



Big data: l'annuncio della «fine della teoria».
Presupposti epistemologici,
implicazioni socio-politiche e ricadute educative

Big data: the announcement of «the end of theory».
Epistemological premises,
socio-political implications and educational repercussions

Pietro Corazza

Università degli Studi di Bologna – pietro.corazza2@unibo.it

ABSTRACT

This article aims at making a contribution towards the comprehension of the Big Data phenomenon, starting from the analysis of an epistemological paradigm that is getting considerable dissemination in the recent years: the one which proclaims “The End of Theory” and the overcoming of the scientific method. The deep socio-political implications of this vision will be examined, then a reflection about the educational repercussions will be developed. Firstly will be considered the opportunities that Big Data may unfold in the field of educational research. Then some of the mainstream representations of Big Data will be analysed, in order to clarify their influences on the possible ways of conceiving and structuring Data Literacy.

Questo articolo intende apportare un contributo alla comprensione del fenomeno dei *Big Data*, a partire dall’analisi di una concezione epistemologica che sta conoscendo una diffusione significativa negli ultimi anni: quella che invoca “la fine della teoria” e il superamento del metodo scientifico. Verranno esaminate le profonde implicazioni socio-politiche di tale visione, per arrivare quindi a riflettere sulle ricadute in ambito educativo. Verranno dapprima considerate le opportunità che i Big Data potrebbero aprire nel campo della ricerca educativa. In seguito saranno analizzate alcune delle narrazioni più diffuse che circolano sui *Big Data*, per esplicitare quali influenze esercitano sui possibili modi di concepire e impostare la *Data Literacy*.

KEYWORDS

Big Data; Epistemological paradigms; Educational Research; Data Literacy; The End of Theory.

Big Data; Paradigmi epistemologici; Educational Research; Data Literacy; La fine della teoria.

Introduzione

Tra le tematiche percepite come maggiormente significative nella nostra società vi è indubbiamente quella riguardante le tecnologie digitali, tanto che è stato addirittura proposto di considerarla la caratteristica distintiva del nostro tipo di società: a questo alludono infatti le definizioni di “società dei dati” o “società dell’informazione” (per una disamina critica di questo concetto si veda Webster, 2014). La portata dei mutamenti introdotti dalle trasformazioni tecnologiche appare tale da coinvolgere ogni aspetto della vita umana, e il mondo dell’educazione non fa certo eccezione: la sensazione è che mutamenti radicali siano già in atto e siano destinati a provocare conseguenze sempre più rilevanti in futuro (Margiotta 2019). E tuttavia l’intensità delle aspettative e delle preoccupazioni rispetto alle trasformazioni in atto spesso si accompagna ad una diffusa mancanza di chiarezza in merito ai fenomeni di cui si parla, rendendo difficile orientarsi tra le varie narrazioni che circolano. Questo articolo intende apportare un contributo alla comprensione del fenomeno dei *Big Data*, come auspicato da Raffaghelli (2017, p. 322), a partire dall’analisi di una concezione epistemologica che sta conoscendo una diffusione significativa negli ultimi anni: quella che invoca “la fine della teoria” e il superamento del metodo scientifico. Verranno esaminate le profonde implicazioni socio-politiche di tale visione, per arrivare quindi a riflettere sulle ricadute in ambito educativo. Verranno dapprima considerate le opportunità che i *Big Data* potrebbero aprire nel campo della ricerca educativa. In seguito saranno analizzate alcune delle narrazioni più diffuse che circolano sui *Big Data*, per esplicitare quali influenze esercitino sui possibili modi di concepire e impostare la *Data Literacy*.

1. *Big Data*: un nuovo paradigma epistemologico?

Di fronte al concetto di *Big Data* possiamo innanzitutto osservare che si tratta non soltanto di un termine decisamente vago, come lo sono tutti i termini utilizzati per comprendere un ampio insieme di fenomeni, ma anche piuttosto povero nella sua capacità di descrivere ciò a cui si riferisce. Infatti la grandezza dei più recenti database, la loro ampiezza in termini di volume delle informazioni, non sembra essere né una novità dal punto di vista storico (infatti enormi raccolte di dati sono disponibili da molto tempo, come ad esempio il censo), né la caratteristica più significativa e distintiva di ciò che indichiamo come *Big Data*. Si tratta infatti di un concetto molto più ampio, che racchiude svariati aspetti al di là del volume. Seguendo la ricostruzione che Kitchin (2013) trae da un ampio esame della letteratura accademica disponibile, la nozione di *Big Data* chiama in causa le seguenti caratteristiche: l’ampiezza del volume (nell’ordine di terabyte o petbyte di dati); la velocità (i dati sono creati in tempo reale o quasi); la varietà (sono compresi dati strutturati e non strutturati); l’aver come obiettivo l’eshaustività (ovvero il tentativo di comprendere intere popolazioni o sistemi); la risoluzione dettagliata e l’identificazione unicamente indessicale; l’intrinseca relazionalità (i dati rimandano a diversi data set che possono essere collegati); la flessibilità (ai data set possono venire aggiunti facilmente altri ambiti e possono crescere rapidamente in dimensioni). Queste novità nel mondo della gestione dei dati hanno condotto gli studiosi a formulare diverse ipotesi interpretative riguardo ai cambiamenti in atto. Jim Gray ha sostenuto che i *Big Data* hanno portato all’avvento di un nuovo paradigma in ambito scientifico (Hey, Tansley, Tolle, 2009): questa concezione è condivisa da numerosi autori, nonostante le critiche che evidenziano i limiti della

nozione kuhniana di “paradigma” scientifico (Kitchin, 2014, p. 3). È opportuno prendere in considerazione l’idea dell’avvento di un nuovo paradigma, nonostante la problematicità di questo concetto, perché ci conduce a considerare la posta in gioco del dibattito epistemologico sui *Big Data*, che ruota attorno alla domanda: quale forma ha il nuovo paradigma scientifico? Qui si fronteggiano due scuole di pensiero: da un lato abbiamo i sostenitori della “fine della teoria”, secondo i quali i *Big Data* avrebbero ormai reso obsoleto il metodo scientifico; dall’altro invece coloro che sostengono che le novità apportate dai *Big Data* siano da inserire all’interno della tradizione che si fonda su tale metodo.

