

Èthos tecnologico e ricerca sociale: datificazione, computazione e sociologia digitale

Biagio Aragona, Cristiano Felaco

Introduzione

Ne "L'immaginazione sociologica", Charles Wright Mills (1959) sosteneva che l'*èthos* tecnologico che stava assumendo la scienza moderna potesse diminuire il livello di stima che le veniva generalmente accordato, poiché ciò che sino ad allora era stato considerato scienza sociale, correva il rischio di essere rubricato semplicemente «come un insieme di frammenti confusi di realtà umana» (p. 47). Wright Mills considerava l'empirismo astratto una delle cause della crisi della sociologia, e vedeva in Paul Felix Lazarsfeld, e nel suo celebre saggio "The Language of Social Science" (1955), la principale fonte di questo decadimento. A suo avviso, il collega che lavorava nella sua stessa università (la Columbia University) confondeva i diversi oggetti di studio della scienza con l'insieme dei metodi che venivano impiegati per studiarli, e tra questi primeggiava l'inchiesta campionaria, con il suo bagaglio di tecniche di interrogazione e di procedure di campionamento e di analisi statistica.

In cinquanta anni, il metodo della ricerca sociale codificato nel volume di Lazarsfeld si è affermato come lo "standard" della ricerca sociale (Marradi, 2007), il modello principale a cui i sociologi quantitativi hanno fatto riferimento. Con il passare del tempo, il bagaglio di tecniche collegate all'inchiesta campionaria è cresciuto a dismisura, insieme all'uso della tecnologia, che è stata impiegata con sempre più intensità per rilevare (attraverso tecniche *computer assisted*), organizzare (in database relazionali e *on-line*) e analizzare (con i più sofisticati packages statistici) i dati di tipo survey. Ma proprio questa stessa dimensione tecnologica che ha caratterizzato il metodo lazarsfeldiano, per alcuni studiosi (Savage & Burrows, 2007), sembra che abbia avviato la "crisi della sociologia empirica". Questa crisi sarebbe stata generata dall'uso diffuso delle nuove tecnologie di comunicazione, trasporto e logistica, che vanno a creare incessantemente dati (*big data, big corpora, open data, linked data, etc.*), sia come effetto diretto e indiretto di attività di tipo burocratico e legislativo, sia come accumulo spontaneo di informazioni derivate dall'uso massiccio della rete internet come strumento di scambio, di relazioni sociali, di conoscenza. Il risultato è stato la disponibilità di innumerevoli fonti empiriche che, anche per il loro carattere massivo, sembrano offrire molte più informazioni che in passato riguardo ai principali oggetti di studio delle survey: comportamenti, atteggiamenti ed opinioni. La progressiva disponibilità del materiale che si accumula in internet, nei *social networks*, o che viene generato da transazioni elettroniche e da dispositivi collegati alla rete, appare quindi destinata a sostituire progressivamente le tradizionali basi empiriche quantitative delle scienze sociali. Ad esempio, le ricerche di mercato vengono oramai realizzate sulle comunità *on-line*, piuttosto che attraverso le inchieste campionarie, e *network* e *sentiment analysis* stanno prendendo il posto dei sondaggi elettorali tradizionali; per non parlare di quanto l'analisi documentale sia cambiata con l'avvento dei *social media* e dei *big corpora* (Amaturo & Aragona 2016).

Se Wright Mills avesse avuto modo di assistere a ciò che è avvenuto in questi ultimi anni nel campo della ricerca sociale, di certo, con l'ironia che contraddistingueva i suoi scritti, avrebbe

notato che l'empirismo che egli chiamava astratto stava per essere progressivamente sostituito da una forma di empirismo ancora più radicale, in cui la dimensione tecnologica non è solo predominante, ma addirittura imprescindibile, costitutiva.

1. Scienza sociale computazionale e big data

Nel 2017, ogni minuto sono state inviate in media 103 milioni di *e-mail* e 15 milioni di messaggi di testo, e effettuate tre milioni e seicentomila ricerche sui motori di ricerca (Domo, 2017). Le grandi catene di supermercati negli Stati Uniti d'America e nel Regno Unito arrivano a registrare fino ad un milione di transazioni all'ora. Internet rappresenta oggi la principale fonte di dati, che provengono sia da tutti i contenuti che vengono di volta in volta caricati sotto forma di testi, immagini, video, url, etc. su social networks e siti web, i cosiddetti *User Generated Contents* (Boccia Artieri 2015), sia dall'*internet of things*, *internet* delle cose (o, più propriamente, *internet* degli oggetti). Gli oggetti comunicano dati su sé stessi e accedono a dati di altri dispositivi. Tra gli esempi, i dati dei sensori, sia fissi, come ad esempio i sistemi di domotica (contacalorie, presenza, incendio, gas), i sensori medici (battito cardiaco, pressione, glicemia, etc.) o quelli per il monitoraggio di fenomeni esterni (meteo, inquinamento, traffico) sia mobili, come i sistemi di localizzazione montati sugli smartphone, oppure i sensori (di parcheggio, di frenata, etc.) installati sulle automobili. In aggiunta, ci sono le tracce che vengono lasciate nella rete internet ogniqualvolta si naviga su un sito o su una particolare pagina web. Un esempio sono i *log* (registri), file che registrano eventi o messaggi che accadono in un sistema operativo, in un software o in un'applicazione. Un ultimo insieme di dati viene prodotto dalle innumerevoli transazioni che avvengono tra dispositivi digitali collegati tra loro, come i dati che si creano ogni volta che effettuiamo una qualsiasi comunicazione (chiamata, sms, videomessaggio) sul nostro telefono cellulare o *smartphone*. Di questo tipo sono anche i dati delle transazioni effettuate attraverso carte di credito e di debito, o i dati generati dai lettori ottici di codici a barre utilizzati nei supermercati.

