

IL LIBRO

Le biblioteche daranno i numeri: alcune riflessioni su *Library analytics and metrics* a cura di Ben Showers

di Chiara Consonni

In God we trust, all others must bring data.
(W. Edwards Deming)

In un mondo in cui sulle prime pagine di quotidiani e riviste convivono titoli accattivanti come *La grande occasione del digitale*¹, e facili allarmismi, *Viviamo in una scatola nera*², non ci si dovrebbe stupire se i bibliotecari impazzissero nel tentativo di decidere cosa fare dei numerosi dati in loro possesso.

Anche la comunità scientifica non è d'aiuto, divisa come è sulle norme che dovrebbero regolare la raccolta e l'utilizzo dei loro dati (pensiamo ad esempio al programma Personal genome project³); tantomeno lo è la normativa nazionale ed internazionale che non riesce a tenere il passo con i tempi nonostante gli sforzi recenti di cui diremo in seguito.

Eppure, se debitamente informati, i bibliotecari potrebbero imparare a sfruttare tutto questo a loro favore.

Il bilancio alla partenza è di grado zero: se da una parte le biblioteche, più o meno consapevolmente, possono contare su set di dati preziosi (sugli abitanti di una determinata area geografica, sui temi più ricercati in un determinato periodo, sui libri più richiesti ecc.), altri e più attivi sono i concorrenti sulla scena. Da anni infatti aziende di commercio elettronico come Amazon, progetti commerciali come Google libri e grandi gruppi edi-

CHIARA CONSONNI, Fondazione Biblioteca europea di informazione e cultura (BEIC), e-mail chiara.consonni@virgilio.it.

Ultima consultazione siti web: 30 novembre 2015.

1 Stefano Righi, *La grande occasione digitale: dai Big data al Big business*, «Corriere della Sera», 22 giugno 2015, p. 17, <http://archivistorico.corriere.it/2015/giugno/22/grande_occasione_digitale_dai_Big_ce_o_20150622_dc874218-18a1-11e5-a475-9b7c410a8d2d.shtml>.

2 Giorgio Fontana, *Viviamo in una scatola nera*, «Internazionale», 16 maggio 2015, <<http://www.internazionale.it/weekend/2015/05/16/internet-privacy-big-data-scatola-nera>>.

3 Sharing personal genome, <<http://www.personalgenomes.org/>>.



toriali come Elsevier, sfruttano i metadati bibliografici, per trarne profitto. Le biblioteche non devono certo puntare a questo, ma possono sempre imparare.

Il libro *Library analytics and metrics: using data to drive decisions and services* curato da Ben Shower⁴ e pubblicato da Facet nel 2014, cerca di fare il punto su questi temi e, soprattutto, di dare degli strumenti alla comunità bibliotecaria per riflettere sui numerosi aspetti legati alla raccolta dei dati ed al loro utilizzo etico.

Dopo una breve introduzione, il volume si articola in sei capitoli, ciascuno focalizzato su un particolare aspetto: *Big and small data*, *Data-driven collection*, *Library impact*, *Beyond numbers*, *Web and social media metrics*, *The risks of analytics*. Ogni capitolo è corredato da una ricca bibliografia e da alcuni casi studio di istituzioni inglesi, americane e australiane. Colpisce ad esempio l'iniziativa delle biblioteche coinvolte nel progetto Copac collection management tools (CCM tools)⁵ che hanno utilizzato i dati raccolti a livello centrale per giustificare lo scarto di parte dei loro materiali e per progettare la conservazione a lungo termine delle collezioni (case 2.2).

Ciò che distingue questa pubblicazione dalle numerose altre dedicate allo stesso tema è l'approccio pragmatico e sintetico che guida il lettore, identificato in primo luogo in direttori di biblioteche e personale coinvolto in progetti di valutazione, a confrontarsi con realtà affini e replicare esperienze di successo, sfruttando gli strumenti a disposizione.