2. «La fine della teoria»: i dati parlano da soli

La concezione che proclama la “fine della teoria” ha suscitato grandi entusiasmi soprattutto nel mondo del business, ma anche in alcuni ambienti accademici, in particolare nel settore della scienza dei dati (*data science*). La sua formulazione più diffusa è quella di Chris Anderson, che, quando era capo redattore della rivista *Wired*, ha pubblicato un articolo intitolato «The End of Theory: The Data Deluge Makes the Scientific Method Obsolete» (Anderson, 2008). Anderson parte dalla constatazione, condivisa in tutti gli ambienti scientifici, che ogni modello è necessariamente imperfetto: «tutti i modelli sono sbagliati, ma alcuni sono utili» recita una massima coniata dallo statistico George Bow (Bow, 1987, p. 74). Secondo Anderson, oggi la prima metà di questa frase resta vera, ma non siamo più obbligati a condividere la seconda: abbiamo una tale abbondanza di dati che possiamo fare a meno dei modelli teorici. Questa affermazione mette in discussione un presupposto cardine del metodo scientifico: quello secondo cui l’esistenza di correlazione tra due variabili non implica che ci sia un rapporto causale tra esse (potrebbe trattarsi di una coincidenza), e dunque non è lecito trarre conclusioni se non si dispone di un modello che spieghi quali relazioni sussistono tra le due variabili. Secondo Anderson invece l’aumento quantitativo della mole di dati avrebbe prodotto anche un mutamento qualitativo nei metodi per analizzarli: non è più possibile creare modelli che consentano di comprendere i dati nella loro totalità, ma non è nemmeno più necessario, poiché «la correlazione è sufficiente», ha preso il posto della spiegazione e l’ha resa superflua. Questo vale non solo per le scienze naturali e matematiche, ma anche per quelle umane: «È finita per ogni teoria del comportamento umano, dalla linguistica alla sociologia. Dimenticate la tassonomia, l’ontologia e la psicologia. Chi lo sa perché le persone fanno quel che fanno? Il punto è che lo fanno, e noi possiamo misurarli con una fedeltà senza precedenti. Se si dispone di abbastanza dati, i numeri parlano da soli» (Anderson 2008). È significativo che gli esempi utilizzati per dimostrare l’efficacia di questo approccio provengano spesso dagli studi di marketing: esso infatti si dimostra molto efficace nell’individuare correlazioni tra gli acquisti, indicando che se una data persona ha apprezzato il libro X, probabilmente, date le sue preferenze passate e quelle di altri consumatori che hanno abitudini simili alle sue, sarà interessata anche al libro Y (Dyche, 2012). È evidente che questo comporta uno slittamento non indifferente negli obiettivi della conoscenza: «solitamente non sappiamo nulla delle cause, perché spesso non ci interessano [...] l’obiettivo è quello di predire, più che di capire il mondo [...] basta che funzioni; la predizione supera la spiegazione» (Siegel, 2013, p. 90).

Nonostante l’attrattiva che questa visione può suscitare, è possibile riscontrare in essa diverse fallacie. La pretesa che i *Big Data* siano in grado di cogliere e rap-

presentare esaustivamente un ambito di ricerca è irrealistica, perché non tiene conto del fatto che ogni campione di dati è modellato dal tipo di piattaforma usata e dall'ontologia adottata per definire cosa è un dato, e dunque inevitabilmente è espressione di un particolare punto di vista e risente dei presupposti ad esso associati: non è possibile collocarsi in un punto di vista onnisciente e infallibile come "l'occhio di Dio", ogni raccolta dati è sempre situata (Kitchin, 2013; Crawford, 2013). La pretesa di fare a meno di teorie e modelli è invalidata dal fatto che anche i *Big Data* non sorgono dal nulla, ma durante ogni fase della loro costituzione entrano in gioco interpretazioni, giudizi e decisioni umane: si sceglie quali fenomeni considerare come "dato" e quali no, quali tra i dati raccolti sono importanti da analizzare e quali no, quali vengono mostrati e quali no (Andrejevic, 2013; Boyd, Crawford, 2012). Come afferma Vis (2013): «Raccontiamo storie attraverso i dati ed essenzialmente si tratta delle storie che abbiamo intenzione di raccontare». L'idea che i dati siano in grado di «parlare da soli», senza il bisogno di un'interpretazione umana, è fuorviante: in assenza di una cornice interpretativa non sarebbe possibile attribuire alcun significato ai dati, non riconoscere questo significa semplicemente evitare di esplicitare qual è la propria cornice di riferimento, nascondendola sotto una pretesa di "neutralità" o "oggettività".

3. Una scienza guidata dai dati

In opposizione alla visione secondo cui i *Big Data* avrebbero reso obsoleto il metodo scientifico, ne esiste un'altra che, pur riconoscendo la radicale portata trasformativa di questa novità tecnologica, la considera integrabile con il metodo scientifico: esso può andare incontro a delle trasformazioni senza dover rinunciare ai suoi fondamenti.

