

Big Data e economia: una riflessione epistemologica

Categories : [Istituzioni e regole](#)

Tagged as : [Mario Cedrini](#), [Menabò n. 105](#)

Date : 2 Giugno 2019

Il “sexy job” di questi anni, avvertiva il *chief economist* di Google Hal R Varian nel 2009 in un’intervista per [McKinsey](#), sarebbe stato quello degli statistici, capaci – in un mondo di dati liberamente disponibili e ubiqui (traduciamo quasi letteralmente) – di comprendere e processare quei dati, estrarre valore da essi, visualizzare e comunicare i risultati più rilevanti della *data analysis*. Il ridotto orizzonte temporale scelto da Varian – dieci anni – accresce il valore della profezia: già nel 2015, il [World Economic Forum](#) annunciava, attendendosi stupore da parte dei lettori, che quello del *data scientist* era il “best job” secondo il criterio sempre più utilizzato del “work-life balance” – il giusto equilibrio di (soddisfazione tratta dal) lavoro e (godimento della) vita privata. L’intervista di Varian illustrava i motivi che rendono il posizionamento dell’analista dei dati in quella classifica tutt’altro che sorprendente: è innegabile che l’avvento dei Big Data abbia avuto un impatto a dir poco violento sulle strategie di marketing aziendali, e che possa rappresentare un’opportunità senza precedenti per accrescere l’efficacia delle politiche pubbliche. Una rivoluzione, si è soliti osservare, ricordando il motto “volume, varietà e velocità”: le analisi Big Data (e dunque la disponibilità dei dati, ma anche delle tecniche) innovano per dimensioni, ovviamente; per rapidità, nella generazione dei dati e nel loro essere sempre attuali; e perché consentono e richiedono di concentrarsi anche su dati non o semi-strutturati, con ciò mostrando un potenziale rivoluzionario di carattere epistemologico sul quale appare ragionevole porre l’attenzione.

E questo è sì sorprendente, perché tra le molte sfide dei Big Data quella che essi sembrano in grado di portare ai modi tradizionalmente utilizzati per generare conoscenza è al tempo stesso la più interessante e la meno (o meno generalmente) dibattuta. “Big Data, fine della teoria”, sentenziava dieci anni orsono [Chris Anderson](#), allora direttore di *Wired*. “Facts by themselves are silent”, asseriva Alfred Marshall nel libro primo dei *Principles of Economics* del 1890; ma i dati? Quelli ubiqui di Varian, i Big Data, non saranno davvero esaustivi in senso forte – permettendo cioè non solo l’analisi di interi sistemi o popolazioni, ma un’analisi senza ipotesi, nella quale la correlazione è ciò che conta, e teorie, modelli, e spiegazioni diventano inutili? Anderson provoca, ma non fa che radicalizzare, semplicemente rendendolo esplicito e cioè cosciente, il sogno ricorrente nelle notti degli scienziati, quello di una ricerca “senza teoria”. E il nuovo empirismo dei Big Data avvicina con tutta evidenza la mappa al territorio – tanto che sono spesso gli algoritmi stessi a far emergere i “punti di interesse”.

C’è un’aria di definitività, naturalmente, che colpisce per la sua capacità... di non colpire a sufficienza. Come quando il vincitore del premio della Banca di Svezia per le scienze economiche in memoria di Alfred Nobel nel 1995, Robert E. Lucas, dichiarò nel suo *Presidential address* del 2003 all’American Economic Association, con soli cinque anni di anticipo sul crollo di Lehman Brothers, che il problema cruciale della prevenzione delle depressioni era stato risolto. Grazie, evidentemente, a un’impostazione – quella delle aspettative razionali e del Dynamic Stochastic General Equilibrium – che ha abituato gli economisti “che contano” all’idea che la loro disciplina possa prescindere dalla teoria (e dal confronto tra teorie alternative) per concentrarsi su modelli empirici capaci di replicare le correlazioni tra i dati. Ma, in fondo, un seminario odierno su tematiche microeconomiche studiate adottando un approccio “ortodosso” alimenta spesso la sensazione che la tipica evocazione iniziale, che precede l’analisi econometrica, della letteratura di riferimento, sia solo la precisazione del contorno della nicchia nella quale ci si chiude, lasciando che l’analisi dei dati (replicabile? Con qualche dubbio) rintracci una regolarità valida per lo specifico database utilizzato, senza porsi troppi problemi sulle condizioni necessarie per la generalizzazione dei risultati.

Menabò di Etica ed Economia

Il fondamentale lavoro di Rob Kitchin (ad esempio in *Big Data and Society*, 2014) mostra tuttavia che l'epistemologia empirista dei Big Data è viziata da una serie di fallacie logiche e narrative: i dati non sono "dati", e anzi riflettono visioni del mondo incorporate nei database, che sono pur sempre campioni, e nelle tecniche utilizzate per analizzarli; il ragionamento scientifico non è scomparso, è semplicemente nascosto dall'algoritmo; i dati continuano a non parlare, e vengono continuamente dotati di senso da operazioni di interpretazione. Infine, un'epistemologia può eliminare la teoria esclusivamente proclamandosi a settica – perdendo essa stessa senso.