Espressioni come "data deluge" (The Economist, 2010), "data revolution" (Kitchin, 2014a), "datafication" (Espeland & Stevens, 2008) e "società *data intensive*" (Amaturo & Aragona, 2017) sono state impiegate per definire quanto l'uso diffuso delle nuove tecnologie nella comunicazione, nei trasporti e nella logistica abbia modificato il ruolo che i dati hanno nelle aziende, nelle amministrazioni e nella società in generale. Alcuni studiosi hanno sostenuto che i nuovi dati abbiano trasformato il modo in cui "viviamo, lavoriamo e pensiamo" (Mayer-Schönberger & Cukier, 2013) e, potremmo aggiungere, facciamo ricerca. I nuovi dati e gli sviluppi computazionali per la loro archiviazione, elaborazione e analisi hanno infatti avuto il doppio effetto, in primo luogo, di richiedere un aggiornamento complessivo del nostro apparato di tecniche di ricerca, e, in secondo luogo, di intervenire sulle questioni fondamentali della scienza sociale e sul futuro della nostra titolarità, in quanto sociologi, ad esprimere punti di vista sulla realtà sociale.

Sul versante delle tecniche, intese come «procedure usate in una data scienza» (Kaplan 1964, p. 19), il *data deluge* ha posto, e continua a porre, grandi sfide sia per la elaborazione e analisi, che per la presentazione dei risultati. In primo luogo, i big data richiedono maggiore

attenzione dei dati tradizionali ai processi di preparazione all'analisi, quelli in cui dati vengono selezionati, ridotti e organizzati per le analisi, per usare un termine molto in voga, la "*curation*" del dato. Quando i dati sono raccolti con tecniche di *web crawling* e *web scraping*, il lavoro di pre-analisi può essere molto faticoso e lungo. Inoltre, la stessa analisi di un numero elevato di records può essere eseguita solo attraverso algoritmi; per cui la difficoltà maggiore sta nello sviluppare processi automatici che possano apprendere dai dati e dalla loro analisi, il cosiddetto *machine learning* (Mitchell 1997). Questa tecnica viene utilizzata come forma di *data mining* e può avvalersi dell'elaborazione del linguaggio naturale, delle reti neurali, degli alberi decisionali, così come della statistica tradizionale. Infine, un altro aspetto tecnico fondamentale per l'analisi dei big data è la visualizzazione. I metodi visuali riescono a rappresentare e comunicare efficacemente la struttura, il modello e i flussi delle variabili e delle loro relazioni. La visualizzazione può essa stessa diventare una tecnica d'analisi, *visual analytics*, guidata da una combinazione di algoritmi e di riflessione teorica (Keim et al., 2010).

Oltre ad avere un impatto sulla "cassetta degli attrezzi" del sociologo, l'innovazione tecnologica dell'indagine sociale ha favorito anche la progressiva richiesta di titolarità a raccogliere e analizzare dati sociali da parte di nuovi studiosi, principalmente informatici e ingegneri, con scarse, o addirittura nulle, conoscenze sociologiche, ma con elevate capacità tecniche. Essi hanno rivendicato a voce sempre più alta il diritto di analizzare la realtà sociale e di trarre conclusioni a partire da queste immense basi dati. Ma estrarre conoscenza da un "diluvio" di dati non è semplicemente un problema tecnico (Floridi, 2012), va a modificare il modo in cui si pongono gli interrogativi di ricerca, e come si cercano le risposte a questi interrogativi. Perciò, la prima, e forse più grave, conseguenza che il proliferare di questi nuovi dati ha avuto sulla ricerca sociale è stata la rinascita di un metodo scientifico guidato dai dati, in opposizione al metodo ipotetico deduttivo (Hempel, 1942) che è tipico della scienza sociale post-positivista. Chris Anderson, provocatoriamente, nel 2008 scriveva che il metodo scientifico sarebbe diventato presto obsoleto: «Petabytes allow us to say: 'Correlation is enough'...Correlation supersedes causation, and science can advance even without coherent models and unified theories». La principale argomentazione che fornisce chi sostiene il modello di scienza guidato dai dati è che, dato che ormai si è in grado di tracciare il comportamento umano con precisione e fedeltà mai vista, l'esplorazione e l'analisi dei nuovi dati è più utile della costruzione di modelli esplicativi. Nello specifico, Lazer, nell'articolo "Life in the network, the coming age of computational social science" pubblicato su *Science* nel 2009 un articolo che ha riscosso grande successo nella comunità scientifica (più di 2565 citazioni) identificò i nuovi dati presenti nella rete Internet come il cuore della *Computational Social Science* (CSS), una disciplina che attraverso tecniche di *data mining* e *machine learning*, applicate a enormi basi di dati, produce analisi sociali su larga scala, e quasi in tempo reale.

