

Assediati dai dati

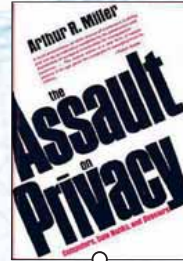
Una storia in continua accelerazione che riguarda anche la clinica e la sanità pubblica



THE SCHOLAR AND THE FUTURE OF THE RESEARCH LIBRARY

Fremont Rider, bibliotecario, ipotizza in un suo libro che il patrimonio delle biblioteche americane raddoppi ogni sedici anni.

1944



ASSALTO ALLA PRIVACY
L'accusa di Arthur Miller in un'opera che lascia il segno.

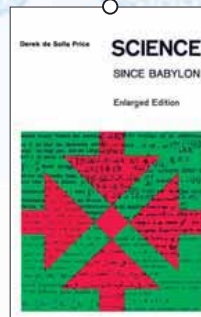
1971

1961

SCIENCE SINCE BABYLON

In questo libro, si tratteggia la crescita della conoscenza scientifica come un processo esponenziale e non lineare. L'autore è **Derek de Solla Price**.

Nell'articolo "Tracking the flow of information", uscito su *Science*, **Ithiel de Sola Pool** afferma che la disponibilità di parole aumenta dell'8,9% ogni anno.



Era il 1662 quando John Graunt pubblicò a Londra il "Natural and political observations mentioned in a following index, and made upon the bills of mortality". Secondo Stephen Mooney, Kimberly Westreich e Abdulrahman El-Sayed era la sintesi di un lavoro da demografo e statistico che anticipava alcuni tratti dell'epidemiologia moderna. L'editoriale "Epidemiology in the era of big data"¹, uscito sulla rivista *Epidemiology*, celebra dunque Graunt, mentre altre note recenti² indicano in John Snow l'antesignano del lavoro sui big data orientato alla salute.

Se è vero che nella locuzione "big data" il termine chiave è il secondo, c'è da dare ragione a chi vede i precursori in quanti hanno saputo lavorare con intelligenza sulle associazioni, muovendo dalle correlazioni alle ipotesi e alla determinazione delle cause. Ciò premesso, quella dei big data è storia recente, con una straordinaria progressiva accelerazione che rischia di rendere superata nel volgere di pochissimo tempo qualsiasi timeline. Volume, variabilità e velocità sono le caratteristiche fondanti dei big data, ma l'affermarsi di questa novità è stato talmente dirimpante che spesso si finisce con l'accettare per big data qualsiasi aggregazione di dati che sia caratterizzata anche solo da uno degli elementi che teoricamente dovrebbero concorrere insieme a definirli. In altre parole, di fronte a "tanti" dati o che giungono da fonti diverse o che ci raggiungono in

un rapido flusso costante crediamo comunque di essere alle prese con i big data.

Al contrario, l'elemento dirimpante è nell'essere i dati generati in modo spontaneo: come il "diluvio" descritto nelle pagine del dossier dell'*Economist* del 2009, la pioggia torrenziale di fatti e cifre non richiede una rilevazione attiva e si modifica l'osservazione tradizionale, affidandone l'analisi ad algoritmi costantemente affinati. È il caso, per esempio, dei dati ambientali, sul clima o sulla qualità dell'aria o delle informazioni che costantemente aggiornano Google Trends che non manca di ricordare che il suo obiettivo "è organizzare l'informazione planetaria

e renderla universalmente accessibile e utile".

L'impressione è di disorientamento: la previsione di Chris Anderson sull'obsolescenza del metodo scientifico conseguente alla resa allo tsunami di dati si traduce nello smarrimento di chi teme l'incapacità di governare un flusso di informazioni in crescita esponenziale. Ci si consola nella fantascienza della *internet of things*, con la speranza di un ecosistema addestrato a risolverci i problemi invece che a complicarci la vita: il frigorifero che ci avverte che il latte è finito, la caldaia che si avvia sollecitata dal meteo, l'automobile che ci porta a destinazione nel più breve tempo possibile informata dagli spostamenti



BIG DATA
Un documento della Nasa, l'agenzia spaziale statunitense, cita per la prima volta l'espressione big data.



