare carta, i nuovi supporti vantano cicli di

durata preoccupanti: da qualche decina di anni per i supporti su disco o su nastro, fino alla breve ed effimera durata – quasi giornaliera – delle pagine web. Per non parlare del rischio di incompatibilità tra i tanti formati finora prodotti, da cui deriva una grande emergenza – al momento priva di soluzioni definitive – rispetto alla trasmissione e preservazione a lungo termine di tutta la memoria contenuta nei testi che velocemente stiamo convertendo in file digitali. E *last but not least*, la questione dell'interoperabilità tra le tante biblioteche digitali, vale a dire la mancanza di uniformità tra gli standard di codifica utilizzati nelle varie campagne di digitalizzazione. Google, ad esempio, nel classificare il proprio materiale sembra che utilizzi standard non proprio aperti e quindi forse potenzialmente incompatibili con quelli di altri sistemi. Tutto questo, in un prossimo futuro, potrebbe diventare un serio ostacolo all'integrazione degli archivi digitali e quindi alla libera individuazione, localizzazione e fruizione dei documenti attraverso Internet. In altre parole, un impedimento in grado di determinare una causa di fallimento o di forte ridimensionamento nel progetto di trasformare la rete in un'immensa biblioteca. Resta comunque il fatto che "l'attrazione fatale" libri, formato digitale e Internet, malgrado scetticismi, traversie e difficoltà varie, sembra irresistibile. "Organizzare l'informazione del mondo" è sempre stata una grande sfida per la cultura; oggi lo è anche per la tecnologia, per il diritto, per il business. Sul risultato finale, per ora non ci è dato di sapere. Di sicuro è che stiamo assistendo a una grande impresa i cui orizzonti sono però ancora tutti da definire.

Abstract

"To organize world information and to make it universally accessible." With this notice Google started off its project of digitization of human knowledge. Goal: make over the Net in an "infinite library". With broadband, developments in digital networks and great rise of web users, the "big hunt" in the Internet substance (the contents) has started. Apart from Google, there are many other big projects of mass digitizations: Million Book Project, Open Content Alliance, Internet Archive, World Digital Library (Library of Congress), European Digital Library. However, in spite of a great economic and technological effort, the results of this big challenge remain uncertain. The problems: copyright issue, digital frailty, risk of incompatibility and lack in uniformity in the code standards.