Le prime reazioni da parte dei sociologi a big data e computazione hanno affrontato la questione della dimensione tecnologica della ricerca sociale come l'ultima delle tante battaglie avvenute all'interno della guerra dei paradigmi della ricerca sociale. Ma si trattava di un modo miope di elaborare il problema, perché è necessario adottare una prospettiva aperta, ma allo stesso tempo critica, sia nell'approccio metodologico, sia nelle premesse epistemologiche. Infatti, anche la ricerca qualitativa è stata influenzata ampiamente dalle

innovazioni tecnologiche della ricerca sociale. Molti studi di comunità avvengono adesso all'interno della rete internet attraverso la *netnography*, e proprio la tecnologia ha permesso lo sviluppo di nuovi filoni di ricerca di stampo costruttivista come lo studio degli assemblaggi di dati (Aragona, Felaco, & Marino, 2018; Kitchin & Lauriault 2014). Inoltre, i contenuti generati dagli utenti *on-line* sui siti internet e sui social networks, seppure rappresentano una mole massiva di dati, vengono interpretati attraverso classificazioni *ex-post*, ed hanno le caratteristiche dei dati profondi, più utilizzati nella ricerca qualitativa e da studiosi di matrice ermeneutica (Boccia Artieri, 2015).

In verità, anche se la versione di CSS portata avanti da Lazer – conosciuta come scienza sociale computazionale “quantitativa” – è diventata in breve tempo il modo principale di realizzare la CSS, questa è nata molto prima dei big data, grazie alla possibilità di simulare fenomeni sociali attraverso i computer. All'inizio non si avvaleva di dati empirici prodotti nel mondo reale, i dati erano costruiti *ad hoc* dagli algoritmi di *machine learning*, che provavano a riprodurre i meccanismi in atto nella realtà sociale. Si trattava di un metodo prevalentemente generativo (Epstein, 2006), basato sulla spiegazione dei fenomeni sociali (Conte et al., 2012), in cui le teorie ricoprivano un ruolo fondamentale. È su questa convinzione che si sono sviluppate nuove proposte di scienze sociali *data intensive*, come ad esempio la *social data science* (Lauro, 2017), che collegano il tecnicismo informatico e le competenze analitiche al dominio di conoscenza (quello sociale) in cui esse sono impiegate; che porta con sé le sue teorie e le sue visioni della realtà. In questo modo, le conoscenze tipiche di una certa scienza si integrerebbero e combinerebbero con le discipline tecniche che sono necessarie per governare big data e computazione. Ovviamente, rimane forte l'*ethos* tecnologico, anche perché sono potenti i regimi discorsivi (Foucault, 1977) che si sono sviluppati nella società e che di volta in volta, presentando posizioni “integrate” o “apocalittiche” (Eco, 1964), contribuiscono a rinvigorirlo.

2. I regimi discorsivi a sostegno dell'*ethos* tecnologico

L'incessante produzione di big data e la conseguente scienza sociale computazionale quantitativa non possono essere spiegati unicamente come il risultato dei cambiamenti socio-tecnici che sono intervenuti nelle nostre società. Esse sono state anche guidate da una serie di facilitatori che hanno sviluppato un potente insieme di argomentazioni, appoggiate da tanti che credono nei benefici che questi nuovi dati possono garantire, e da gruppi di interesse che hanno obiettivi specifici affinché queste argomentazioni vengano alimentate. Tutto questo ha prodotto ciò che Foucault (1977) chiamerebbe un “regime discorsivo” sui big data che plasma, disciplina e riduce la complessità del fenomeno attraverso una serie di argomentazioni – che seppure diversificate, e a volte anche divergenti – rappresentano un insieme di ragionamenti concatenati, che favorisce l'adozione di nuove tecnologie.

Il primo fattore trainante dei big data è sicuramente di tipo economico e consiste nella promessa per le aziende di una maggiore efficienza e di maggiori profitti. Nel settore del commercio, Manyika et al. (2011) hanno identificato alcune attività aziendali, come il marketing, il *merchandising* e la logistica, in cui i big data possono determinare un aumento delle vendite e un miglioramento dell'efficienza. Ad esempio, le grandi catene di

supermercati generano una immensa quantità di dati che collega le vendite ai clienti attraverso carte fedeltà e carte di credito, e che può essere usata per escogitare nuove strategie dei prezzi, lanciare campagne pubblicitarie, identificare nuovi prodotti. Piuttosto che condurre inchieste campionarie *one-shot* che rileverebbero intenzioni o resoconti di acquisto, i dati delle transazioni elettroniche rilevano esplicitamente cosa i clienti hanno davvero acquistato, facilitando la segmentazione dei consumatori, la profilazione dei loro comportamenti di acquisto, la selezione dei canali di marketing da utilizzare per raggiungerli. Il secondo insieme di argomentazioni a sostegno dei big data si è invece consolidato nelle istituzioni governative, a tutti i livelli. È dagli anni novanta, all'interno della prospettiva delle politiche basate sull'evidenza (Stoker & Evans, 2016), che i dati vengono concepiti come conoscenza di governo. Desrosières (1998) ha notato che i dati hanno una natura politica perché svolgono da sempre un ruolo fondamentale nella formulazione, nel monitoraggio e nella valutazione delle politiche pubbliche. I big data perciò non sono visti come una novità, ma come l'ultimissima gamma di tecnologie che può ampliare e migliorare l'operato dei governi attraverso l'adozione di pratiche computazionali sempre più vaste e tempestive (Kitchin, 2014a). In questo contesto, sono due le retoriche principali che giustificano il ricorso ai big data. La prima riguarda la possibilità di risparmiare o guadagnare di più: perché si aumenta l'efficienza operativa avendo più dati a disposizione; perché si può aumentare il gettito fiscale potendo incrociare differenti fonti di dati; perché è possibile allocare i fondi con più precisione; perché si possono realizzare delle analisi di contesto più raffinate. Un'altra retorica che serve a promuovere l'uso dei big data nella sfera pubblica riguarda la sicurezza e la lotta al crimine. In tutti i paesi la prevenzione e la sicurezza si è sviluppata facendo uso delle nuove tecnologie: intercettazioni telefoniche, telecamere a circuito chiuso, fotografie aeree scattate da droni e immagini satellitari. Non a caso le agenzie di intelligence sono all'avanguardia nello sviluppo di strumenti per l'analisi dei big data e stanno investendo tanto in tecnologie e apparecchiature, come ad esempio gli 1.7 miliardi di dollari messi a disposizione dal governo degli Stati Uniti d'America nel 2017 per un centro in Utah per elaborare e immagazzinare dati intercettati dallo Stato.