Consideriamo in primo luogo il campo delle scienze matematiche e naturali. In quest'ambito, i sostenitori di un tipo di scienza guidata dai dati (*data-drive science*) affermano che questa differisce da quella tradizionale per il fatto che le sue ipotesi «nascono dai dati», piuttosto che «dalla teoria» (Kelling, Hochachka, Fink, 2009, p. 613). Si tratta di utilizzare un procedimento che combina processi di induzione e deduzione. Alcune modalità induttive vengono incorporate nel processo di definizione del progetto di ricerca: una analisi preliminare dei dati aiuta a generare l'ipotesi, ma la ricerca non si ferma qui, e prosegue impiegando procedimenti deduttivi. A sua volta il processo di induzione non avviene in un vuoto di teoria, ma è situato nel contesto di quadri di riferimento in continua evoluzione. Diversi autori sostengono che questo modello di scienza guidata dai *Big Data* sia destinato a diventare il nuovo paradigma scientifico, perché è in grado di originare intuizioni che la scienza tradizionale, impostata su un modello teorico-deduttivo, non sarebbe in grado di generare: quest'ultima era particolarmente adatta ad un contesto dove i dati erano scarsi e la potenza di calcolo debole, ma ora il nuovo modello si rivela più efficace (Kelling et al., 2009; Loukides, 2010; Miller, 2010).

Nell'ambito delle scienze umane la situazione è diversa, perché esse sono sempre state caratterizzate da una pluralità di presupposti teorici differenti, il che impedisce di parlare di un paradigma scientifico unitario. Pertanto qui la domanda non può essere se i *Big Data* siano responsabili della comparsa di un nuovo paradigma, ma come le novità da essi apportate si inseriscano nel variegato panorama che caratterizza le scienze umane. Per poter riassumere le principali posizioni in merito, occorre in primo luogo distinguere tra la ricerca empirica e quella teoretica.

Nel campo della ricerca empirica, durante il Novecento si sono confrontate principalmente due prospettive: quella quantitativa e quella qualitativa. I *Big Data* si presentano come un'opportunità estremamente allettante per chi si occupa di ricerca quantitativa: le cosiddette "scienze sociali computazionali" (*computational social sciences*) sostengono di permettere il passaggio da «studi basati su dati scarsi ad altri basati sull'abbondanza di dati; da fotografie statiche a dispiegamenti dinamici; da aggregazioni grossolane a risoluzioni elevate; da modelli relativamente semplici a simulazioni maggiormente complesse e sofisticate» (Kitchin, 2014b, p. 3). Secondo i sostenitori di questo approccio, esso consentirebbe alla ricerca quantitativa di migliorarsi rispetto ad alcuni punti di debolezza che le sono stati tradizionalmente attribuiti: sarebbe possibile rispondere alle critiche di riduzionismo ed universalismo sviluppando rappresentazioni più variegata e sfumate dei fenomeni umani, e sottoponendole a verifica in molti contesti differenti (Kitchin, 2014a). I critici tuttavia ribadiscono la permanenza di carenze significative. Innanzitutto, come si è già visto in precedenza, individuare correlazioni non equivale a spiegarle: perciò, per una ricerca che non intenda aderire ai principi epistemologici espressi dai sostenitori della "fine della teoria", l'individuazione di correlazioni non costituisce l'esito finale ma piuttosto il punto di partenza dell'indagine, al quale dovranno presumibilmente seguire ulteriori raccolte di dati. Inoltre l'approccio alle scienze umane basato sui *Big Data* corre il rischio di ricadere in concezioni deterministiche del comportamento umano: ad esempio Alex Pentland, che è considerato uno degli scienziati informatici più influenti al mondo, ha spiegato come il suo utilizzo dei *Big Data* si fondi sul presupposto secondo cui «il tipo di persona che sei è determinato principalmente dal tuo contesto sociale, perciò se posso osservare alcuni dei tuoi comportamenti posso dedurre il resto semplicemente comparandoti con le persone che fanno parte della tua cerchia» (Pentland, 2012). Più in generale, a questo approccio possono essere rivolte le stesse critiche che sono state tradizionalmente mosse ai metodi quantitativi: tanto le società quanto le persone sono entità complesse e sfaccettate, che agiscono in modi anche imprevedibili e difficilmente riconducibili a logiche ben definite, e di conseguenza non possono sempre essere comprese attraverso leggi e correlazioni statistiche. Inoltre lo sguardo quantitativo non è in grado di cogliere tutto ciò che sfugge ai suoi parametri di misurazione numerica, trascurando così molti degli elementi a cui solitamente gli esseri umani attribuiscono importanza, come ad esempio le esperienze, le credenze, i significati, i dilemmi etici. Tuttavia, riconoscere i limiti dei metodi quantitativi non significa negare la loro utilità: possono fornire un importante complemento rispetto ai metodi qualitativi, che dal canto loro offrono analisi molto approfondite ma in grado di rappresentare solamente contesti ristretti. Alcuni studiosi hanno proposto infatti di sviluppare le scienze sociali computazionali, che come si è detto si basano sui *Big Data*, adottando ed estendendo i presupposti epistemologici utilizzati nell'ambito della statistica radicale e dell'approccio critico alla geografia informatizzata (GIS): si tratta di metodi che utilizzano le tecniche quantitative mantenendo la consapevolezza dei loro limiti epistemologici, servendosi di cornici interpretative provenienti dalla teoria critica, le quali ricordano come ogni indagine non sia mai neutrale bensì inevitabilmente condotta da un punto di vista situato, carico di valenze politiche (Kitchin 2014a).

Nel campo della ricerca teoretica, i *Big Data* si collocano all'interno della discussione attorno alle *Digital Humanities*. Si tratta di un ambito che è nato occupandosi della digitalizzazione di testi e materiali, oltre che della cura e dell'analisi dei dati digitali. Gli sviluppi successivi della disciplina hanno fatto emergere due distinte correnti di pensiero tra i sostenitori di questo approccio. La prima afferma

