

Digital Books and the Impact on Libraries

Peter Brantley editor, "Library Trends", 57 (2008), 1, ISSN 0024-2594

In tempi molto recenti è esploso nel mondo digitale il tema degli e-books.

Complici alcuni progetti di digitalizzazione di massa come quello di grande impatto lanciato nel 2005 da Google, oggi conosciuto sotto il nome di Google Book Search, o quello di poco successivo dell'Open Content Alliance e grazie allo sviluppo vorticoso degli applicativi che vanno sotto il nome di Web 2.0, il libro digitale sta vivendo una fase di grande crescita sul mercato editoriale, scientifico e non, e di rinnovato e "partecipato" interesse da parte dei lettori. Non possono non tenerne conto editori e biblioteche che si interrogano sul futuro del libro e dei suoi *stakeholders* in questo bel volume monografico curato da Peter Brantley dal titolo *Digital books and the impact on libraries* dedicato dalla rivista "Library Trends" ai libri elettronici e al loro impatto sulla realtà delle biblioteche.

I sette contributi contenuti nel volume affrontano il tema degli e-books da un punto di vista molto più vicino alla realtà editoriale che non a quella delle biblioteche in verità, ma proprio questa peculiarità rende la lettura di questo volume particolarmente affascinante per chi, invece, lavora in biblioteca. Non che ci sia una grande differenza concettuale tra i temi cari agli editori e quelli cari ai bibliotecari. Talvolta però gli interessi divergenti e le ottiche contrastanti finiscono con il prevalere creando una barriera comunicativa tra i primi ed i secondi, laddove invece il digitale necessiterebbe di ampiezza di vedute e azioni sinergiche. Il parallelismo tra editori e biblioteche è condotto in modo severo da Michel Jensen nel suo contributo *Cultural tenacity within Libraries and Publishers*. Jensen critica in modo eguale editori e biblioteche per essere entrambi lontani dal mondo dei lettori. I primi, secondo Jensen, conoscono le dinamiche di mercato ma non i singoli consumatori; le seconde, parimenti, si focalizzano troppo sulle esigenze di alcuni gruppi di utenti, ma in virtù dei limiti imposti dalle normative sulla privacy e per il fatto di non essere soggetti alle leggi di mercato sono anch'essi lontani dalle reali esigenze dei singoli. Il digitale sta cambiando molto velocemente i ruoli e, secondo l'autore, per affrontare il cambiamento editori e bibliotecari dovranno imparare ad ascoltare i lettori più di quanto non abbiano mai fatto in passato. Le biblioteche prestando attenzione alle folksonomie, gli editori valutando attentamente ciò che i lettori scrivono nei blog personali o

su siti quali Amazon e LibraryThing.

Jason Epstein nel suo articolo *The end of the Gutenberg era* parte da un paragone tra l'era di Gutenberg e quella attuale. Anche nel presente, infatti, così come accadde nel Quattrocento, non è possibile prevedere quali saranno le conseguenze culturali, sociali ed antropologiche del passaggio sempre più spinto dal libro in formato cartaceo a quello in formato digitale. Ciò che sembra certo al momento è che il digitale ha creato nuovi mercati (la "lunga coda" di Anderson) per una pluralità di lingue (non solo quella inglese) e nuovi modelli di distribuzione – si pensi alla formula del *print on demand* – a costi contenuti rispetto alla carta. Non tutti i libri sono e saranno letti in futuro in formato elettronico, dato che in alcuni casi il libro cartaceo continuerà ad avere una portabilità superiore al libro elettronico così come, nonostante la tendenza ad una crescita della scrittura collaborativa, non tutti potranno improvvisarsi scrittori, ma non c'è ragione di credere che il mezzo digitale possa danneggiare il futuro del libro che invece potrà giovare di un mercato potenzialmente molto ampio, in quanto su scala mondiale. Il tema della digitalizzazione viene affrontato da Juliet Sutherland nel suo contributo *A mass digitization primer*. L'autrice ha una vasta e consolidata esperienza nel campo della digitalizzazione di testi e confronta tra loro alcuni progetti quali i ben noti Google Book Search e Open Content Alliance e altri meno noti quali il Posner Memorial Collection, il Cuneiform Digital Library Initiative o il progetto della biblioteca dell'Università di

Oxford Early Manuscripts at Oxford. Dall'esame comparativo dei differenti progetti di digitalizzazione emergono sostanzialmente due ordini di problemi per questo tipo di iniziative: quello relativo alla qualità delle immagini e quello del riconoscimento ottico dei caratteri (OCR).

Quanto al primo è evidente che la quantità di testi digitalizzati va sempre a scapito della qualità. L'autrice rivela di avere scoperto non pochi volumi in Google Book Search con pagine mancanti o, per altri versi, con pagine illeggibili, mentre loda ad esempio la qualità e l'alta definizione delle immagini nei manoscritti digitalizzati dalla Oxford Library.