Un altro elemento che ha facilitato la diffusione dell'analisi dei big data riguarda il legame, da sempre esistito, tra sviluppo tecnologico e miglioramento degli stili di vita. In questo regime discorsivo, le ICT (*Information Communication Technologies*) fornirebbero soluzioni per migliorare l'ambiente in cui viviamo, rendendolo più sostenibile, più pulito e più efficiente. L'ultimo esempio di questo legame sono le cosiddette *smart cities*, intese come luoghi ad elevata presenza di strumentazione con capacità computazionale, in grado di regolarne le attività quasi in tempo reale. Il passaggio dal "dumb" allo "smart" non ha riguardato solo le città, ma anche sistemi e dispositivi che utilizziamo quotidianamente e che oggi possono compiere una serie di scelte in modo autonomo e automatico attraverso l'esecuzione di algoritmi. Un buon esempio sono le automobili, provviste di numerosi dispositivi digitali che raccolgono dati sui percorsi e di sensori (di parcheggio, di frenata, etc.), che intervengono in aiuto del guidatore. Questi dati possono essere trasferiti all'azienda che ha prodotto l'auto per effettuare diagnosi in tempo reale o alle compagnie di assicurazione per localizzare e ricostruire le dinamiche di eventuali sinistri. Inoltre, con il passaggio ai sistemi "intelligenti", i dispositivi sono programmati per raccogliere quanti più dati possibili e sono anche diventati ubiqui, cioè capaci di muoversi insieme alle persone.

Molto legato al discorso "smart" c'è quell'insieme di argomentazioni che invece può essere ricondotto all'*openness*. È dalla creazione del web che l'obiettivo principale dei tecnologi informatici è aprire a tutti le potenzialità della rete. A volte per mera strategia di marketing, altre volte perché è insita nella mentalità dei creatori di Internet (come Tim Berners Lee) l'idea che il web debba essere diffuso, democratico e aperto a tutti. La retorica dell'apertura, da quando l'OCSE nel 2008 invitò i governi ad aprire i propri dati, ha visto la realizzazione di tanti progetti di *open data* sia negli USA che in Europa, che mirano a valorizzare il patrimonio informativo dei governi e degli enti a tutti i livelli. I dati aperti sono anche una conseguenza dell'importanza che la trasparenza e l'*accountability* (l'obbligo per un soggetto di rendere conto delle proprie decisioni, e di essere responsabile dei risultati conseguiti) stanno riscuotendo in questi tempi. Non è un caso che gli attori più determinati a rendere pubblici i loro dati siano anche quelli più affetti dalla crisi di fiducia che sta investendo la società: politici, istituzioni pubbliche, aziende nei settori dell'energia, dei trasporti, del credito, dell'ambiente, etc. (Chignard 2013).

D'altro canto, un insieme, anche numeroso, di critiche all'uso dei big data viene avanzato su più fronti. Proprio sulle *smart cities*, ad esempio, Kourtit e colleghi (2012) hanno osservato che sono state promosse dalle maggiori compagnie di servizi per software e hardware e che sono state adottate con entusiasmo dalle istituzioni a tutti i livelli. Lyon (2007), per questo motivo, ha sollevato dubbi etici e politici, perché l'uso estensivo dei big data porterebbe ad una forma di governo tecnocratica in cui il ruolo degli algoritmi sarebbe decisivo. Inoltre, un ulteriore rischio è che i big data potrebbero accelerare il processo di privatizzazione della sfera pubblica, perché questo tipo di dati è in possesso delle principali aziende di software e hardware (IBM, CISCO, Microsoft, Intel, Siemens, Oracle, SAP, etc.) e proviene dalle maggiori aziende di comunicazione e logistica (Amazon, Facebook, Google, etc.). Un esempio di tecnocratismo che viene spesso citato dai detrattori dei big data è l'*anticipatory governance*, una forma di profilazione che viene impiegata per prevedere comportamenti futuri e guidare gli interventi da effettuare (Kitchin & Lauriault, 2014). Questa forma di profilazione è molto diffusa per regolare i controlli negli aeroporti, dove i passeggeri vengono classificati per vari livelli di pericolosità, con il risultato che ricevono trattamenti diversi ai controlli, sulla base di ciò che potrebbero o non potrebbero fare (Dodge & Kitchin, 2004; Amoore, 2006). Altri sostengono invece che l'analisi delle tracce in rete, e l'uso dei sensori e dell'*Internet of things* nei contesti urbani e nella sfera pubblica, possa generare degli effetti da *big brother*, in cui la pervasività della tecnologia diverrebbe uno strumento di controllo. Esempi in tal senso ce ne sono già, basti pensare alle attività di sorveglianza di massa messe in campo dall'Agenzia per la Sicurezza Nazionale statunitense (NSA) con i servizi di intelligence di altri paesi, sia nei confronti di cittadini e istituzioni statunitensi che stranieri. L'NSA ha collezionato metadati sulle telefonate effettuate attraverso tutti i gestori statunitensi e una speciale divisione dell'agenzia, chiamata *Follow the Money*, raccoglie dati sulle transazioni finanziarie di importanti istituti internazionali come Visa, Mastercard e SWIFT. Attraverso il programma di sorveglianza PRISM, l'NSA ha accesso diretto ai server di molte delle principali aziende dell'informatica statunitense, quali Microsoft, Google, Yahoo!, Facebook, Apple, YouTube e Skype. L'agenzia monitora quindi le attività degli utenti, compresi scambi di messaggi, foto e video, conservando in particolare le liste di indirizzi utente usate nei servizi e-mail e di messaggistica istantanea. Il trattamento di questa

immensa mole di dati viene giustificata per questioni di sicurezza e legata alla lotta al terrorismo e alla salvaguardia dello stato.

