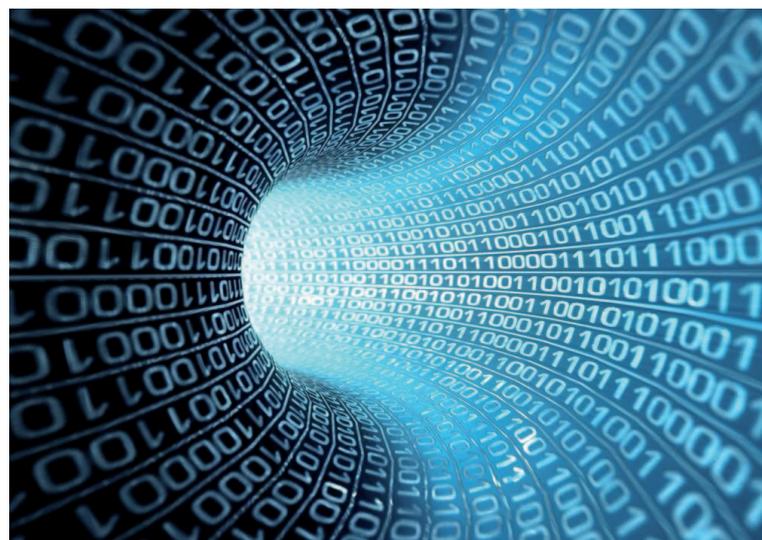


# Biblioteche e Big Data

FABIO DI GIAMMARCO  
digiammarco@tiscali.it

Lo “Sloan Digital Sky Survey”<sup>1</sup> è un progetto astronomico molto importante che prevede la mappatura dello spazio profondo. Finora sono stati catalogati circa 100 milioni di stelle, 1 milione di galassie e 100 mila quasar. Un’impresa titanica con conseguente produzione di un’enorme massa di dati. Per gestire la complessità di questi “Big Data” sono arrivate in aiuto le biblioteche della John Hopkins University di Baltimora.<sup>2</sup> Da una recente inchiesta risulta che i ricercatori di Oxford – solo nel 2012 – avrebbero generato almeno 3 petabyte (3 milioni di gigabyte) di dati. In pratica, il doppio delle capacità attuali del sistema centrale dell’università. Dell’organizzazione di questa mole di dati si stanno occupando le biblioteche della Bodleian<sup>3</sup> mediante implementazione di un sistema di storage (a pagamento) modulato in base alle diverse esigenze: da livelli completamente chiusi per le informazioni sensibili come quelle di carattere medico fino a *data sets* aperti con accesso pubblico e utilizzo di metadati. Iniziativa simile quella dell’UC Curation Center della California Digital Library<sup>4</sup> che ha cominciato anche lei a fornire servizi d’assistenza e supporto per l’intero ciclo dei dati. Secondo poi uno studio dell’Università del Tennessee su oltre 100 biblioteche universitarie, risulta che almeno il 40% è impegnato a sviluppare programmi per supportare gli scienziati nelle procedure di *Big Data curation*.

Da tutte queste iniziative appare chiaro come il mondo della ricerca trova difficoltà a gestire il “diluvio di dati” da esso stesso generato. Difficoltà ulteriormente confermate anche da un altro sondaggio che attesta proprio il difficile rapporto tra ricercatori e gestione dei dati: da un lato il desiderio maggioritario di accedere a un sistema facile di condivisione per l’utilizzo dei dati altrui, dall’altro l’ammissione d’incapacità (anch’essa maggioritaria) nel creare le condizioni minime di condivisione per permettere un accesso semplice ai propri. Insomma, sembra che con l’esplosione dei Big Data il mondo scientifico – in cerca di un supporto – chieda aiuto. Una richiesta che sta spingendo le biblioteche, in primo luogo quelle accademiche e di ricerca, verso un nuovo salto di paradigma: dalla *data curation* alla *Big Data curation*. Com’è capitato più volte in passato, un altro veloce e pervasivo ri-adattamento incalza.



Tanto è che la NISO (National Information Standards Organization)<sup>5</sup> si è affrettata a organizzare un convegno dal titolo “Libraries and Big Data”,<sup>6</sup> e nella presentazione ha messo in chiaro che il trattamento e soprattutto l’analisi di questa nuova megasfera informativa non può non competere alle biblioteche in quanto istituzioni specializzate riguardo l’intero ciclo dell’informazione, uniche competenti nella creazione di tassonomie, nel progettare schemi di metadati e organizzare sistemi di information retrieval ad hoc.

Il mondo bibliotecario, confidando nel suo prestigioso passato e nella sua recente digitalizzazione, aspira anche questa volta a fare quello che negli ultimi secoli ha sempre fatto: preservare e organizzare le informazioni per renderle universalmente disponibili alla ricerca e alle future generazioni. Ma la rincorsa tecnologica si fa sempre più difficile: un salto ad ostacoli con l’asticella posizionata via via sempre più in alto. Preservare, gestire e “far parlare” i Big Data sono le sfide che le biblioteche dovrebbero affrontare. Una prova tecnologicamente e concettualmente complessa: non si tratta solo di dotarsi di nuove e costose infrastrutture, ma soprattutto di ri-adattare il sistema biblioteca in funzione di un approdo inedito, vale a dire il passaggio dalla forma testo all’ambiente dati.