Per ciò che riguarda il riconoscimento OCR va detto che, nonostante l'evoluzione tecnologica ed il grado di sofisticazione raggiunto dai lettori ottici, nessun OCR riesce a fare una lettura del testo priva di errori, soprattutto nel caso si debbano sottoporre a riconoscimento ottico i caratteri utilizzati nei libri antichi.

Per i testi moderni i principali problemi di riconoscimento sono da mettere in relazione con la lettura degli spazi bianchi e della punteggiatura.

La Sutherland discute la possibilità della correzione manuale dei caratteri. La scelta di ricorrere a questo tipo di correzione dipende sempre dalle risorse umane e finanziarie disponibili per ogni progetto di digitalizzazione, dai tempi e dalla quantità di materiale che si intende digitalizzare. La correzione manuale infatti è un'operazione decisamente lunga e costosa.

Una soluzione affascinante per risolvere il problema dei costi della correzione ma-



nale è quella ideata da Luis von Ahn della Carnegie Mellon University che ha sviluppato un servizio denominato ReCaptcha (<http://recaptcha.net/>) che utilizza l'intelligenza collettiva della rete per la correzione manuale su base volontaria dei testi digitalizzati nell'ambito del progetto Internet Archive.

È un monito agli editori il contributo di Sara Lloyd, dal titolo *A book publisher's manifesto for the twenty-first century: how traditional publishers can position themselves in the changing media flows of a networked era*. L'autrice che lavora per la casa editrice Pan Macmillan (UK) rivolge uno sguardo decisamente critico al mondo dell'editoria tradizionale che ancora non è riuscita a cogliere in pieno la portata del cambiamento che il mezzo digitale impone. Al con-

trario Google grazie al suo progetto di digitalizzazione di massa ha intuito che la partita del digitale va giocata intorno al contenuto e alla sua accessibilità immediata ed illimitata. Ciò vale in modo particolare per il libro elettronico la cui natura sta cambiando molto velocemente sotto la spinta di blog, wiki, forum, siti che consentono forme varie di auto-pubblicazione ecc, ovvero di strumenti che sfruttano la "sapienza collettiva" del Web 2.0.

Nella rete i confini del libro elettronico sfumano e diventano sempre meno stabili: i lettori leggono, annotano, commentano, discutono in forum. L'autore rivede il testo e pubblica la nuova versione di un'opera. Accade per esempio sul sito dell'Institute for the Future of the Book.

I lettori non sono più solo

passivi fruitori del contenuto, ma assumono un ruolo attivo nella creazione del testo. Da *consumers* diventano *producers* e *consumers* al tempo stesso, ovvero *prosumers*.

Quale può essere il ruolo degli editori in questo contesto vorticosamente *in fieri*? La Loyd propone agli editori di rompere con i vecchi schemi e di assumere un atteggiamento proattivo di fronte al cambiamento. L'autrice lancia l'idea di costruire piattaforme editoriali che siano in grado di caratterizzarsi in maniera forte sul mercato, proponendosi come polo di attrazione per autori e lettori, di concentrare insieme i contenuti di più editori, di valorizzare i testi con un sistema di linking citazionale (sul modello di ciò che accade per le riviste grazie a sistemi quali CrossRef e a piattaforme quali Ingenta Connect), di sfruttare i nuo-

vi media e gli strumenti del Web 2.0.

È necessario sviluppare accordi in merito ai DRM, accrescere le competenze, proporre modelli commerciali flessibili ed in linea con le diversificate esigenze del mercato.

La sfida per gli editori resta aperta.

Il contributo successivo, *The role of self-publishing in libraries*, scritto da Laura Dowson, commenta la crescita del fenomeno dell'auto-pubblicazione. Grazie a siti quali Lulu.com, iUniverse, Booksurge o, più semplicemente, attraverso i blog personali, la rete si popola di autori più o meno improvvisati. Quale deve essere – si chiede l'autrice – la reazione da parte delle biblioteche rispetto a questa ennesima estrinsecazione digitale dell'intelligenza collettiva?

Gran parte di questa mole di scrittura creativa è naturalmente da etichettare come “vanity press”, ma non tutto ciò che viene autopubblicato deve essere rifiutato in quanto qualitativamente scadente.

Alcuni volumi possono avere un interesse per la storia locale ed è per questo che negli Stati Uniti alcune *public libraries* cominciano a prestare attenzione anche a questo tipo di pubblicazioni, inserendole nelle proprie collezioni digitali. La valutazione va fatta, ovviamente, titolo per titolo. Differente è il discorso per le biblioteche accademiche che si rivolgono ad un'utenza di tipo specialistico e non sono interessate a volumi auto-pubblicati a meno che non vengano esplicitamente segnalati da docenti.

Si discostano dal tema principale i due ultimi contributi del volume.