Un'ultima serie di argomentazioni critiche proviene, infine, da coloro che ritengono che i big data siano il risultato di un modello di business che considera le informazioni come parte centrale delle pratiche economiche, in cui i dati stessi diventano merce di scambio per raggiungere potenziali clienti, sul web e fuori. A questo punto la raccolta intensiva delle tracce in rete sarebbe realizzata con il solo obiettivo di acquisire profitto e tutta la retorica sui big data rappresenterebbe un modo per alimentare i profitti delle principali aziende che si stanno avvantaggiando delle potenzialità offerte dalla *digital economy*.

Sia che ci si schieri con i detrattori della *data revolution* sia con gli entusiasti, comunque il risultato è che tutte queste retoriche contribuiscono a rinforzare e rinnovare di volta in volta il ruolo della tecnologia, intesa non solo come nuova capacità strumentale all'indagine empirica, in grado di produrre nuove conoscenze sui comportamenti e le relazioni sociali, ma anche come oggetto di studio della sociologia in sé.

3. Le sociologie digitali e il loro metodo

Dopo alcuni anni dall'avvento dei big data e della computazione, nel 2014, gli stessi autori che avevano evocato la crisi della sociologia empirica hanno pubblicato un articolo "oltre la crisi" sulla rivista *Big Data and Society* (Savage & Burrows, 2014). Infatti, indubbiamente, la diffusione senza precedenti di grandi basi dati provenienti da una miriade di fonti dà la possibilità di fare ricerca in molti modi, sia migliorando tecniche che venivano già usate in passato (ad esempio l'analisi del contenuto e l'analisi delle reti), sia creando nuove tecniche propriamente computazionali come il *machine learning*, le reti neurali, la *sentiment analysis*, la *natural language processing*, e molte altre. Allo stesso tempo, si aprono anche nuovi campi di ricerca sociologica critica, che mirano a valutare le conseguenze sociali dei processi di digitalizzazione, e di conseguente datificazione, di vari settori della società: dall'attività governativa e amministrativa all'istruzione, dall'economia, ai media.

Una prima applicazione dei big data riguarda gli studi urbani e, in particolare, l'analisi delle *smart cities*. La presenza pervasiva di dispositivi, sensori e telecamere nelle produce big data che permettono analisi in tempo reale della vita della città che possono essere impiegati per orientare e strutturare nuove modalità di gestione e di pianificazione strategica degli assetti urbani (Batty, 2013; Townsend, 2013; Kitchin, 2014b, 2018). L'analisi dei big data è stata estesa anche al turismo (Chareyron et al., 2014; Heerschap et al., 2014), soprattutto per ricostruire automaticamente i percorsi principali dei turisti a partire dalle tracce digitali lasciate sui social media.

Un'altra applicazione riguarda l'uso dei big data per le previsioni sugli esiti elettorali (Ceron et al., 2014; Schoen et al., 2013) o per comprendere le opinioni dei cittadini su questioni rilevanti (Tumasjan et al., 2010; Birmingham & Smeaton, 2011). Un esempio di quest'ultimo tipo di uso è l'*e-consultation*, che ha l'obiettivo di adattare le politiche ai commenti dei suoi beneficiari. L'*e-consultation* è stata utilizzata in Italia per i servizi sanitari (Ferro et al., 2013). Nel 2008, la Regione Piemonte ha lanciato un servizio di telemedicina per alcune aree montuose scarsamente abitate. Fu condotta una campagna comunicativa attraverso i social

media, dove i pazienti e i medici comunicavano pubblicando testi e video su *Facebook*, *Twitter* e *Youtube*, esprimendo le loro opinioni rispetto al servizio di telemedicina. Le piattaforme sono usate anche come forme di partecipazione alla vita pubblica. Un esempio è il network *Mysociety* (<https://www.mysociety.org/>). Fondato nel Regno Unito, coinvolge i cittadini nella vita pubblica attraverso applicazioni come *TheyWorkForYou* (<https://www.theyworkforyou.com>), che mira a controllare l'azione politica e *Fixmystreet* (<https://www.fixmystreet.com/>), dove i cittadini possono segnalare un problema indicando un'area specifica attraverso il codice postale e una mappa e descrivendo il problema. Gli utenti possono controllare in tempo reale, attraverso foto e video, cosa le autorità hanno fatto per risolvere il problema.