Definizioni

La raccolta e analisi dei dati, in realtà, non sono una novità per le biblioteche che fin dall'inizio del secolo scorso hanno raccolto dati sui loro utenti e le loro ricerche per progettare servizi sempre più efficaci⁶. Le biblioteche sono inoltre state tra i primi istituti culturali ad affrontare le sfide del digitale, scontrandosi con le durezze del Web, ma cogliendone anche i benefici, consapevoli che la sempre più vasta disponibilità di contenuti digitali imponeva loro di essere presenti, pena l'esclusione dai flussi di ricerca dell'informazione. Quello che sembra invece mancare alle biblioteche di oggi è la capacità di adattarsi ai nuovi scenari e sfruttarne a pieno il potenziale.

Prima di proseguire, è importante chiarire cosa si intenda con *'analytics'* e *'metrics'*. Con quest'ultimo termine si indicano i criteri con cui i dati vengono quantificati e misurati, ovvero quali sono i parametri secondo i quali le informazioni vengono raccolte ed organizzate e secondo quale unità di misura. Mentre con *analytics* si intende la fase di analisi dei dati e la loro interpretazione, solitamente attraverso delle rappresentazioni grafiche come tabelle o grafici. L'utilizzo dei due anglicismi rimane d'obbligo, poiché non esiste un corrispettivo in italiano.

4 Attualmente *digital transformation manager* presso il Cabinet Office inglese. In passato ha lavorato per Jisc su progetti legati al futuro delle biblioteche, all'usabilità e alla digitalizzazione di materiali dopo essersi laureato presso la University College London nel 2008.

5 CCM tool è un progetto finanziato da Jisc dal 2011 che coinvolge più di novanta biblioteche inglesi e irlandesi nell'utilizzo di un sistema centralizzato di gestione delle biblioteche e nella presentazione delle registrazioni in un unico catalogo.

6 Ad esempio, già nel 1914 la American Library Association aveva preparato e distribuito un modulo cartaceo di raccolta dei dati per documentare e confrontare il possesso delle biblioteche pubbliche americane e stava procedendo a crearne di simili per le biblioteche speciali, come documentato nel report presentato alla Medical Library Association nel 1915 da James F. Ballard, bibliotecario presso la Boston Medical Library <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC234666/?page=6>>.

Le implicazioni legali ed etiche

Un altro aspetto di difficile traduzione nel contesto italiano è quello legato alle conseguenze legali ed etiche dell'utilizzo di dati, soprattutto di natura personale, in un contesto di servizio pubblico come quello delle biblioteche. Questo argomento è affrontato nel sesto capitolo, ma qui verrà trattato subito per sottolinearne l'importanza.

A prescindere dalla consapevolezza dei bibliotecari, le biblioteche sono depositarie di informazioni personali riguardanti i loro utenti, che solitamente comprendono dati di natura anagrafica, relativi ai prestiti e alle sessioni aperte nei computer disponibili in loco, fino ai video delle telecamere di sorveglianza.

I dati raccolti, se incrociati, possono portare alla creazione di dati sensibili, ovvero dati la cui raccolta e trattamento sono soggetti al consenso dell'interessato e all'autorizzazione preventiva del Garante per la protezione dei dati personali. Tra i dati sensibili, e quindi soggetti al d.lgs. 30/06/2003, n. 196 (*Codice in materia di protezione dei dati personali*) vengono infatti elencati l'origine razziale ed etnica, le convinzioni religiose, le opinioni politiche e la vita sessuale.

La stessa Comunità europea è da anni impegnata su questo fronte e dal 2017 dovrebbe entrare in vigore il *General data protection regulation* (GDPR)⁷ che andrà a sostituire l'attuale e ormai inadeguato⁸ *Data protection directive* (dir. CE 1995/46)⁹ datato 24 ottobre 1995.

Questi dati vengono solitamente forniti senza esitazione dagli utenti che vedono nelle biblioteche un ente pubblico che proteggerà quanto dichiarato. Dall'altra parte i bibliotecari raramente si preoccupano di questi aspetti, poiché raccolgono i dati direttamente dagli utenti e quindi con il loro consenso esplicito.

Come sottolineano i consulenti legali, autori del caso studio (6.1), le questioni sono in realtà più di natura etica che legale, infatti le biblioteche, facendo parte della pubblica amministrazione, seguono la normativa in vigore e sono solitamente protette in caso di ripercussioni legali.