Intanto, le biblioteche soffrono di un taglio generalizzato delle risorse. Secondo l’ALA<sup>7</sup> negli USA hanno

perso negli ultimi anni centinaia di milioni di dollari di finanziamenti. Sempre per i tagli, in Gran Bretagna nel 2012 sono state chiuse più di duecento biblioteche.<sup>8</sup> E nel resto d'Europa, a causa dei noti programmi d'austerità fiscale, le cose non vanno di certo meglio. Da noi, il Ministero per i beni e le attività culturali ha visto in pochi anni dimezzato il proprio budget.<sup>9</sup> E poi, oltre le ristrettezze economiche, c'è anche da fare i conti con la concorrenza, vale a dire con il business della *Big Data curation* dove operano società di *data repository* aggressive e performanti che possono mettere in campo potenze di fuoco tipo GenBank.<sup>10</sup>

Inoltre, lo spostamento dalle collezioni di libri ai Big Data, non significa solo nuovi costi infrastrutturali, ma vuol dire progettare e supportare futuri servizi per l'accesso ai dati. Una sfida che si gioca nel trovare soluzioni rispetto a una serie di problematiche concernenti: "curation", archiviazione, ricerca, condivisione, trasferimento, visualizzazione e analisi dei dati. Ma, non basta. L'impatto dei Big Data promette di essere ancora più profondo. Infatti, una delle conseguenze dei necessari grossi impegni economici, tecnologici e operativi potrebbe essere il rivolgimento degli attuali assetti bibliotecari. La rivoluzione dell'"esplosione dei dati" non sembra proprio alla portata delle piccole biblioteche, e questo gap potrebbe significare il declino dei sistemi bibliotecari distribuiti con un ritorno – inaspettato – al modello basato sulla centralizzazione dei servizi.

C'è però anche chi – in linea con le indicazioni della NISO – vede nel "salto di paradigma" lo schiudersi di nuove opportunità per le biblioteche. Non soltanto professionalità dei bibliotecari nel classificare e presentare i Big Data ripuliti del loro contenuto "grezzo", ma anche l'utilizzo dei set di dati per migliorare le funzionalità stesse delle biblioteche. Ad esempio, sfruttando il "diluvio informativo" per il potenziamento e ottimizzazione delle politiche e performance interne: dalle decisioni sull'impiego degli spazi pubblici al monitoraggio dei materiali e strumenti di studio fino all'utilizzo dei "big data tools" per ampliare la propria offerta informativa. Nuove attività facilitate anche dal fatto che la gestione dei Big Data non è più cosa esclusiva delle grandi organizzazioni, ma possibile anche per strutture più piccole e questo grazie ai minori i costi dei server e *house data*, e alla disponibilità delle applicazioni open source.

Infine, un altro fattore incentivante, è la possibilità per le biblioteche di rivestire un ruolo importante nei progetti governativi sui Big Data". In particolare, riguardo le politiche avviate dall'amministrazione Obama,

James R. Jacobs, bibliotecario di Stanford<sup>11</sup>, sostiene che le biblioteche – a partire dagli standard di metadati e dalle strategie di conservazione delle informazioni – possono candidarsi a partner fondamentali nei processi di costruzione di sistemi pubblici basati sugli Open Data e Big Data. Un esempio in tal senso è quello dei *data sets* relativi alla ricerca finanziata dal governo federale. Dopo la decisione delle autorità di renderli pubblici e quindi completamente accessibili, William Michener – coordinatore e-science del sistema bibliotecario dell'Università di New Mexico<sup>12</sup> – ha fatto presente l'importanza di conservarli correttamente per poi renderli disponibili a tutta la comunità scientifica. Compito che secondo Michener solo le biblioteche saranno in grado di assolvere al meglio.

---

---

#### NOTE

<sup>1</sup> <<http://www.sdss.org/>>.

<sup>2</sup> <<http://www.jhu.edu/>>.

<sup>3</sup> <<http://www.bodleian.ox.ac.uk/bodley>>.

<sup>4</sup> <<http://www.cdlib.org/services/uc3/>>.

<sup>5</sup> La NISO è un'organizzazione statunitense senza scopo di lucro che sviluppa, mantiene e pubblica standard tecnici relativi ad applicazioni bibliografiche e bibliotecarie. È stata fondata nel 1939, costituita come associazione culturale non-profit nel 1983, ed ha assunto il suo attuale nome nel 1984 (da "Wikipedia").

<sup>6</sup> <[http://www.niso.org/news/events/2013/webinars/data\\_curation](http://www.niso.org/news/events/2013/webinars/data_curation)>.

<sup>7</sup> <[www.ala.org/](http://www.ala.org/)>.

<sup>8</sup> <<http://www.theguardian.com/books/2012/dec/10/uk-lost-200-libraries-2012>>.

<sup>9</sup> Passato dai 2,7 miliardi di euro del 2001 (lo 0,37 per cento del bilancio totale dello stato) a 1,5 miliardi previsti per il 2013 (appena lo 0,2 per cento del bilancio dello Stato).

<sup>10</sup> <<http://www.ncbi.nlm.nih.gov/genbank/>>.

<sup>11</sup> <<http://library.stanford.edu/>>.

<sup>12</sup> <<http://library.unm.edu/>>.

DOI: 10.3302/0392-8586-201310-007-1

#### ABSTRACT

The research community finds it difficult to manage the "data deluge" by itself generated. It seems that with the explosion of "Big Data", the scientific world asks for help, a request that is pushing libraries towards a new paradigm shift: from "data curation" to the "Big Data curation". Thus, the libraries can maintain their traditional commitment: to preserve information and to make it available.