Anche nel campo dell'istruzione i big data e la computazione stanno consentendo l'organizzazione, l'analisi e la visualizzazione di indicatori chiave delle performance delle attività di ricerca, di didattica e amministrative. Ad esempio, nelle piattaforme di *web learning*, le analisi in tempo reale consentono di personalizzare l'esperienza formativa degli utenti, datificando il processo formativo attraverso feedback continui e personalizzazione del materiale di studio, e prevedendo i risultati formativi probabili (Ferguson, 2012; Mayer-Schönberger & Cukier, 2013). Esempi in questo campo sono *The Learning Curve* messa a punto da Pearson con l'ausilio di IBM e la piattaforma *Data Futures*, che è stata lanciata dal Ministero dell'Istruzione britannico per creare un sistema tempestivo e efficiente di raccolta dati sull'istruzione accademica (Williamson, 2016).

Insomma, la raccolta, analisi ed elaborazione di dati, reti e relazioni attraverso tecniche computazionali riesce a creare nuovi punti di contatto tra la vita sociale e la ricerca sociale. È ciò che sostengono quegli studiosi che hanno riassunto l'insieme delle innovazioni portate da computazione e big data e dai regimi discorsivi che li sostengono, con un unico termine, "digitale", e che portano avanti una sociologia digitale (Lupton, 2014; Orton-Johnson & Prior, 2013), o meglio delle "sociologie digitali" (Daniels, Gregory & Cotton, 2016). Nella sociologia digitale si sostiene un impegno attivo da parte dei sociologi nel confrontarsi con la dimensione tecnologica dell'indagine sociale, e nel definire un metodo digitale (Rogers, 2013; Marres, 2017) che solleva, e allo stesso tempo prova a risolvere, nuove e vecchie questioni che hanno origine nelle diverse tradizioni metodologiche ed epistemologiche che convivono nella nostra disciplina. Questo metodo più che seguire derive neo-empiriste, conserva i principali dogmi del metodo scientifico post-positivista, ma promuove l'uso congiunto di induzione e deduzione. Per i fautori delle sociologie digitali, la conoscenza scientifica viene perseguita utilizzando tecniche computazionali "guidate" per scoprire ipotesi da sottoporre a successivo controllo empirico. Il processo è guidato in quanto le teorie esistenti sono impiegate per direzionare lo svolgimento della scoperta, piuttosto che, come vorrebbe la scienza computazionale quantitativa, sperare di identificare tutte le relazioni esistenti in un database, e assumere che queste siano in qualche modo sensate. Invece, il modo in cui i dati sono costruiti o rianalizzati è guidato da alcuni assunti, sostenuto dalla conoscenza teorica e pratica e dall'esperienza su come le tecnologie e le loro configurazioni siano in grado di produrre materiale empirico valido e pertinente per la ricerca. L'analisi dei dati non viene effettuata attraverso ogni possibile tecnica di *data mining*, ma invece le decisioni tecniche vengono indirizzate da riflessioni teoriche che sono prese per capire come meglio analizzare un nuovo database in modo da ricavarne informazioni

che siano interessanti e preziose anche per ulteriori ricerche. In pratica, il metodo utilizzato è abducente (Pierce, 1883) e mira ad inserire risultati inaspettati in un quadro interpretativo. La sociologia digitale assume che l'uso di metodi analitici più o meno computazionali sia diventato lo standard della ricerca sociologica, ma allo stesso tempo, non lo ritiene un imperativo. Può essere utile per produrre nuove letture dei fenomeni sociali, ma gli studiosi devono rimanere in guardia quando decidono di inserirli nella loro cassetta degli attrezzi, tanto che, secondo Marres (2017), la loro capacità metodologica andrebbe testata ancora. Le tecniche dovrebbero essere discusse apertamente, valutandone gli impatti sui disegni di ricerca, sulla formulazione degli interrogativi e, quando serve, sulle strategie di test delle ipotesi. Inoltre, il loro impiego andrebbe affrontato criticamente e con creatività, pronti a confrontare le nostre cornici empiriche con le riconfigurazioni in divenire delle infrastrutture, dei dispositivi, delle piattaforme e delle interfacce digitali (Jacomy, 2015); tutto ciò con la consapevolezza che gli oggetti di indagine della sociologia non sono dati una volta e per tutte, ma che sia la stessa attività di ricerca a costruirli. Big data e computazione da soli non garantiscono un rafforzamento dell'attività di ricerca, ma anzi possono del tutto inficiare i risultati delle indagini. Sia nella ricerca sociale che utilizza il digitale come strumento, sia in quella che considera il digitale come oggetto di studio, dati e metodi tradizionali dovrebbero continuare ad essere semplicemente fondamentali, per aggiungere senso e significato ai diversi risultati scoperti a livello più macro (boyd & Crawford, 2012).

Infatti, non sono pochi i limiti dei big data e delle tecniche computazionali. Il tema della rappresentatività e del campionamento è in genere una delle prime questioni di metodo che riguardano l'uso dei *Big Data*. È molto diffuso il problema della sottocopertura che si verifica quando alcune unità della popolazione sono sistematicamente escluse dalla ricerca. Sul web parte della popolazione è irraggiungibile per definizione, perché può non avere accesso alla rete, oppure perché molti semplicemente sono consumatori passivi delle informazioni contenute nella rete piuttosto che utenti che partecipano attivamente nel web 2.0. Inoltre, come dimostra un'ampia letteratura sul *digital divide* (Di Maggio & Hargittai, 2001; Van Dijk & Hacker, 2003), l'accesso alle piattaforme social può essere segmentato in relazione a variabili sociodemografiche come nazionalità, età, genere, titolo di studio e reddito, e ciò comporta un errore sistematico di sottostima di interi strati della popolazione. Un ulteriore limite riguarda la validità, cioè la rispondenza del dato a ciò che il ricercatore intende rilevare. Grande attenzione deve essere impiegata nell'operazionalizzazione dei costrutti teorici. Le definizioni operative non sono adeguatamente documentate, per cui si deve provare a ricostruire *ex-post* quali sono i costrutti latenti sottostanti. Un'ultima questione riguarda l'accesso. Sono i *providers* che determinano quali sono i processi per accedere ai dati. Alcune ricerche, infatti, potrebbero essere viste con maggior favore di altre, delineando di fatto delle iniquità nell'accessibilità in ragione della disponibilità economica o dei contatti intrattenuti con i providers. Insieme a questo, un'ulteriore questione da tenere in considerazione è la possibilità di accedere ai microdati direttamente dalle interfacce attraverso le API o con tecniche di *scraping*. Tutto ciò può avere conseguenze sociali rilevanti in un sistema di produzione dei dati che è principalmente nelle mani delle grandi multinazionali della logistica e della comunicazione.