che le nuove tecniche quali conteggio, creazione di grafici, mappatura e lettura a distanza siano in grado di conferire un certo grado di oggettività e di rigore metodologico alle discipline umanistiche, che prima ne erano sprovviste (Moretti, 2005; Ramsay, 2010). La seconda tipologia di sostenitori crede invece che le nuove tecniche non debbano introdurre un approccio empirico o positivista nell'ambito umanistico, né rimpiazzare i tradizionali metodi di interpretazione e costruzione teorica, ma piuttosto fungere da supporto per la realizzazione di studi su vasta scala, che sarebbero stati impossibili senza l'ausilio delle tecniche computazionali (Berry, 2011). Oltre agli entusiasti, esistono però anche autori critici rispetto alle *Digital Humanities*: alcuni sostengono che le tecniche di lettura a distanza impediscano lo sviluppo dell'attenzione necessaria ad lettura approfondita (Culler, 2010); altri sottolineano come determinati artefatti culturali, come ad esempio le opere letterarie, non possano essere considerati alla stregua di semplici dati, poiché il loro significato non può essere compreso se sono scollegati dal contesto di appartenenza, e inoltre posseggono alcune qualità che sono ineffabili (Marche, 2012). Dunque le discipline umanistiche e le scienze sociali, quando si discute sulle modalità di integrazione dei *Big Data* all'interno delle proprie metodologie di ricerca, presentano argomentazioni piuttosto simili nei loro tratti fondamentali. La principale differenza risiede nel fatto che le *Digital Humanities* si servono delle indagini statistiche per scopi principalmente descrittivi, mentre le *computational social sciences* cercano anche di individuare rapporti causali, orientati a formulare spiegazioni e previsioni dei comportamenti umani: questo rende le loro pretese più delicate, e le critiche più pesanti.

4. Le implicazioni socio-politiche della “fine della teoria”

Nelle pagine precedenti è stata presa in considerazione la tesi della «fine della teoria», secondo la quale l'avvento dei *Big Data* avrebbe condannato il metodo scientifico al tramonto. Essa si basa su una componente descrittiva, in quanto presenta il funzionamento di alcuni procedimenti conoscitivi che si stanno attualmente sviluppando grazie ai *Big Data*, per trarne una visione normativa: eleva tali procedimenti a modello da imitare. In questo senso l'annuncio della «fine della teoria», più che una constatazione, si presenta come un programma da realizzare: delinea un progetto tutt'altro che concluso, che va portato a compimento. E tale progetto si estende ben oltre l'ambito epistemologico: presenta infatti alcune implicazioni socio-politiche di notevole rilevanza.

In primo luogo, occorre chiedersi che tipo di sapere potrebbe emergere dalla prospettiva auspicata da Anderson. Il metodo scientifico presuppone che il sapere sia, almeno in linea di principio, sottoponibile alla verifica e alla critica: a chiunque è permesso indagare i presupposti teorici e i metodi utilizzati, una volta impadronitisi delle competenze necessarie ad orientarsi entro le coordinate della disciplina presa in esame. Invece nella visione di Anderson i presupposti teorici non possono essere analizzati, perché egli afferma che sono scomparsi. E tuttavia, se adottiamo la prospettiva dei critici della «fine della teoria», possiamo affermare che i presupposti non sono affatto scomparsi (poiché come si è detto la raccolta e l'analisi dei dati implicano sempre prese di posizione), ma semplicemente vengono occultati. Questo occultamento può avere conseguenze pesanti: diversi studi testimoniano come i presupposti che modellano la gestione dei *Big Data* in diversi casi abbiano generato discriminazioni basate sul sesso, l'etnia e lo status socio-economico, penalizzando alcune categorie di persone rispetto alla possi-

bilità di trovare un lavoro, all'accesso al credito, alle offerte assicurative, all'equità nel trattamento penale (O'Neil, 2016; Eubanks, 2017). Quando i presupposti vengono nascosti sotto la retorica della «fine della teoria», viene meno la possibilità di esaminarli e sottoporli a critica: in questo modo diventa difficile rendersi conto delle discriminazioni, e ancor più attivarsi per contrastarle. Anche per quanto riguarda i metodi attraverso cui i dati vengono processati, la narrazione che si accompagna alla «fine della teoria» tende all'occultamento: gli algoritmi utilizzati per analizzare i dati vengono spesso descritti come “scatole nere”, alludendo al fatto che le procedure che consentono di rintracciare correlazioni all'interno di enormi moli di dati sarebbero troppo complesse per essere comprese dalla mente umana. In realtà, sebbene alcuni algoritmi siano effettivamente estremamente complessi, questo non è vero per tutti: spesso il concetto di “scatola nera” viene utilizzato strumentalmente per raggruppare sotto lo stesso termine sia i sistemi complessi che quelli semplici, al fine di mantenere opaco un algoritmo che in realtà sarebbe comprensibile, «perché la trasparenza metterebbe a rischio segreti industriali o svelerebbe le scelte discutibili dei proprietari del sistema» (Foryciarz, Leufer, Szymielewicz, 2020). E in ogni caso, anche quando gli algoritmi sono effettivamente complessi, è comunque possibile mettere in luce alcuni aspetti rilevanti del loro funzionamento:

Il punto chiave è che le spiegazioni delle decisioni automatizzate dovrebbero essere svincolate dalla capacità dell'opinione pubblica di capire come funzionano gli algoritmi. [...] Dobbiamo capire le scelte, le valutazioni e i compromessi fatti dalle persone che hanno progettato il sistema e che influenzano il comportamento dell'algoritmo. Per questo livello di comprensione non c'è bisogno di aprire la scatola nera tecnica (Foryciarz et al., 2020)

Dunque esiste la possibilità di richiedere trasparenza rispetto ai presupposti e ai metodi che orientano gli algoritmi. Ma se si diffonde una visione che nega l'esistenza di tali presupposti in nome della «fine della teoria», la non trasparenza degli algoritmi viene percepita come una caratteristica inevitabile. Così si profila una concezione del sapere che potremmo definire *oracolare*: il responso dell'algoritmo è insindacabile, perché le modalità attraverso le quali viene prodotto rimangono inaccessibili. Ciò appare in linea con la «mitologia» che secondo Boyd e Crawford aleggia attorno al concetto di *Big Data*, ovvero «la credenza diffusa che ampi insiemi di dati siano in grado di offrire una forma superiore di intelligenza e conoscenza, in grado di generare intuizioni che erano precedentemente impossibili, aventi un'aura di verità, oggettività e accuratezza» (Boyd, Crawford 2012, p. 663). Se i *Big Data* sono percepiti come capaci di offrire una conoscenza superiore a quella umana, finalmente oggettiva e accurata, può in fondo apparire ragionevole rinunciare a vagliare criticamente i presupposti e i metodi che vengono utilizzati: perché non dovremmo affidarci ciecamente a quello che dicono gli algoritmi?

