

Paolo Natale

Il terzo paradigma. Lo studio del web

(doi: 10.53227/101177)

Rivista di Digital Politics (ISSN 2785-0072)

Fascicolo 1, gennaio-aprile 2021

Ente di afferenza:

()

Copyright © by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati.

Per altre informazioni si veda <https://www.rivisteweb.it>

Licenza d'uso

Questo articolo è reso disponibile con licenza CC BY NC ND. Per altre informazioni si veda <https://www.rivisteweb.it/>

Paolo Natale

Il terzo paradigma

Lo studio del web^{*}

THE THIRD PARADIGM. THE STUDY OF THE WEB

If the digital revolution has produced an epochal change in society in general, the same thing has happened, albeit more slowly, also in socio-political research and in the tools that the researcher can use in his or her analysis. This paper aims to clarify, first of all, the type of data we find on the internet (digitized data, virtual data and digital data), distinguishing in particular between «provoked-data» and «user-generated data». The latter are the ones that present absolutely new characteristics in the history of research, being generated solely by the will of users to leave traces of themselves on social media. We cannot approach these data through one of the two traditional paradigms of quantitative and qualitative research (respectively, positivism and interpretivism), which both provided for a strong, albeit different, interaction between scholar and social actors being studied. In fact, digital data, Big Data, have a genesis completely free of intrusiveness, they are not '*solicited*' in any way by the researcher, and the approach to their study therefore responds to a sort of «third way», a different and original paradigm, whose distinctive features are presented and discussed here, together with the most relevant tools of web analysis.

KEYWORDS *Web Analysis, Methodological Paradigms, Digital Data, Big Data, Virtual Data.*

A trent'anni dalla nascita di internet e del web, dopo oltre quindici dall'avvento di Facebook e poi degli altri social media, possiamo dire di trovarci nel mezzo di una vera e propria «rivoluzione digitale», un cambiamento per certi versi epocale della società in generale. Una rivoluzione ancora più radicale e invasiva del computer, che sta alla pari con quella dell'automobile, del telefono, della radio o della televisione. Qualcosa che ha cambiato la vita, e continua a cambiarla sempre più radicalmente, ogni giorno che passa, grazie alle costanti innovazioni che entrano a far parte degli strumenti e delle applicazioni che la Rete fornisce: una sorta di «quarta rivoluzione» (Floridi 2014).

Paolo Natale, Dipartimento di Scienze Sociali e Politiche - Università degli Studi di Milano - Via Conservatorio, 7 - 20122 Milano. email: paolo.natale@unimi.it, orcid: 0000-0002-2903-4234.

^{*} Questo articolo presenta una versione ridotta, rivista e aggiornata dei materiali presentati nel volume *Web e Social Media*, Santarcangelo di Romagna, Maggioli, 2017.

E se questo è vero per la vita quotidiana di milioni di individui, lo è anche, sia pur con un iter più lento, per la ricerca politica e sociale, per i metodi di analisi che alla vita del web si possono associare. Internet e i social media non hanno certo fatto piazza pulita dei fondamenti metodologici delle scienze politiche, così come non hanno cancellato dalle nostre vite i rapporti umani faccia a faccia. Sondaggi, interviste ed etnografie *offline* continuano a rappresentare risorse essenziali per lo studio dei fenomeni politici, e noi (Covid permettendo) continuiamo a uscire a cena con gli amici e a fare passeggiate. Ciò non significa, però, che nulla sia cambiato.

In una realtà dove il livello di densità sociale è in costante incremento, con una numerosità di comunicazioni e interazioni interpersonali impensabile nel recente passato, le strade pullulano di smartphone e – non a caso – i dati digitali diventano un oggetto di analisi privilegiato. Così come non sorprende più le schiere di volti chini a scrutare WhatsApp sui mezzi pubblici, l'adattamento di tecniche di ricerca convenzionali ai contesti e dati della Rete, allo stesso modo, va considerato un fatto *normale*.

Nuove abitudini tecnologicamente guidate prendono piede, mentre approcci metodologici e statistici innovativi – dalla sentiment analysis al machine learning – diventano sempre più mainstream. Questo accade perché le piattaforme online, le pratiche di comunicazione che ospitano e persino i metodi che mirano a estrarne le informazioni sono tutti prodotti di una medesima realtà sociale che, inevitabilmente, si influenzano a vicenda.