Questo non deve però sollevare i bibliotecari dal preoccuparsi di questi aspetti; anzi, è proprio perché, come detto, la normativa è al momento inadeguata rispetto ai rapidi mutamenti del digitale, che deve subentrare l'etica professionale.

Anonimizzare le sessioni di ricerca sui computer installati in biblioteca, esporre ben in evidenza nei locali e sul sito secondo quali modalità i dati vengono raccolti, trattati e soprattutto protetti, sono esempi di buone pratiche di facile attuazione.

Se la raccolta dei dati non sembra presentare eccessive difficoltà, è il loro trattamento che ha spesso scoraggiato anche i bibliotecari armati delle migliori intenzioni. Shower al riguardo presenta però un caso interessante: cosa succederebbe se uno studente, dopo aver fallito un esame, accusasse l'istituto di formazione di non aver fatto tutto quello che era in suo potere per prevenirne il fallimento? Perché la biblioteca e l'istituto responsabile non hanno condotto studi sui comportamenti che pregiudicano il successo degli esami? Perché non sono intervenuti? (pag. 50).

⁷ *General data protection regulation*, «Wikipedia: the free encyclopedia», <https://en.wikipedia.org/wiki/General_Data_Protection_Regulation>.

⁸ European Commission, *Why do we need an EU data protection reform?* 2012, <http://ec.europa.eu/justice/data-protection/document/review2012/factsheets/1_en.pdf>.

⁹ *Data protection directive*, «Wikipedia: the free encyclopedia», <https://en.wikipedia.org/wiki/Data_Protection_Directive> .

Al di là del sofisma celato nell'esempio¹⁰, ha senso ricordare le conseguenze etiche e le possibili ricadute sul livello del servizio che la scelta di non raccogliere ed utilizzare i dati può comportare.

La raccomandazione è quella di iniziare a costruire un approccio etico condiviso tra i professionisti, avviando un dibattito pubblico e creando un contesto nel quale gli utenti siano messi in grado di prendere decisioni consapevoli e ragionate in merito ai loro dati.

Usare i dati

Ma per cosa possono essere utilizzati i dati? La risposta può essere in parte riassunta nel titolo dei due capitoli centrali nel volume: per dimostrare l'impatto e il valore della biblioteca e gestire al meglio le collezioni.

Come evidenziato in numerosi studi¹¹, le biblioteche sono ampiamente riconosciute come bene pubblico indispensabile per la comunità. Lo studio congiunto di Ipsos Mori, MLA e Shared Intelligence condotto nel 2010¹² dice che solo il 3% degli intervistati (target identificato in adulti residenti in Inghilterra senza ulteriori distinzioni) ritiene le biblioteche un servizio poco o affatto importante a fronte di un 35% che lo dichiara un servizio indispensabile.

A questa opinione diffusa non corrisponde però un pari utilizzo: più del 30% degli intervistati dichiara infatti di non aver visitato la biblioteca da più di un anno e il 15% si definisce un non utilizzatore.

Di fronte a questo quadro piuttosto contraddittorio e alle continue minacce di tagli alle risorse, diventa indispensabile quantificare il valore sociale e culturale delle biblioteche ed interrogarsi su quali siano effettivamente i servizi utilizzati. Facendo riferimento al precedente studio, risulta ad esempio chiaro che gli utenti delle biblioteche pubbliche inglesi nel 2010 apprezzavano la possibilità di prendere in prestito dei libri (76%) e di utilizzare i computer collegati ad internet (20%), mentre avrebbero gradito una revisione degli orari del servizio.

I tre casi studio presentati nel capitolo (3.1, 3.2, 3.3) sono relativi a biblioteche universitarie che hanno cercato di dimostrare l'esistenza di una correlazione positiva tra la frequentazione delle biblioteche e il successo accademico.

10 Di questo aspetto si è occupata anche Elena Maceviciute che nella sua recensione per *Information Research* (<<http://www.information.net/ir/reviews/rev535.html>>) si dichiara in profondo disaccordo con lo scenario presentato da Shower:

I would violently disagree that access to any amount of data and wonderful tools of analysis should take away the responsibility of those who are 'measured' and investigated [...]. The monitoring and understanding students, lecturers, researchers, managers and any organization as a whole does not imply that duties of those mentioned are disappearing or turn into some obscure rights to be saved from trouble, just because we can collect data about what they do and how they behave.