A ciò va aggiunta una considerazione sul rapporto tra le discipline. Si può osservare come la figura del *data scientist* sia attualmente presentata come una delle professioni più richieste e incisive nel panorama lavorativo (Davenport, Patil, 2013). Tuttavia questa alta considerazione professionale, se accompagnata alla narrazione della «fine della teoria», tende a promuovere l'idea che essere in grado di “maneggiare i numeri”, ovvero di destreggiarsi con gli strumenti della statistica, sia sufficiente per poter interpretare i dati: questo rischia però di affidare l'interpretazione a soggetti che ignorano le conoscenze disciplinari proprie dell'ambito in cui si svolge la ricerca. Ad esempio si è assistito al fenomeno per cui un cospi-

cuo numero di fisici si è dedicato allo studio delle città, cercando di individuare presunte leggi in grado di descriverle: spesso le ricerche sono state impostate ignorando le teorie più significative che la scienza sociale ha prodotto nell'ultimo secolo, dando così luogo ad analisi che peccano di riduzionismo e funzionalismo, ricadendo in ingenuità positiviste che erano scomparse dagli studi sociologici da circa mezzo secolo (Bettencourt, Lobo, Helbing, 2007; Kitchin, 2014a). Un altro esempio proviene dal mondo delle scienze naturali, dove vi sono biologi che lamentano di come vengano affidate responsabilità di ricerca a bio-informatici che hanno una conoscenza ristretta e discutibile della biologia (Strasser, 2012). Assi-stiamo in questi casi a una dinamica che tende ad auto-alimentarsi: il paradigma della «fine della teoria» spinge ad affidare sempre maggiori responsabilità nella ricerca ai *data scientist*, a prescindere dalla loro conoscenza dei modelli teorici propri delle varie discipline, così facendo vengono portate avanti ricerche caratterizzate da una scarsa consapevolezza rispetto ai propri presupposti teorici, che contribuiscono a loro volta a promuovere e normalizzare un'attitudine che abita a trascurare la pratica riflessiva.

A questi elementi vanno infine aggiunte alcune considerazioni riguardo agli squilibri di potere che caratterizzano l'attuale gestione dei *Big Data*. Infatti, sebbene buona parte dell'entusiasmo associato al concetto di *Big Data* derivi dall'idea che è possibile avere accesso ad enormi quantità di dati, è necessario chiedersi: chi effettivamente ha accesso ai dati? La risposta è che «soltanto le aziende tecnologiche hanno accesso a quantità veramente ampie di dati» (boyd e Crawford 2012, p. 663). In contrasto rispetto alle speranze diffuse all'alba di internet, che vedevano nel Web un sistema incline a promuovere la decentralizzazione (sia da un punto di vista economico che di possibilità di espressione culturale), siamo attualmente costretti a constatare che l'attuale sistema economico favorisce la concentrazione di risorse, e quindi di capacità di influenza, nelle mani di poche grandi imprese tecnologiche (Hindman 2018). Ciò significa che ai ricercatori indipendenti è negata sia la possibilità di svolgere ricerche sui grandi database delle aziende, sia di vagliare gli studi che queste commissionano al proprio interno. In altre parole, non solo si profila la tendenza ad un modello di saper di tipo oracolare, ma coloro che possiedono e gestiscono gli oracoli sono imprese private. Se davvero la conoscenza passa e passerà sempre di più attraverso di *Big Data*, e questi non sono – attualmente – accessibili pubblicamente, ma piuttosto appartengono a privati, emerge una questione preoccupante circa gli obiettivi stessi a cui la conoscenza si indirizza: infatti è ragionevole supporre che le aziende tendano a perseguire i propri interessi, e di conseguenza ad orientare anche le azioni di ricerca verso questo fine.

5. Ricadute su ricerca educativa e Data Literacy

Le questioni sollevate sono rilevanti per il mondo della pedagogia da due punti di vista: da un lato si inseriscono nel dibattito sui paradigmi della ricerca educativa, dall'altro rientrano nel campo di riflessione della *Data Literacy*. Per quanto riguarda il primo ambito, attualmente queste questioni occupano ancora un ruolo piuttosto marginale nel dibattito sui paradigmi della ricerca educativa, tuttavia è prevedibile che in futuro acquisteranno una rilevanza sempre maggiore. In tal senso, una riflessione epistemologica approfondita rispetto alle novità che i *Big Data* prefigurano appare fondamentale. Questo contributo ha voluto porre l'attenzione su uno specifico paradigma, quello che annuncia «la fine della teoria»,

per evidenziarne da un lato le fragilità epistemologiche, dall'altro le preoccupanti implicazioni socio-politiche. È auspicabile che coloro che non condividono i presupposti e gli intenti di una simile visione siano consapevoli e preparati a contrastarne la diffusione, che negli Stati Uniti appare già avviata (si veda Prensky 2009). Con ciò non si vuole certo sostenere che sia necessario opporsi ad ogni tipo di innovazione metodologica, anzi: le opportunità di trasformazione, anche radicale, potrebbero rivelarsi feconde, a patto che sia mantenuta viva un'attitudine riflessiva e critica rispetto ai presupposti teorici e agli obiettivi, impliciti ed espliciti, dei metodi che si utilizzano. Le opportunità riguardano principalmente chi si occupa di ricerca quantitativa: in questo ambito una sfida consisterà nello sviluppare metodologie che siano in grado di sfruttare l'ampiezza e la precisione offerte dai *Big Data*, mantenendo tuttavia alta la consapevolezza dei limiti dell'approccio quantitativo alla vita umana, così come dell'inevitabile parzialità di ogni presupposto e di ogni criterio metodologico. In questo senso, potrebbe rivelarsi utile un confronto con le riflessioni e i metodi sviluppati nell'ambito della statistica radicale e dell'approccio critico alla geografia informatizzata, menzionati in precedenza.