Lo studio dei dati digitali è una pratica di ricerca relativamente giovane, se comparata con gli approcci metodologici più consolidati (Airoldi 2017a). Ogniquale volta un metodo innovativo guadagna terreno, sempre ci si divide in entusiasti e scettici – a seconda delle convinzioni. Questa tendenza è particolarmente evidente nel caso dei metodi di ricerca online, i quali non sono che una conseguenza minore del più dirompente processo di digitalizzazione e datafication alimentato dalla diffusione di internet e dei social media, che ha sconvolto pressoché ogni aspetto dell'esistenza umana.

Le sfide portate dall'analisi dei dati digitali ad approcci collaudati nell'ambito della ricerca sociale, qualitativa così come quantitativa, sono senza pari. Parole chiave come «campionamento», «accesso al campo», «etica della ricerca», «rappresentatività» sono state interamente risemantizzate, mentre nuovi termini hanno fatto il loro debutto nel gergo comune – un esempio su tutti: «Big data» (Mayer-Schoenberger e Kenneth 2013).

Il vento del cambiamento aveva inizialmente gonfiato le fila degli scettici, aprioristicamente convinti dell'inaffidabilità di ricerche effettuate online. Successivamente, sono stati gli entusiasti a prendere il sopravvento, guidati

dall'ingenua certezza che l'*internet research* e il *data mining* avrebbero fatto tabula rasa dei classici metodi della ricerca politica e sociale (Airoldi 2017b).

Oggi si può sicuramente affermare che le cose si siano maggiormente relativizzate: sappiamo che gli strumenti per studiare la Rete servono, e continuiamo a rimanere convinti che anche gli strumenti per studiare il comportamento degli individui fuori dalla Rete siano altrettanto importanti.

1. I dati della Rete

Prima di analizzare in profondità la rivoluzione che ha portato la Rete nei tradizionali paradigmi della ricerca superando, o meglio, affiancando l'approccio «positivista» e quello «interpretativista», e di riassumere le caratteristiche di questa «terza via», è bene introdurre una distinzione, per fare chiarezza su quale sia l'oggetto di studio precipuo del web. Secondo quanto suggerito da Rogers (2015), coesistono in Rete tre tipi di dati: digitalizzati, virtuali e digitali.

Per «dati digitalizzati» si intendono quei materiali che, provenienti da fonti diverse, sono stati riversati sul web in formato digitale (dunque: libri, programmi televisivi, video musicali, film, articoli di giornali o di riviste scientifiche, fotografie, eccetera). Lo studio e l'analisi di questi prodotti non è sostanzialmente differente da quello che generalmente si effettua offline, come ad esempio nell'analisi del contenuto testuale, ma è soltanto molto più semplice, più rapida e meno onerosa dal punto di vista economico.

Per «dati virtuali» si intendono invece tutte quelle informazioni che si possono ricavare dall'utilizzo sul web dei tradizionali metodi di ricerca, sia quantitativa che qualitativa (dunque, dai *focus group* ai sondaggi in Rete). Se la genesi della produzione di questo tipo di dati è dunque differente, non lo è il risultato ottenuto: un questionario online è ad esempio del tutto simile ad un questionario postale, trattandosi entrambi di questionari auto-compilati, senza l'ausilio dell'intervistatore. Così come un *focus group* effettuato online, attorno ad un tavolo virtuale, anziché seduti ad un tavolo reale.

Dal punto di vista prettamente metodologico, per questi due primi tipi di dati cambia poco per il ricercatore averli ottenuti attraverso il web ovvero da una raccolta effettuata offline. L'approccio sarà sostanzialmente simile e può facilmente venir ricondotto ad uno dei due tradizionali paradigmi di ricerca.

I «dati digitali», al contrario, sono di tutt'altro tipo. Consistono infatti nelle «tracce» lasciate in Rete da parte degli utenti che vi accedono: sono i tweet, i like, i commenti, le visite ai diversi siti, i post scritti da chi accede ai social media. Costituiscono un corpus del tutto differente, un materiale empi-

rico geneticamente inedito, non riconducibile ad alcun altro oggetto di ricerca sociale.