Fondazione del primo grande portale di e-commerce.

QUANTA INFORMAZIONE?
Peter Lyman e Hal R. Varian della UC di Berkeley pubblicano **How much information?**: il primo studio che quantifica in termini di storage informatico il totale dei dati disponibili su carta, cd, dvd e supporti magnetici.

GENOMA
Completamento del primo draft del genoma umano.

INFLUWEB
In Belgio e Olanda viene lanciata la prima piattaforma web per il monitoraggio dell'influenza.

Tim O'Reilly nel libro **What is Web 2.0?** scrive: "The data is the next Intel inside".

Viene creato **Hadoop**, un framework open source per la raccolta di big data di Apache.

ORIGINI
In una conferenza, il chairman di Google, **Erick Schmidt**, afferma che ogni due giorni è prodotta una eguale quantità di dati di quella prodotta dalle origini della civiltà al 2003.

RIVOLUZIONE
Esce su *Wired* l'articolo di **Chris Anderson**: "The end of theory: the data deluge makes the scientific method obsolete".

NATURE
Numero speciale sui big data.

100 MILIONI UTENTI SU FB
Il social network di Mark Zuckerberg raggiunge una prima significativa milestone.

Special issue della **Harvard Business Review** dedicato a: "Big data: the management revolution".
La **NPR** dichiara "big data" espressione dell'anno.



Il governo Usa lancia la **Partnership for Resilience and Preparedness** che intende favorire l'accesso a una piattaforma open source per il controllo del cambiamento climatico.

BIG DATA EUROPE
Apre il sito dell'Unione europea.

1994 1997 2000 2003 2005 2008 2010 2012 2015 2016

1983 1996 1998 2001 2004 2006 2009 2011 2013 2014

EXPEDIA
Nasce il primo sito di turismo online.

LE TRE V
Esce il paper di Doug Laney del Meta Group che introduce i criteri chiave di **volume, variabilità e velocità**.

TWITTER
Apre il primo canale di microblogging online. L'ideatore è **Jack Dorsey**.
Decolla la *digital epidemiology* con **Healthmap**.



THE ECONOMIST
Pubblica il dossier sulla presenza ubiquitaria dei dati: **Data, data everywhere**.

DATA.GOV
Apre il portale degli open data del governo Usa.

TITOLI
Nella libreria di Amazon, il catalogo di libri sui big data supera le 250 unità.

BIG DATA. SEIZING OPPORTUNITIES, PRESERVING VALUES
Esce il report della Casa Bianca.

Chiude **Google flu trends**.



GOOGLE
All'università di Stanford, **Sergey Brin e Larry Page** costruiscono il motore di ricerca che cambierà il web.



FACEBOOK
Nasce la rete sociale di **Mark Zuckerberg**.

MC KINSEY
Esce il report **Big data: the next frontier for innovation, competition, and productivity**.

di chi ha lasciato traccia sullo stesso nostro percorso. In questo modo riportiamo i big data a un livello più rassicurante: quello della nostra spesa o dello scaldabagno. Spaventati dall'evidenza che i big data "si riferiscono a quello che si può fare su larga e non su piccola scala, che dà nuove suggestioni o crea nuove forme di valore, con modi che cambiano i mercati, le organizzazioni, le relazioni tra cittadini e governi"³.

Mettendo da parte la preoccupazione principale: l'evidenza che internet è lo spazio non governato più grande del mondo, il più grande esperimento di anarchia mai tentato nella storia⁴. F

1. Mooney SJ, Westreich DJ, El-Sayed AM. Epidemiology in the era of big data. *Epidemiology* 2015;26:390-4.
2. Khoury MJ, Ioannidis JP. Big data meets public health. *Science* 2014;346:1054-5.
3. Mayer-Schonhuber V, Cukier K. Big data: A revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Harcourt, 2013.
4. Schmidt E, Cohen J. The new digital age. New York: Knopf, 2013.