Conclusioni

Per quanto si possa condividere o meno la proposta di una, o addirittura più di una, sociologia digitale, è indubbio che le società *data intensive*, e i loro processi di digitalizzazione e di datificazione, impattino profondamente sul metodo della ricerca sociale. In primo luogo, perché rafforzano le fonti empiriche a disposizione dell'indagine sociale, ma anche perché favoriscono l'interdisciplinarietà tra diverse comunità scientifiche interessate all'analisi della realtà sociale, incoraggiando l'integrazione di dati e di tecniche. Una sapiente miscela tra teoria sociologica, computazione e nuovi dati può perciò contribuire a migliorare l'analisi dei processi sociali, se adeguatamente integrata agli approcci tradizionali della ricerca sociale, sia quantitativi (survey e esperimenti), sia qualitativi (interviste ed etnografia). Dopo un decennio in cui big data e computazione sono stati visti alternativamente come il nuovo oro delle scienze sociali (Lazer, 2009; Mayer-Schonberger & Cuckier, 2013), o come una pericolosa nuova forma di quantofrenia (boyd & Crawford, 2012), un gruppo, ormai numeroso, di studiosi (Ruppert, 2015; O'Sullivan, 2017; Halford & Savage, 2017) ha delineato la possibilità di costruire un terreno comune in cui sia possibile fronteggiare le sfide e, al contempo, avvantaggiarsi delle opportunità che i nuovi dati e la computazione pongono. È solo in questa cornice di metodo mista e critica che l'attuale carattere tecnologico dell'indagine sociale può essere riportato proficuamente all'interno dell'alveo della nostra disciplina. Altrimenti, riprendendo la preziosa metafora Millsiana citata all'inizio di questo lavoro, tutti questi dati digitali rimarrebbero semplicemente «un insieme di frammenti confusi di realtà umana» (Wright Mills, 1959, p.47).

Bibliografia

- Amaturo, E., & Aragona, B. (2016). La "rivoluzione" dei nuovi dati: quale metodo per il futuro, quale futuro per il metodo?. In F. Corbisiero, & E. Ruspini (a cura di), *Sociologia del futuro. Studiare la società del ventunesimo secolo* (pp. 25-50). Lavis: Wolters Kluwer Italia.
- Amaturo, E., & Aragona, B. (2017). Introduction. In C. Lauro, E. Amaturo, M. G. Grassia, B. Aragona, & M. Marino (Eds.), *Data Science and Social Research: Epistemology, Methods, Technology and Applications* (pp.1-9). Heidelberg: Springer-Verlag.
- Amoore, L. (2006). Biometric borders: Governing mobilities in the war on terror. *Political Geography*. vol. 25, pp. 336-351.
- Amoore, L. (2009). Algorithmic war: Everyday geographies of the war on terror. *Antipode*, 41, pp. 49-69.
- Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired magazine*, 16(7). Retrived from http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.
- Aragona, B., Felaco, C., & Marino, M. (2018). The Politics of Big Data Assemblages. *Partecipazione e Conflitto*. vol 11(2), pp. 448-471.
- Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), pp. 274-279.
- Birmingham A. & Smeaton A. F. (2011). *On using Twitter to monitor political sentiment and predict election results*. 2011 Workshop on Sentiment Analysis where AI meets Psychology (SAAIP).
- Boccia Artieri, G. (2015). *Gli effetti sociali del web. Forme della comunicazione e metodologie della ricerca online*. Milano: FrancoAngeli.
- boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), pp. 662-679.
- Ceron A., Curini L., Iacus S. M. & Porro G. (2014). "Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France". *New Media & Society*, 16(2), pp. 340-358.