Per meglio differenziare i dati virtuali da quelli digitali, entrambi prodotti sul web, si fa riferimento in particolare alla dicotomia tra *provoked data* e *user-generated data*; con il primo termine si intendono sostanzialmente le risposte fornite in Rete, dagli utenti, a seguito di specifiche domande o richieste formulate dal ricercatore, mentre nel secondo caso ci si riferisce ad informazioni generate in maniera autonoma da un individuo, su Facebook, con un tweet, su una chat, e così via. Il primo tipo di dati nasce quindi sulla base del classico cortocircuito stimolo-risposta, tipico delle ricerche offline, mentre il secondo nasce spontaneamente dalla volontà esclusiva dell'individuo stesso, «auto-stimolato».

Analizzare quindi i «dati digitali» significa in primo luogo cercare di comprendere come si muovono, e come interagiscono, gli utenti che popolano i diversi ambiti di internet. Fare ricerca online significa proprio questo, occuparsi delle principali modalità attraverso cui studiare la Rete, e ciò che accade nella Rete, soprattutto in quella parte dove sono protagonisti gli individui, gli attori sociali immersi nelle costanti interazioni con il resto del mondo e che, attraverso questo processo, ci dicono qualcosa su di loro, sui loro atteggiamenti, sulle loro opinioni, sul loro stile di vita. Ma il percorso che ha portato ai dati digitali è stato piuttosto lungo.

2. I dati digitalizzati

Per molti di noi oggi appare normale andare in vacanza portando con sé una piccola libreria portatile, dentro un cosiddetto «e-reader», dove possono venir stipati centinaia e centinaia di volumi, sotto forma di *e-book*. Un tempo ovviamente non era così. Ma i primi vagiti dell'era computerizzata, negli anni Settanta-Ottanta del secolo scorso, hanno avuto come uno degli obiettivi quello di condensare tutto il sapere, prodotto e in produzione, sotto la forma di dati fruibili e ispezionabili: è stato il lungo processo della digitalizzazione. Il passaggio dall'analogico al digitale ha rappresentato di fatto la prima grande rivoluzione collegata alla fruizione dei media (libri, appunto, giornali, brani musicali, registrazioni radio-televisive, e così via), che da quel momento in poi, poco alla volta, sono andati a costruire una immensa biblioteca di dati digitalizzati.

Inutile sottolineare come, per un ricercatore interessato alle analisi socio-politiche, abituato a compiere i suoi studi di *content analysis* (l'analisi del contenuto) o di *discourse analysis* (l'analisi del discorso) a partire da testi cartacei o

da registrazioni, il materiale che si rendeva disponibile nelle banche-dati computerizzate diventava una fonte inesauribile di documentazione empirica, un gigantesco archivio virtuale.

Nel corso del tempo, queste tecniche prima circoscritte solamente ad un gruppo ristretto di studiosi, a causa degli elevati costi per procurarsi e trasformare il materiale di base, al fine di renderlo leggibile agli specifici software su computer, divennero appannaggio di molti ricercatori, che diedero vita ad una sorta di rinascita di una metodologia rimasta di nicchia per molti anni.

Ma, come è stato più volte sottolineato, si trattava ancora di utilizzare la Rete a partire da «dati» che non erano nati in Rete, non erano creati dal web, la cui funzione era solamente di supporto, per ridurre i costi ed il tempo necessario per entrare in possesso di raccolte digitali, sulle quali svolgere poi le successive analisi.

Soltanto una decina d'anni più tardi, dopo i primi esperimenti «di nicchia», si cominceranno ad effettuare su vasta scala analisi del contenuto il cui oggetto di studio non erano più, o non solo, i dati digitalizzati, ma anche i dati digitali, quelli nati in Rete, propri dei social media.

Internet stava però già modificando in maniera essenziale la vita del cittadino comune e dello studioso, del ricercatore, ma lo faceva a partire da ambienti *diversi* dalla Rete stessa, da eventi ad essa estranei che venivano riversati sul web attraverso la loro digitalizzazione (Negroponte 1995). Il web fungeva da mero ausilio logistico, se così si può dire. Permetteva di accorciare i tempi della ricerca, fosse quella di un buon ristorante, del film preferito ovvero dell'analisi del discorso di un politico, del contenuto di un testo, di un articolo di giornale, dell'opera omnia di James Joyce.