11 Ad esempio: Sara Chiessi, *What are libraries worth? A way to access the impact of Italian public libraries on users' lives and society*. Novembre 2011, <<http://www.ifla.org/files/assets/library-theory-and-research/Projects/researcher-librarian-report-chiessi.pdf>> e Anna Galluzzi, *Libraries and public perception: a comparative analysis of the European press*. Amsterdam: Elsevier, 2014.

12 *Attitudes to public libraries in England: topline results*. 15 ottobre 2010, <<https://www.ipsos-mori.com/Assets/Docs/MLA%20what%20do%20public%20want%20from%20libraries%20-%20topline%20-%20nov%202010.PDF>>.

I dati raccolti possono inoltre dare un valido aiuto nella gestione delle collezioni. Nel testo vengono citate strategie quali la *patron-driven acquisition*¹³, che prevede l'acquisto di nuovi titoli solo a seguito di un'esplicita richiesta da parte di un utente, e la *gamification*¹⁴, ovvero una simulazione in forma di gioco per coinvolgere gli utenti nelle attività dell'istituzione e avere un loro feedback¹⁵, oltre che l'adozione di *tool* per la condivisione dei dati, come nel caso della già citata esperienza di CCM tools (case 2.2).

Le possibilità non si limitano a questo, come dimostrano l'esperienza della Open university library (case 1.1), che ha sviluppato un *recommendation system* specializzato per i propri studenti, e quello della University of North Carolina (case 4.2), che dopo aver analizzato i dati di utilizzo degli spazi fisici della biblioteca (area relax, caffetteria, depositi ecc.), sta valutando una loro ristrutturazione.

Oltre i numeri

Proprio sull'aspetto della necessità, Shower sottolinea il rischio di porsi solo domande a cui sia facile rispondere senza interrogarsi sull'utilità della domanda stessa («street light effect»¹⁶), ad esempio se sia davvero sufficiente contare gli individui entrati nell'edificio per avere il numero totale degli utenti, o se invece non vadano conteggiati anche coloro che si sono connessi in rete per consultare il catalogo o navigare sul sito, e soprattutto se basti un numero alto di accessi per stabilire la qualità del servizio.

Un altro aspetto da ricordare è l'importanza di non limitarsi ai dati quantitativi, più facili da raccogliere ed analizzare, quando si cerca di dare risposta a domande complesse, come quelle relative alle abitudini degli utenti, ma di includere anche dati qualitativi, come quelli che è possibile raccogliere attraverso focus group ed interviste, per avere un quadro più preciso rispetto al contesto in cui il servizio viene erogato e alle motivazioni di una determinata scelta. Parafrasando Shower: «non vogliamo sapere solo cosa gli utenti fanno, ma anche capire quale sia stata la loro esperienza» (pag. 80).

Sempre nel tentativo di garantire una corretta interpretazione dei dati, Shower sottolinea la necessità di concentrare l'attenzione sugli *'small data'*, i dati locali. Negli ultimi anni infatti, la disponibilità di set di dati sempre più grandi, generati secondo standard sempre più interoperabili, i cosiddetti *'big data'*, ha distolto l'attenzione dalla dimensione locale, l'unica in grado di garantire un'analisi specializzata e non soggetta a semplificazioni, senza sminuire l'importanza di confrontarsi con altri soggetti. La raccolta e analisi dei dati è infatti un tema trasversale non solo agli istituti culturali (i cosiddetti MAB – musei, archivi e biblioteche), ma anche agli enti sociali, alle scuole e a tutte le attività economiche in generale.

Un altro aspetto trasversale e vitale nell'esistenza di ormai tutte le istituzioni è la presenza online. Presidiare la componente fisica non è infatti più sufficiente; è necessario assicurarsi una visibilità anche nel mondo digitale, attività questa più complessa per la dinamicità tipica della rete. Con l'avvento del Web 2.0¹⁷ la presenza di un sito

13 Rick Anderson, *What patron-driven acquisition (PDA) does and doesn't mean: an FAQ*, «The scholarly kitchen», 31 maggio 2011, <<http://scholarlykitchen.sspnet.org/2011/05/31/what-patron-driven-acquisition-pda-does-and-doesnt-mean-an-faq>>.