Gli autori del primo contributo, *Digital text collections, linguistic research data and mashups: notes on legal situation*, sono ricercatori di linguistica computazionale e informatica legale in università tedesche e propongono una riflessione sul tema legale e della protezione del diritto di autore in ambiente digitale. In modo particolare gli autori partono dalla descrizione dei corpora linguistici per approfondire le problematiche poste dall'assemblaggio di dati di diversa tipologia e provenienza sul Web. È il caso dei *mashups* che per le loro caratteristiche intrinseche sono assimilabili ai corpora linguistici. I *mashups* sono diventati molto comuni in rete grazie agli strumenti del Web 2.0. Ne esistono di privati, creati per uso personale, oppure di com-

merciali. Alcuni servizi di pubblica utilità come il sito “Chicago crime map” (<http://chicago.everyblock.com/crime/>) mantenuto dalla polizia di Chicago, o il “Rentometer” (<http://www.rentometer.com/>) che visualizza il costo medio dei fitti negli Stati Uniti sono anche basati su *mashups* di dati. I creatori di *mashups* devono in primo luogo stipulare un contratto con il provider dei dati che intendono utilizzare. In svariati casi il provider esclude a priori la possibilità di un riutilizzo commerciale dei dati. È il caso ad esempio di Google Maps. I conflitti legali che riguardano queste “collezioni di dati remixati” sono comunque molteplici e strettamente legati alle differenti normative nazionali sul diritto di autore. In Europa due articoli che hanno un impatto sulla creazione di mashups sono l'art. 2 e l'art. 3 della Direttiva 2001/29/CE “Armonizzazione di taluni aspetti del diritto d'autore e dei diritti connessi nella società dell'informazione”. Eccezioni e limitazioni al diritto di autore sono anche contenute nell'articolo 5 della stessa direttiva.

Negli Stati Uniti, invece, limitazioni al copyright sono contenute negli articoli 108-122 del Copyright Act. Sempre nel Copyright Act, l'articolo 107 incorpora il principio del *fair use* che consente l'utilizzo di materiale protetto da copyright sulla base dei seguenti criteri:

- scopo e caratteristiche dell'utilizzo (commerciale o non);
- natura del materiale protetto da copyright;
- quantità di materiale utilizzato in proporzione all'estensione dell'opera considerata nella sua unitarietà.

In Germania il *fair use* è

regolato dal German Copyright Act.

Per altri aspetti i *mashups* possono rientrare sotto la normativa che protegge le basi di dati (per l'Europa la Direttiva 96/9/CE per gli Stati Uniti dall'art. 101 del Copyright Act) e sotto quella che protegge i marchi di fabbrica (*trademarks*), nel caso di riutilizzo di loghi, pagine Web o *screenshots* che contengano marchi.

I problemi legali che riguardano i corpora linguistici sono meno complessi rispetto a quelli dei *mashups* in quanto i corpora nascono ad esclusivo scopo di ricerca e generalmente sono risorse digitali molto stabili. Tuttavia la complessità di questo scenario in continua evoluzione impone un ripensamento della normativa sul diritto di autore e rende indispensabile lo sviluppo di protocolli aperti che consentano, agli utenti, ai creatori di *mashups* e ai data provider di regolare l'accesso ai loro contenuti. L'ultimo contributo, *Information discovery in ambiguous zones of research* di Suzana Sukovic, affronta il problema di come l'informazione scientifica viene organizzata e quindi fruita dagli studiosi in rete.

Biblioteche e editori spesso organizzano l'informazione in modo gerarchico come se ogni elemento fosse una “monade”, mentre il modello organizzativo della rete suggerisce un approccio rizomatico all'informazione che dovrebbe essere organizzata, così come viene fruita, in modo molto simile al sistema nervoso umano.

L'autrice pone particolare attenzione al caso delle scienze umane. Per queste discipline infatti l'integrazione di risorse, fonti, formati e media è particolarmente signifi-

cativa e va supportata in modo adeguato attraverso l'utilizzo di standard, protocolli e metadata per la descrizione e la gestione amministrativa dei documenti elettronici.

Usando come esempio la rappresentazione di un “limone”, la Sukovic si chiede come sia possibile costruire un sistema informativo coerente, aggiornato ma strutturato sulla base di un contenuto dalle rappresentazioni molteplici e mutevoli.

Non tutti i temi connessi con il libro elettronico vengono approfonditi nel volume curato da Peter Brantley che pone molti dubbi e offre poche risposte alle domande di biblioteche, editori, autori e lettori circa il futuro del libro.

Del resto, i comportamenti di scrittura e di lettura degli utenti che mutano vorticosamente e i progetti di digitalizzazione di massa avviati da Google o dall'OCA rendono vana qualsiasi plausibile previsione.

Per cercare delle risposte biblioteche, consorzi, editori e provider dovranno lavorare in sinergia, concentrandosi sulla moltiplicazione dell'offerta di contenuti digitali, semplificando le modalità di accesso agli stessi e rendendo pervasiva la disponibilità delle risorse con un occhio di riguardo per il trionfo della comunicazione mobile.

Ancora una volta è Google che ci mostra la strada da seguire con il recente accordo concluso ad ottobre 2008 con l'Associazione degli editori americani (Association of American Publishers) per la digitalizzazione di massa e la commercializzazione in rete di opere fuori commercio e protette da copyright.

Maria Cassella

Università degli studi di Torino
maria.cassella@unito.it