Ma allora come definire i tratti della rivoluzione dei Big Data? Indagando piuttosto quelli della "*Data-driven science*", una sorta di nuovo paradigma non paradigmatico, che trae forza dal ricombinare metodi induttivi, deduttivi e abduuttivi – le intuizioni, e le ipotesi stesse, sono sì generate dai dati, ma è la teoria (esistente) che guida il processo di ricerca e scoperta di nuova conoscenza, e dunque anche quello di generazione, gestione e analisi dei dati. E questi ultimi rispondono a una logica abduuttiva – l'induzione è dunque valida in modalità *context-specific*, e le intuizioni generate servono a generare nuove ipotesi, che poi saranno testate.

La *Data-driven science* appare quindi qui, e per sua natura, un approccio attento ai limiti dell'indagine e ai possibili abusi di un modello di ricerca scientifica "datocentrico". Che il Lakatos Award del 2018 sia stato assegnato a Sabina Leonelli (è suo il neologismo appena utilizzato) per un suo recente lavoro sulla *Data-Centric Biology* (2016) è anche un riconoscimento del fatto che i Big Data sono innanzitutto uno strumento per riflettere sull'epistemologia stessa. I lettori italiani del saggio *La ricerca scientifica nell'era dei Big Data* (2018) restano impressionati dalla quantità di questioni aperte che la costruzione dei database, i loro "viaggi", la loro conservazione, prima ancora che il loro utilizzo, secondo Leonelli, pongono agli scienziati; e sono indotti a soffermarsi sugli aspetti critici, piuttosto che su quelli positivi, dei Big Data – tutt'altro che esaustivi: i Big Data danno informazioni "su *molto poco* e in maniera che tende a precludere, o comunque a rendere difficoltoso, ogni tipo di opposizione costruttiva. L'idea che i Big Data racchiudano una rappresentazione completa della realtà è un'illusione che sta distruggendo lo spirito critico con cui i ricercatori affrontano l'analisi e l'interpretazione dei dati empirici", osserva Leonelli (pp. 63-64) opponendo alla visione "rappresentativa" dei dati una "relazionale", nella quale non sono i dati a rappresentare la realtà, ma i modelli, e cioè determinati modi di ordinare i dati stessi; modi dai quali gli scienziati derivano conoscenza, avendo posto i dati in relazione con "elementi cruciali per l'interpretazione" (p. 1111) quali il *quaesitum* dell'analisi e i suoi presupposti concettuali. Si noti che lo studio di Leonelli si conclude (coerentemente) con una difesa del pluralismo nella produzione della conoscenza. Noi aggiungiamo che alla normatività del carattere desiderabile di una pluralità di approcci si accompagna la natura descrittiva di un pluralismo che è per certi versi nei fatti, quando si consideri che alla pretesa di esaustività dei Big Data si oppone idealmente uno sviluppo delle scienze – sociali in particolare – fortemente segnato dalla specializzazione nella ricerca.

Un'osservazione, quest'ultima, che si potrebbe utilizzare per richiamare un utilizzo non banale (è per altro lo stesso punto di arrivo di Kitchin) delle novità che i Big Data possono promettere nel campo della conoscenza. Il cammino della scienza è in realtà un cammino di specializzazione; e una disciplina come l'economia – contrariamente all'immagine che tende a offrire di sé, incentrata sulla modellistica di equilibrio e sul formalismo – riflette questa peculiare modalità di progresso scientifico nella varietà di programmi di ricerca (più o meno) alternativi al *core* neoclassico attualmente compresenti in un *mainstream* sempre più grande e aperto al contributo delle altre discipline. La *data-driven science* può aiutare a mappare l'evoluzione storica della scienza economica, ri-tracciando le strade interrotte e quelle poco illuminate, indagando l'alternanza di periodi di monismo e pluralismo, e le radici dell'attuale frammentazione. Un approccio quantitativo alla storia dell'economia (si veda in merito una recente *issue* del *Journal of Economic Methodology*, 2018), che utilizzi le tecniche Big Data del tipo "distant reading" (Moretti, *La letteratura vista da lontano*, 2005; *Distant Reading*, 2013), e che dunque focalizzi l'attenzione non sull'unicità dei contributi più rilevanti ("close reading"), ma sul vasto *corpus* di lavori scientifici prodotti dagli economisti nel corso del tempo (un esempio è il *topic modeling*, una tecnica probabilistica di *text-*

Menabò di Etica ed Economia

mining fondata sul criterio della cooccorrenza delle parole nei testi), appare in grado, opportunamente guidata dal “tradizionale” approccio proprio della storia del pensiero economico, di far emergere la struttura per così dire latente e mutevole di una disciplina caratterizzata da spinte centrifughe inimmaginabili fino a tempi recenti. Emergerebbero così *pattern* ricorrenti (la forma visibile *oltre* i dati) illustrativi della dimensione sociale, e tacita, del lavoro degli economisti e ciò suggerirebbe che i Big Data, quando sono considerati non alternativi, ma complementari, alle prospettive tradizionali, possono produrre quel salto di qualità epistemologico che, invece, il riduzionismo di un approccio “datocentrico” trasformerebbe in un passo indietro.