14 *Gamification*, «Wikipedia: l'enciclopedia libera», <<https://it.wikipedia.org/wiki/Gamification>>.

15 Un esempio è Librarygame, <<http://librarygame.co.uk/index.html>>.

16 *Effetto lampione*, «Wikipedia: l'enciclopedia libera», <https://it.wikipedia.org/wiki/Effetto_lampione>.

17 *Web 2.0*, «Wikipedia: l'enciclopedia libera», <https://it.wikipedia.org/wiki/Web_2.0>.

istituzionale non è infatti più sufficiente ed è invece indispensabile creare un insieme più variegato di identità sulle varie piattaforme, identità che devono essere tra loro coerenti e rispettare lo spirito originale dell'ente.

Anche il rapporto con l'utente remoto deve essere curato quanto quello del mondo fisico; studi su *community engagement*¹⁸, *brand management*¹⁹ e *Web presence*²⁰ sono diventati termini comuni nel lessico di chi se ne occupa²¹. Interessante ad esempio il caso studio del progetto Let's get real (5.2), coordinato da Culture24, un'organizzazione inglese che cura diverse piattaforme culturali online, con la partecipazione, tra gli altri, del Tate Museum e della British Library, che nella sua seconda fase si è concentrato sul concetto di *digital engagement* e su come questo si rifletta sulle attività istituzionali.

Fare rete e condividere

Due elementi di fondamentale importanza quando si parla di *data analysis* emergono solo indirettamente nel volume: la necessità di fare rete e la condivisione dei dati per creare un sistema nazionale ed internazionale di valutazione. Questo si dimostra particolarmente vero nel caso italiano dove, al di là di alcune sperimentazioni virtuose soprattutto in campo universitario²², non esiste ancora una politica professionale diffusa e condivisa.

Di fronte alle contingenze del periodo storico, ma soprattutto ai continui cambiamenti nel mondo del digitale, diventa di vitale importanza il coordinamento sia a livello istituzionale, per condividere risorse e *best practices*, ma anche a livello professionale, per creare una rete attiva di discussione e sviluppo.

Parallelamente alla condivisione delle forze e abilità è indispensabile mettere a disposizione di tutti i dati raccolti e parzialmente processati (meglio elaborati oppure trattati).

Senza dover citare il discusso caso di WikiLeaks²³, sono molte oggi le agenzie governative e le organizzazioni non a scopo di lucro che hanno creato delle piattaforme online per il recupero e il riutilizzo libero dei loro dati a beneficio di quanti vogliono sviluppare nuovi servizi o fare del *data-driven journalism*²⁴.

18 *Community engagement*, «Wikipedia: the free encyclopedia», <https://en.wikipedia.org/wiki/Community_engagement>.

19 *Brand management*, «Wikipedia: l'enciclopedia libera», <https://it.wikipedia.org/wiki/Brand_management>.

20 *Web presence*, «Wikipedia: the free encyclopedia», <https://en.wikipedia.org/wiki/Web_presence>.

21 In Italia se ne occupano ad esempio Marco Goldin (*In media ref...*, <<https://inmediaref.wordpress.com/>>) e Valeria Baudo (*Strategia di engagement e misurazione dei risultati in comunità di pratica online*, <http://www.academia.edu/4742268/BAUDO_Strategia_di_engagement_e_misurazione_dei_risultati_in_comunit%C3%A0_di_pratica_online>).

22 Un esempio è il Gruppo interuniversitario per il monitoraggio dei Sistemi bibliotecari di ateneo (GIM), costituitosi nel 2000 allo scopo di individuare metodologie comuni per misurare e valutare i servizi delle biblioteche universitarie italiane (<<http://www.gimsba.it/>>).

23 *What is Wikileaks?* 7 luglio 2011, <<https://wikileaks.org/About.html>>.

24 *Data-driven journalism*, «Wikipedia: the free encyclopedia». 20 luglio 2015, <https://en.wikipedia.org/wiki/Data-driven_journalism>.

