

# Se l'Eldorado dei big data diventa una discarica di pregiudizi



di **Andrea Granelli**

**D**ov'è la Vita che abbiamo perduto vivendo? / Dov'è la saggezza che abbiamo perduto sapendo? / Dov'è la conoscenza che abbiamo perduto nell'informazione? (T. S. Eliot, cori dalla "Roccia"). In un mondo che cambia, l'informazione è la bussola per orientarsi: non cadere nelle trappole delle novità apparenti, capire ciò che sta tramontando davvero e individuare – prima degli altri – il nuovo, la sua forma, la sua direzione, i luoghi dove inizierà a manifestarsi. Il dato non è marginale ma tende – nell'era della conoscenza – a diventare un fattore produttivo competitivo. Inoltre i dati non sono solo numerici; ci sono anche, per esempio, le descrizioni dei comportamenti (dei clienti, dei concorrenti, ...), oppure le affermazioni – i verbatim (i "quote" per la cultura anglosassone) – fatte per esempio alla front-line di un'azienda o durante un'intervista etnografica. Anzi la dimensione qualitativa del dato sta assumendo una rilevanza crescente e il titolo suggestivo del celebre libro di Lindstrom – "Small data" – non solo coglie appieno questo aspetto ma lo contrappone dialetticamente al mondo vasto ma superficiale dei big data. E poi ci sono i segnali deboli, gli indizi che normalmente non hanno la dignità di dato (per la loro parzialità, incompletezza e talvolta la loro ambiguità) ma che sono sempre più importanti per anticipare il futuro, per prefigurare quello che sarà. I grandi investigatori, ma anche i grandi medici – pensiamo al dr House – sanno trasformare le tracce in conoscenza "obiettiva" (che consente loro di identificare con precisione un colpevole o diagnosticare una malattia). Ma da solo, il dato non diviene informazione. Per trasformare i dati grezzi in ciò che Edward Tufte – il padre delle rappresentazioni grafiche – chiama "differenza che fa la differenza", serve lo scarto informativo, quel quid in grado di orientare le decisioni, di modificare la conoscenza che abbiamo delle cose e farci agire trasformando. Non tutti i dati possono svolgere questo ruolo: c'è quindi sempre il rischio di riempirci di dati inutili, vuoti, che non ci dicono niente, che non ci servono a prendere decisioni, a risolvere emergenze, a imboccare nuovi percorsi. Vi è anche un altro rischio, forse più insidioso: come ha osservato Ronald Coase, premio Nobel per l'economia nel 1991, «se tortureremo a sufficienza i dati, confesseranno tutto». La grafica – con le sue capacità suggestive e il suo mettere in luce alcune cose e marginalizzarne delle altre

può diventare un pericolosissimo strumento di manipolazione, ancora più insidioso perché ha per definizione l'apparenza di un insieme di dati obiettivi e "tecnici". Il dato deve dunque aiutare a decidere, a capire meglio; ma dobbiamo ricordarci che l'uomo è un Giano bifronte: si orienta, decide, usando un misto di credenze e di fatti oggettivi, usa processi razionali ma si lascia contaminare dalle emozioni e dai pre-giudizi. La decisione viene orientata da un misto di dati da correlare e credenze che vengono organizzate in un ragionamento, più o meno esplicito. L'argomentazione è dunque un processo ibrido dove entrano numeri "obiettivi", credenze, luoghi comuni, ragionamenti. Dobbiamo sempre tenere a mente questo aspetto per ricordarci che la qualità del dato è condizione necessaria ma non sufficiente per una buona decisione. Per trasformare il dato in informazione ci vuole dunque discernimento, competenza retorica – che consente di costruire ragionamenti persuasivi – e quell'abilità che Steve Jobs chiamava "connettere i puntini". Detto in altre parole ci vuole una vera e propria "cultura del dato". Bisogna dunque reintrodurre i metodi della retorica anche nel ragionamento fattuale (e scientifico), per contrastare gli eccessi della cultura dei numeri – che tende a scambiare l'apparente obiettività con la rinuncia alla comprensione. In un famoso articolo di Wired del 2008 – chiamato emblematicamente "La fine delle teorie", l'allora direttore Chris Anderson osservava: «gli scienziati hanno sempre contato su ipotesi ed esprimimenti. [...] Di fronte alla disponibilità di enormi quantità di dati questo approccio - ipotesi, modello teorico e test - diventa obsoleto. [...] C'è ora una via migliore. I petabytes ci consentono di dire: "la correlazione è sufficiente". Possiamo smettere di cercare modelli teorici. Possiamo analizzare i dati senza alcuna ipotesi su cosa questi possano mostrare. Possiamo inviare i numeri nel più grande insieme di computer [cluster] che il mondo abbia mai visto e lasciare che algoritmi statistici trovino modelli [statistici] dove la scienza non può. [...] La correlazione soppianta la causalità e le scienze possono avanzare addirittura senza modelli teorici coerenti, teorie unificate o una qualche tipo di spiegazione meccanicistica». Una sana cultura del dato deve quindi combattere due derive pericolose: l'inflazione del dato che premia la quantità rispetto alla qualità, che lo considera – banalizzandolo – il nuovo "oro nero" e la venerazione algoritmica che tende a dare al computer il potere assoluto di comprendere i dati e di decidere al posto nostro.