Proponiamo ora una riflessione che si colloca nell'ambito della *Data Literacy*, nella convinzione che questa non possa orientarsi semplicemente alla trasmissione di competenze informatico-statistiche, ma debba includere anche una riflessione relativa alle implicazioni epistemologiche, etiche e sociali legate al mondo dei dati (Maybee, Zilinski 2015; Raffaghelli, 2017). All'interno di questa prospettiva, vorrei soffermarmi sull'importanza di una riflessione critica riguardante le narrazioni attraverso le quali vengono rappresentati i *Big Data*. Il presupposto su cui si fonda il mio discorso è il seguente: la nostra relazione con le tecnologie non è mai binaria, ovvero determinata unicamente dal rapporto tra l'individuo e il dispositivo tecnologico, ma è sempre triadica, cioè passa anche attraverso i discorsi e le pratiche attraverso le quali veniamo socializzati alle tecnologie stesse (Caronia 2002). In altre parole, è importante considerare le tecnologie non soltanto come cose con le quali interagiamo, ma anche come *parole*, ovvero come oggetto dei nostri discorsi: il modo in cui descriviamo le tecnologie ha un forte impatto sulle nostre modalità di rapportarci ad esse. Per questo motivo, risulta particolarmente importante avere consapevolezza delle narrazioni alle quali siamo esposti, e che a nostra volta veicoliamo e riproduciamo. Spesso infatti le espressioni che circolano nei discorsi riguardo alle tecnologie traggono origine, e a loro volta sostengono, visioni epistemologiche e socio-politiche. Perciò intendendo esaminare alcune narrazioni diffuse al fine di mettere in evidenza come esse siano espressione delle concezioni che sono state analizzate in precedenza.

In primo luogo consideriamo la tendenza a descrivere i dati qualcosa di intrinsecamente positivo e dotato di valore, da cui nasce l'idea generica che «più dati abbiamo, meglio è» (Lupton, 2015, p. 94). Questa si accompagna all'immagine dei dati come qualcosa di oggettivo, neutrale, affidabile. Spesso si sente parlare di «dati grezzi», sottintendendo con questo aggettivo che si tratta di dati «puri», non ancora contaminati da alcun tipo di interpretazione o punto di vista parziale; ma, come si è visto, anche prima che arrivino ad essere interpretati i dati non sono mai «puri», perché già le modalità della loro raccolta implicano prese di posizione: per questo alcuni autori hanno affermato che l'espressione «dati grezzi» è in sé fuorviante, è «un ossimoro» (Gitelman, 2013). In opposizione a questo tipo di metafore, ci sono studiosi che hanno coniato termini volti ad esplicitare i limiti e le carenze che possono caratterizzare gli insiemi di dati. Si parla quindi di «dati sporchi» per indicare raccolte dati che sono incomplete o presentano errori, e necessitano quindi di essere «ripulite» per poter essere utilizzate (Waterman, Hendler

2013); oppure di “dati corrotti” per indicare dati che sono stati alterati in modi che non erano nelle intenzioni dei loro creatori (Boellstorff, 2013). Un’altro genere di metafora è quella che associa i dati alle idee di circolazione e movimento: Internet viene rappresentata come un’“autostrada” lungo la quale le informazioni sono finalmente libere di viaggiare senza restrizioni. Tale immagine coglie certamente qualcosa di vero, in quanto Internet è stata progettata come un sistema decentralizzato nel quale è praticamente impossibile bloccare o censurare un messaggio (Abbate, 2000). Tuttavia, questa raffigurazione non tiene conto degli sbarramenti, delle resistenze e delle zone di esclusione che pure caratterizzano la rete: come si è visto gran parte dei *Big Data* è posseduta da aziende private, e inoltre permangono significative disparità di accesso e di possibilità a vari livelli (Sartori digital divide). Secondo alcuni autori la retorica della libera circolazione risponderebbe più agli impliciti presupposti neoliberali che dominano buona parte del discorso sulle tecnologie, piuttosto che a una effettiva descrizione di come è organizzata la rete (Lupton, 2015, p. 109). Infine, un’ultima famiglia di metafore sui dati è quella legata all’area semantica dell’acqua: i *Big Data* sono spesso descritti usando i termini “flusso”, “corrente”, “fuoriuscita”, “fiume”, “mare”, “oceano”. In parte, queste immagini rimandano alla già citata idea della libera circolazione, richiamando anche l’immaginario legato alla “società liquida” introdotto da Bauman (2011). Ma soprattutto è interessante notare un’evoluzione nel tipo di metafore acquatiche usate in relazione alla rete: se nei primi anni di diffusione del Web si utilizzava spesso il termine “navigare in rete”, ora, nonostante questa espressione non sia del tutto scomparsa, siamo più spesso esposti a parole quali “inondazione”, “diluvio”, “tsunami” di dati. Il cambiamento è evidente: se l’idea di navigare rimandava al piacere della scoperta, alla libertà di esplorazione, alla capacità di orientarsi e di cavalcare le onde con destrezza, i termini più recenti trasmettono invece la sensazione che i dati siano una marea incontrollabile, troppo grande e pervasiva per essere gestita, che minaccia costantemente di travolgerci (Lupton, 2015, p. 107). Quest’ultima considerazione è particolarmente ricca di significato e di implicazioni pedagogiche. Infatti la metafora della navigazione è in linea con una pedagogia che invita ad esplorare l’ambiente digitale, sostenuta dalla fiducia che sia possibile sviluppare gli strumenti per orientarsi in sicurezza ed autonomia. Al contrario, le profezie di inondazione trasmettono l’idea che la rete sia un luogo minaccioso, che le persone comuni non saranno mai in grado di muoversi autonomamente al suo interno: non resta quindi che affidarsi agli algoritmi, alle magiche “scatole nere” che, possedendo un’intelligenza superiore alla nostra, sono in grado di districarsi abilmente dove non possiamo, per offrirci in modo semplice e veloce la risposta che stavamo cercando. Che questo avvenga al prezzo di affidare i propri dati, e di delegare la propria facoltà di giudizio, alle aziende che possiedono tali algoritmi, appare tutto sommato un prezzo accettabile da pagare: in fondo, se siamo di fronte a un’inondazione non abbiamo molte alternative, occorre aggrapparsi alle ancore che ci vengono offerte. Perciò non a caso l’espressione «diluvio di dati» compare nel titolo dell’articolo di Anderson: fa parte di un immaginario che è perfettamente funzionale alla visione che egli intende promuovere.