- Chareyron, G., Da-Rugna, J., & Raimbault, T. (2014). *Big data: A new challenge for tourism*. In Big data (Big data), 2014 IEEE international conference on (pp. 5-7). IEEE.
- Chignard, S. (2013). A brief history of Open data. *Paris Tech Review*.
Retrieved from <http://www.paristechreview.com/2013/03/29brief-history-open-data/>.
- Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., V. Loreto, S. Moat, J. -P. Nadal, A. Sanchez, A. Nowak, A. Flache, M. San Miguel, D. Helbing (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1), pp. 325-346.
- Daniels, J., & Gregory, K., Cottom, T.M. (Eds.). (2016). *Digital sociologies*. Bristol: Policy Press.
- Desroisères, A. (1998). *The politics of large numbers: a history of statistical reasoning*. Cambridge: Harvard University Press.
- DiMaggio, P., & Hargittai, E. (2001). From the 'digital divide' to 'digital inequality': Studying Internet use as penetration increases. Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University, 4(1), pp. 4-22.
- Dodge, M., & Kitchin, R. (2004). Flying through code/space: the real virtuality of air travel, *Environment and Planning A*, 36, pp. 195-211.
- Domo (2017). *Data Never Sleeps 5.0*. Retrieved from <https://www.domo.com/learn/data-never-sleeps-5>.
- Eco, U. (1964). *Apocalittici e integrati: la cultura italiana e le comunicazioni di massa*. Milano: Bompiani,
- Epstein, J.M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*, Princeton: Princeton University Press.
- Espeland, W.N., & Stevens, M.L. (2008). A Sociology of Quantification. *European Journal of Sociology*, 49(3), pp. 401-437.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), pp. 304-317.
- Floridi, L. (2012). Big data and their Epistemological challenges. *Philosophy and technology*, 25(4), pp. 435-7
- Foucault, M. (1977). *Discipline and Punish*. London: Allen Lane.
- Halford S., & Savage, M. (2017). Speaking Sociologically with Big Data: symphonic social science and the future of big data analytics. *Sociology*, 51(6), pp. 1132-1148.
- Heerschap, N., Ortega, S., Priem, A., & Offermans, M. (2014, May). *Innovation of tourism statistics through the use of new big data sources*. In 12th Global Forum on Tourism Statistics, Prague, CZ.
- Hempel, C. G. (1942). The function of general laws in history. *The Journal of Philosophy*, 39(2), pp. 35-48.
- Jacomy, B. (2015). *Une histoire des techniques*. Paris: Le seuil.
- Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Sciences*. San Francisco: Chandler.
- Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the information age solving problems with visual analytics*. Eurographics Association.
- Kitchin, R. (2014a). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin, R. (2014b). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1-14.
- Kitchin, R. (2018). The realltimeness of smart cities. *TECNOSCIENZA: Italian Journal of Science & Technology Studies*, 8(2), pp. 19-42.
- Kitchin, R. & Lauriault, T. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work. *Social Science Research Network*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Kourtit, K., Nijkamp, P., & Arribas-Bel, D. (2012). Smart cities perspective – a comparative European study by means of self-organizing maps. *Innovation*, 25(2), pp. 229-246.
- Lauro, C. (2017). Preface. In C. Lauro, E. Amato, M. G. Grassia, B. Aragona, & M. Marino (Eds.), *Data Science and Social Research: Epistemology, Methods, Technology and Applications* (pp.v-vi). Heidelberg: Springer-Verlag.
- Lazarsfeld, P. F., & Rosenberg, M. (1955). *The language of social research: a reader in the methodology of the social sciences*. New York, NY, US: Free Press.
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Jebara, T. (2009). Life in the network: the coming age of computational social science. *Science*, 323(5915), pp. 721-723.
- Lupton, D. (2014). Beyond techno-utopia: Critical approaches to digital health technologies. *Societies*. vol. 4, pp. 706-711.

- Lyon, D. (2017). *Surveillance Studies: An Overview*, Cambridge: Polity.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
- Marradi, A. (2007). *Metodologia della ricerca sociale*. Bologna: Il Mulino.
- Marres, N. (2017). *Digital sociology: The reinvention of social research*. Cambridge: John Wiley & Sons.
- Mayer-Schönberger V., & Cukier K. (2012). *Big Data: A revolution that transforms how we work, live, and think*. Boston: Houghton Mifflin Harcourt.
- Mills, C. W. (1959). *L'immaginazione sociologica*. trad. it. Milano: Il Saggiatore, 1962.
- Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. Cambridge: MA: MIT.
- O'Sullivan, D. (2017). Big Data: why (oh why?) this computational social science?. Retrived from <https://escholarship.org/uc/item/0rn5n832>.
- Orton-Johnson, K., & Prior, N. (Eds.). (2013). *Digital sociology: Critical perspectives*. London: Palgrave Macmillan.
- Peirce, C.S. (Ed.). (1883). *Studies in logic*. Boston, MA: Little, Brown and Company.
- Rogers, R. (2013). *Digital methods*. Cambridge: MIT press.
- Ruppert, Evelyn. 2015. Big Data Economies and Ecologies. In: Linda McKie and Louise Ryan, eds. An End to the Crisis of Empirical Sociology? Trends and Challenges in Social Science Research. London: Routledge, pp. 12-26. ISBN 9781138828674.
- Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, 41(5), pp. 885-899.
- Savage, M., & Burrows, R. (2014). After the Crisis? Big Data and the Methodological Challenges of Empirical Sociology. *Big Data and Society*, 1(1), pp. 1-6.
- Schoen H. & Gayo-Avello D., Takis Metaxas P., Mustafaraj E., Strohmaier M., Gloor P. (2013). The power of prediction with social media. *Internet Research*, 23(5), 528-543.
- Stoker G., & Evans M. (Eds.). (2016). *Evidence-based Policy Making in the Social Sciences: Methods that Matter*. Bristol: Policy Press.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G. & Welp, I. M. (2010). *Predicting elections with twitter: What 140 characters reveal about political sentiment*. In 2010 AAAI 4th International Conference on Web and Social Media (ICWSM), pp. 178-185.
- Townsend, A. M. (2013). *Smart cities: Big data, civic hackers, and the quest for a new utopia*. WW Norton & Company.
- The Economist, (2010). *The data deluge: Businesses, governments and society are only starting to tap its vast potential*. Print edition. Retrived from <http://www.pewinternet.org/2015/01/09/demographics-of-key-social-networking-platforms-2/>.
- Van Dijk, J., & Hacker, K. (2003). The digital divide as a complex and dynamic phenomenon. *The information society*, 19(4), pp. 315-326.