Due esempi sono il progetto del governo del Regno Unito *Data.gov.uk*²⁵, avviato nel 2010 a seguito di una lunga campagna promossa, tra gli altri, dal *Guardian*²⁶, che chiedeva la pubblicazione dei dati raccolti secondo modalità aperte e gratuite. Ad oggi il portale dà accesso a più di 19.000 banche dati. Il secondo esempio è il portale *Crisis trends*²⁷ nel quale sono pubblicati i dati relativi agli oltre 10 milioni di sms ricevuti dall'organizzazione *Crisis Text Line*²⁸, il cui obiettivo è fornire una discreta consulenza ad adolescenti americani che sperimentano violenze domestiche, bullismo, disordini alimentari o simili problematiche. Questi dati, come la stessa CEO – Nancy Lublin – dichiara sono stati resi pubblici per sensibilizzare le amministrazioni pubbliche locali ed aiutarle a ridefinire le politiche giovanili e per la formazione del personale scolastico e giudiziario²⁹.

Quest'ultima osservazione ci porta poi ad un'altra importante riflessione, ovvero la necessità di includere questo campo di studi nei corsi di laurea e nel regolare aggiornamento professionale bibliotecario dato il ruolo che può giocare nel futuro delle biblioteche.

Per concludere, come dice Palfrey:

Questo periodo di transizione per le biblioteche sarà duro e costoso. Richiederà una gran dose di pensiero innovativo, di spiacevoli compromessi e cambiamenti in molti lavori e per molte persone. Ma pagherà degli ottimi dividendi alla società, se fatto bene, così come gli investimenti nelle biblioteche si sono dimostrati validi – e hanno ripagato profumatamente – per secoli³⁰.

Nel caso specifico dei *data analysis and metrics*, l'auspicio resta quello del titolo.

Articolo proposto il 1 dicembre 2015 e accettato il 15 dicembre 2015.

ABSTRACT

AIB studi, vol. 56 n. 1 (gennaio/aprile 2016), p. 103-110. DOI 10.2426/aibstudi-11415

CHIARA CONSONNI, Fondazione Biblioteca europea di informazione e cultura (BEIC), e-mail chiara.consonni@virgilio.it.

Le biblioteche daranno i numeri: alcune riflessioni su *Library analytics and metrics* a cura di Ben Showers

Oggi le biblioteche possono contare, più o meno consapevolmente, su una serie di dati preziosi relativi agli utenti e ai loro interessi. Tuttavia sono ancora troppo poche le istituzioni che utilizzano questi dati per migliorare i servizi offerti. Il libro *Library analytics and metrics: using data to drive decisions and*

25 *Data.gov.uk: opening up Government*, <<https://data.gov.uk/>>.

26 *Free our data: make taxpayers' data available to them*, <<http://www.freeourdata.org.uk/>>.

27 *Crisis trends*, <<http://www.crisistextline.org/trends/>>.

28 *Crisis text line*, <<http://www.crisistextline.org/>>.

29 Nancy Lublin, *The heartbreaking text that inspired a crisis help line*, «TEDWomen», maggio 2015, <http://www.ted.com/talks/nancy_lublin_the_heartbreaking_text_that_inspired_a_crisis_help_line>.

30 John Palfrey, *BiblioTech: why libraries matter more than ever in the age of Google*. New York: Basic Books, 2015, p. 38.

services cerca di fare il punto su questo tema, offrendo alla comunità bibliotecaria degli strumenti per riflettere sui numerosi aspetti legati alla raccolta dei dati e al loro utilizzo. In questa recensione vengono analizzati in particolare gli aspetti legali ed etici e le modalità di raccolta e utilizzo dei dati, facendo riferimento al contesto italiano. Viene inoltre sottolineata l'importanza di fare rete e di condividere i dati per creare un sistema nazionale e internazionale di valutazione.

Libraries at sixes and sevens: a few considerations on the book *Library analytics and metrics*, edited by Ben Showers

Although libraries can count on precious data sets covering users' personal data and habits, not very many are profiting from the potential of analytics to make informed decisions and ultimately drive new services. *Library analytics and metrics: using data to drive decisions and services* delves into some of the key issues in the use of data analytics and metrics, accompanying each chapter by short, practical case studies. This review focuses on the legal and ethical risks, and possible ways to uncover new insights and exploiting their use with an Italian perspective. The vital need of networking and sharing data is also stressed.