

Grandissimi numeri

di Luca De Biase

Nel 2000 il 25% delle informazioni registrate nel mondo era in formato digitale: il 75% era analogico, su carta, pellicola, plastica magnetizzata, e così via. Nel 2013 il 98% delle informazioni registrate nel mondo è in formato digitale: i vari supporti analogici, dalla carta alla plastica, si dividono il restante 2 per cento. Sono i calcoli di Martin Hilbert, che insegna alla Annenberg School for Communication and Journalism, della università della California del Sud. I dati digitali raddoppiano ogni tre anni e nella prima decade del nuovo millennio hanno dunque sorpassato i dati analogici che crescono molto lentamente a una velocità straordinaria, spiazzando abitudini, mentalità e business. Generando una trasformazione fondamentale per la convivenza e la conoscenza. Interpretare questa trasformazione è un compito talmente urgente che la ricerca in materia è incessante e ossessiva. Tentare di fermarla è futile.

Il cambiamento non riguarda soltanto il modo di comunicare, elaborare e immagazzinare le informazioni. Ma anche il modo di comprenderle. La ricerca sulla società umana era un tempo condotta sulla base di un insieme molto limitato di dati, sicché la bravura dei ricercatori stava nella capacità di estrarre il massimo di conoscenza possibile dalle statistiche, rispettando le fonti e approfondendo continuamente i metodi dell'analisi. Oggi i dati utilizzabili non sono più tanto scarsi. Ogni attività svolta online da due miliardi e mezzo di persone che usano internet lascia tracce digitali. E lo stesso avviene a chi va in macchina con il navigatore satellitare, chi paga con le carte di credito, chi segue la tv digitale o va in giro col telefonino, passeggia davanti a un portone difeso da una telecamera, e molto altro ancora. I dati che così si accumulano servono allo svolgimento di quelle attività, ma possono essere utilizzati per mille altri scopi sconosciuti. Lo sanno bene a Google e Facebook, lo sanno le compagnie di telecomunicazioni e le banche, lo sanno gli agenti dell'Nsa e quelli dei governi iraniano e cinese. Lo sanno centinaia di startup e centri di ricerca in tutto il mondo. Ma le conseguenze di questo cambiamento sono state appena intraviste. Di chiaro e sintetico, questo fenomeno ha per ora soprattutto un nome: Big data.

Viktor Mayer-Schönberger, che insegna all'Oxford Internet Institute, e Kenneth Cukier, giornalista dell'Economist, hanno raccolto con straordinaria capacità di sintesi un'enorme quantità di informazioni, un chiaro percorso di visioni e un buon numero di preoccupazioni in materia in un libro intitolato appunto "Big da-

ta. Una rivoluzione che trasformerà il nostro modo di vivere – e già minaccia la nostra libertà" (va detto che nel titolo originale – Big Data: A Revolution That Will Transform How We Live, Work, and Think – il tema della minaccia era molto meno evidente che nella traduzione italiana di Garzanti). Il libro sarà presentato da Viktor Mayer-Schönberger al Festival della Letteratura di Mantova il 5 settembre.

La conoscenza generata dai Big data, dicono gli autori, è caratterizzata dalla possibilità di lavorare non sui dati campionari ma sull'intero universo dei dati, allo scopo di trovare non relazioni di causa-effetto ma correlazioni probabilistiche, con la conseguenza di arrivare a vedere tutto ciò che avviene nel mondo digitale come una miniera di informazioni. Di valore immenso.

Utilizzando i Big data, sono nate startup che calcolano le probabilità di ritardo dell'aereo, valutano la convenienza delle offerte dei vari centri commerciali, indicano quali storie sono più interessanti sui giornali online e spingono i giornalisti ad approfondirle, correlare pubblicità e promozioni agli spostamenti delle persone: Decide.com, FlyOnTime.us, Prismatic e altri sono aziende che si sono sviluppate su questo genere di idee. Il MedStar Hospital Center di Washington ha analizzato tutti i dati relativi ad anni e anni di ricoveri per scoprire quali erano le condizioni che rendevano più probabile la reiterazione dei ricoveri. Microsoft, Google, Amazon, stanno diventando partner di una quantità di iniziative del genere.

Non è in nessun modo la fine della teoria e del metodo scientifico, come Chris Anderson ex direttore di Wired aveva pensato: perché le domande da porre ai dati, l'interpretazione delle correlazioni, il rispetto dei dati restano centrali nella generazione di conoscenza. La novità è che i dati possono parlare di più perché non sono solo campioni: ma esprimono il comportamento dell'intero universo preso in considerazione.

Il fatto è che le piattaforme proprietarie dei dati e coloro che sono autorizzati a utilizzarli hanno un potere immenso rispetto a chi è semplicemente un generatore di fatti registrati digitalmente. La privacy è messa in discussione. L'ipercontrollo è un rischio. Ma nel digitale ogni difetto è opportunità per creare qualcosa di meglio. All'Università di Pisa, Fosca Giannotti e Dino Pedreschi lavorano all'idea di Big data costruiti in modo da rispettare la privacy by design. La consapevolezza è il primo passo per il progresso.

Li chiamano Big Data: sfidano vecchie abitudini di ricerca, promettono conoscenze prima impensabili e offrono enormi opportunità

Il 98% dei dati è registrato in digitale: nel libro di Cukier e Schönberger, le conseguenze