Conclusione

Si è visto come la tesi della «fine della teoria», più che una descrizione dell’effettivo funzionamento delle pratiche scientifiche attuali, rappresenti un progetto: rappresenta quello che i suoi sostenitori vorrebbero che diventasse il nuovo pa-

radigma conoscitivo. Chi lo promuove tende a presentarlo come un processo destinato alla realizzazione: se ne parla come se fosse il naturale sviluppo della conoscenza nell'epoca dei *Big Data*, l'unico disponibile per coloro che hanno il coraggio di confrontarsi con i mutamenti in atto, senza rimanere «aggrappati alle vecchie vie» (Anderson, 2008). Ma, come si è visto, non si tratta affatto dell'unica opzione possibile: esistono diverse strade che il metodo scientifico può imboccare per sfruttare le opportunità offerte dai *Big Data* senza dover necessariamente rinunciare ai suoi fondamenti. La questione si presenta quindi nei termini di un confronto tra diversi paradigmi epistemologici che tentano di affermarsi. Alcuni autori sostengono che nell'ambito delle scienze matematiche e naturali, se è vero che i *Big Data* tenderanno a mettere in discussione il paradigma scientifico tradizionale incentrato sulla teoria, l'esito più probabile è che il metodo si evolva in direzione delle *data-driven sciences* (mantenendo quindi i fondamenti del metodo scientifico), poiché il modello della «fine della teoria» appare epistemologicamente fragile (Kitchin, 2014a). Nell'ambito delle scienze umane la questione appare invece più aperta, in ragione del fatto che esiste un pluralismo di approcci non riconducibili ad un paradigma unitario. Ciò significa che anche la ricerca educativa sarà chiamata a decidere quale strada imboccare. L'auspicio è che sia in grado di sfruttare le opportunità offerte dai *Big Data*, mantenendo però un'attitudine riflessiva che eviti il rischio di ricadere in concezioni epistemologiche riduzioniste simili a quella promossa dai sostenitori della «fine della teoria».

Sono state analizzate le implicazioni socio-politiche del modello di conoscenza promosso dalla «fine della teoria»: esso tende a un tipo di sapere oracolare, che non può essere sottoposto al vaglio della critica perché presentato come superiore alle capacità mentali umane, ma viene accettato come autorevole perché ammantato di un'aura di oggettività e accuratezza. Tale progetto tende ad elevare le competenze informatico-statistiche al di sopra delle altre conoscenze teoriche, ritenute ormai superate, rischiando così di dissipare il sapere accumulato dalle altre discipline. A ciò si aggiunge il fatto che la proprietà dei *Big Data* è principalmente concentrata nelle mani delle grandi aziende tecnologiche, che si ritrovano così ad esercitare un potere notevole, rischiando di piegare la conoscenza prodotta dai dati in direzione dei propri obiettivi economici.

Si è osservato come la visione dei sostenitori della «fine della teoria» si esprima e si propaghi anche grazie ad alcune narrazioni di uso comune: «più dati abbiamo, meglio è», «i dati rappresentano in modo oggettivo e accurato la realtà», «siamo sommersi dai dati». In opposizione ad questa concezione, credo sia possibile e auspicabile promuovere un approccio diverso rispetto al mondo della rete. In linea con una pedagogia fondata sull'esercizio del pensiero critico e problematizzante (Bertin 1968), credo non ci si possa arrendere all'idea di rinunciare alla propria autonomia di pensiero: meglio quindi rappresentare la rete come un mare che, pur presentando un certo grado di complessità, può essere navigato ed esplorato, non come uno tsunami dal quale si può solamente venire travolti. Occorre inoltre rifiutare le rappresentazioni dei *Big Data* come depositari di conoscenza intrinsecamente neutrale ed affidabile, e degli algoritmi come «scatole nere» assolutamente incomprensibili: è invece possibile esercitare una vigilanza critica sulla loro struttura e sul loro funzionamento. Credo infine che sia fondamentale, di fronte alle narrazioni che descrivono determinati sviluppi tecnologici come se fossero naturali, inevitabili e inarrestabili, educare (ed educarsi) a coltivare incessantemente la capacità di immaginare mondi diversi. Paulo Freire chiama «situazioni-limite» quelle «che si presentano agli uomini come determinanti storiche, schiaccianti, di fronte alle quali non esisterebbe altra alternativa se non adat-

tarsi»: «nel momento in cui [gli uomini] le percepiscono non più come una frontiera tra l'essere e il nulla, ma tra l'essere e l'“essere di più”, si fanno sempre più critici nella loro azione», fino a scoprire, al di là delle situazioni-limite, «le possibilità ancora *inedite* di azione» (Freire, 2002, p. 94). Questo vale quando ci si confronta con le narrazioni come quelle della «fine della teoria», che mirano ad imporre la propria visione come se fosse l'unica possibile, ma anche per quelle analisi critiche che sono abili nel mettere in luce le ombre e le ingiustizie che caratterizzano il nostro mondo, ma le descrivono come talmente potenti e pervasive da apparire invincibili.

Riferimenti bibliografici

- Abbate, J. (2000). *Inventing the Internet*. Cambridge, MA: MIT Press.
- Anderson, C. (2008, June 23). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. <https://www.wired.com/2008/06/pb-theory/> (consultato il 13/02/2020)
- Andrejevic, M. (2013). *Infoglut: How too Much Information is Changing the Way We Think and Know*. New York: Routledge.
- Bauman, Z. (2011). *Modernità liquida*, Bari: Laterza.
- Berry, D. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine* 12. <http://www.culturemachine.net/index.php/cm/article/view/440/470> (consultato il 13/02/2020)
- Bertin, G. M. (1968). *Educazione alla ragione. Lezioni di pedagogia generale*. Roma: Armando.
- Bettencourt, L.M.A., Lobo, J. Helbing, D. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences* 104(17), 7301–7306.
- Boellstorff, T. (2013). Making big data, in theory, *First Monday*, 10. <http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750> (consultato il 13/02/2020)
- Box, G.E.P., Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces*. Hoboken: Wiley & Sons.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15 (5), 662-679.
- Caronia, L. (2002). *La socializzazione ai media. Contesti, interazioni, pratiche educative*. Milano: Guerini.
- Crawford, K. (2013, April 1). The hidden biases of big data, *Harvard Business Review Blog*. <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/> (consultato il 13/02/2020)
- Culler, J. (2010). The closeness of close reading, *ADE Bulletin*, 149, 20-25.
- Davenport, T. & Patil, D. (2013). Data scientist: the sexiest job of the 21st century. *Harvard Business Review Magazine*. <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century> (consultato il 13/02/2020)
- Dyche, J. (2012, November 20). Big data 'Eureka!' don't just happen. *Harvard Business Review Blog*. http://blogs.hbr.org/cs/2012/11/eureka_doesnt_just_happen.html (consultato il 13/02/2020)
- Eubanks, V. (2017). *Automating Inequality. How High-Tech Tools Profile, Police and Punish the Poor*. New York: St. Martin's Press.
- Foryciarz, A., Leufer, D., & Szymielewicz, K. (2020, January 17). Black-Boxed Politics: Opacity is a Choice in AI Systems, *Medium*. <https://medium.com/@szymielewicz/black-boxed-politics-cebc0d5a54ad> (consultato il 13/02/2020)
- Freire, P. (2002). *Pedagogia degli oppressi*. Torino: EGA.
- Gitelman, L. (ed.) (2013). *'Raw Data' is an Oxymoron*. Cambridge, MA: MIT Press.
- Hey, T., Tansley, S. & Tolle, K. (2009). Jim Grey on eScience: A transformed scientific method. In Hey, T., Tansley, S. & Tolle, K. (eds.), *The Fourth Paradigm: Data-Intensive Scientific Discovery* (pp. xvii–xxx). Redmond: Microsoft Research.

- Hindman, M. (2018). *The Internet Trap. How The Digital Economy Builds Monopolies and Undermines Democracy*. Princeton: Princeton University Press.
- Kelling, S., Hochachka, W., Fink, D., (2009). Data-intensive Science: A new paradigm for biodiversity studies, *BioScience*, 59(7), 613–620.
- Kitchin, R. (2013). Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3), 262–267.
- Kitchin, R. (2014a). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1, 1-12.
- Kitchin, R. (2014b). The real-time city? Big data and smart urbanism, *GeoJournal*, 79, 1-14.
- Loukides, M. (2010, June 2). What is data science? *O'Reilly Radar*. <http://radar.oreilly.com/2010/06/what-is-data-science.html> (consultato il 13/02/2020)
- Lupton, D. (2015). *Digital Sociology*. London and New York: Routledge.
- Marche, S. (2012, October 28). Literature is not data: Against digital humanities, *Los Angeles Review of Books*. <https://lareviewofbooks.org/article/literature-is-not-data-against-digital-humanities/> (consultato il 13/02/2020)
- Margiotta, U. (2019). Responsabilità pedagogica e ricerca educativa: intelligenza collaborativa, formazione dei talenti e tecnologie dell'artificiale. *Formazione e Insegnamento*, 1, 13-18.
- Maybee, C., & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. *Proceedings of the Association for Information Science and Technology*, 52(1), 1–4. <http://doi.org/10.1002/pr2.2015.1450520100108>
- Miller, H.J. (2010). The data avalanche is here. Shouldn't we be digging?, *Journal of Regional Science*, 50(1), 181–201.
- Moretti, F. (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.
- O'Neil, C. (2016). *Weapons of Math Destruction. How Big Data Increase Inequality and Threatens Democracy*. New York: Penguin.
- Pentland, A. (2012, August 30). Reinventing society in the wake of big data, *Edge*. https://www.edge.org/conversation/alex_sandy_pentland-reinventing-society-in-the-wake-of-big-data (consultato il 13/02/2020)
- Prensky, M. (2009, January 3). H. sapiens digital: From digital immigrants and digital natives to digital wisdom, *Innovate: Journal of Online Education*, 5(3). <https://nsuworks.nova.edu/innovate/vol5/iss3/1> (consultato il 13/02/2020)
- Raffaghelli, J.E. (2017). Alfabetizzare ai dati nella società dei big e open data: una sfida formativa. *Formazione e Insegnamento*, 3, 299-324.
- Ramsay, S. (2010). *Reading Machines: Towards an Algorithmic Criticism*. Champaign: University of Illinois Press.
- Siegel, E. (2013). *Predictive Analytics*. Hoboken: Wiley.
- Strasser, B.J. (2012). Data-driven sciences: From wonder cabinets to electronic databases, *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 85-87.
- Vis, F. (2013). A critical reflection on big data: considering APIs, researchers and tools as data makers. *First Monday*, 10. <http://firstmonday.org/ojs/index.php/fm/article/view/4878/3755> (consultato il 13/02/2020)
- Waterman, K. & Hendler, J. (2013). Getting the dirt on big data. *Big Data*, 1 (3), 137-140.
- Webster, F. (2014). *Theories of the Information Society*. London and New York: Routledge.