Il passo successivo dell'evoluzione di internet fu l'utilizzo della Rete come supporto, o succedaneo, delle indagini offline, prima nella ricerca quantitativa e poi in quella qualitativa. L'uso di strumenti computer-assisted era già entrato con forza negli anni Novanta nella gestione delle interviste telefoniche (il Cati, *Computer assisted telephone interview*) e in quelle *face-to-face* (il Capi, le *personal interview*), velocizzando nettamente le antiche modalità di raccolta delle informazioni delle indagini demoscopiche e favorendone la frenetica e, a volte, incontrollata diffusione (Natale 2004; 2009).

I risultati di quelle indagini si trasformavano agevolmente in dati digitalizzati e, grazie allo sviluppo della Rete, andavano ad alimentare una crescente mole di informazioni all'interno delle nascenti banche dati, grazie all'attività sia di istituti universitari che, ancor più, degli istituti di ricerca di mercato, come la Nielsen, da subito all'avanguardia in molti paesi nel monitoraggio dei consumatori.

Con l'ulteriore diffusione di internet e con il suo crescente utilizzo da parte di un numero sempre maggiore di utenti, a cavallo del nuovo millennio anche in Italia si iniziò a sperimentare, con risultati soddisfacenti, la ricerca demoscopica tramite web, con questionari standardizzati compilati autonomamente dal campione di cittadini estratti ad hoc per l'indagine.

3. I dati virtuali nella ricerca quantitativa: la web survey

I risultati ottenuti in questo modo, come abbiamo detto, possono essere chiamati «dati virtuali», poiché differiscono da quelle rilevati offline, dal punto di vista della loro genesi, ma non sono prodotti nativi della Rete, come nel caso dei «dati digitali».

Come noto, due sono le modalità prevalenti di effettuazione di una survey: l'intervista quantitativa e il questionario auto-compilato, ovvero un sondaggio in cui è prevista la presenza di un intervistatore e uno in cui non è prevista la sua presenza. Se le premesse metodologiche e le logiche sottostanti ad entrambe le modalità di raccolta dei dati e delle informazioni sono del tutto simili, variano invece in maniera sostanziale da una parte l'impatto dello strumento di raccolta sul soggetto «indagato», dall'altra la forma e la gestione dello strumento stesso.

Per quanto riguarda questi due ultimi elementi, i limiti di un questionario auto-compilato sono ben noti: deve essere completamente auto-esplicativo ed è auto-gestito da chi dovrà rispondere, sia nei tempi sia nei modi della sua compilazione. Per quanto riguarda invece l'impatto sull'intervistato, l'indagine basata sulla auto-somministrazione evita di scontrarsi con le distorsioni provocate dall'influenza dell'intervistatore (il cosiddetto «effetto Hawthorne», quel fenomeno per cui l'osservato muta il proprio comportamento o le proprie risposte a causa della presenza di un osservatore o di un intervistatore «esterno»). Essendo uno strumento *freddo*, nel questionario auto-compilato il rispondente ha meno possibilità di essere condizionato dalla desiderabilità sociale, per cui questa modalità di raccolta delle informazioni garantisce un più elevato distacco ed una maggiore oggettività nei risultati del sondaggio.

Nella lunga storia delle indagini demoscopiche, l'utilizzo del questionario auto-compilato si è prevalentemente associato con la rilevazione postale che, rispetto all'intervista *face-to-face* (l'altra modalità più diffusa per la raccolta delle informazioni, prima dell'avvento dell'intervista telefonica) porta con sé, come vedremo tra breve, alcuni decisivi vantaggi, ma anche molti svantaggi. Vantaggi e svantaggi del tutto simili allo strumento che, con l'avvento di inter-

net, ha poco alla volta sostituito l'utilizzo dell'indagine postale: il cosiddetto Cawi (*Computer assisted web interviewing*) o web-survey, vale a dire la raccolta delle informazioni (i «dati virtuali») tramite questionari inviati attraverso la Rete.

Le due modalità di raccolta, dal punto di vista delle implicazioni metodologiche, sono pressoché simili: dopo aver estratto un campione di nominativi dall'universo di riferimento, nel primo caso viene inviato ad un indirizzo di posta un questionario, che verrà compilato e rispedito al mittente (l'istituto o l'ente che conduce il sondaggio); nel secondo caso si invia ad un indirizzo di posta elettronica il link ad un questionario che verrà compilato online.

Il vero svantaggio – insuperabile – della rilevazione effettuata tramite posta risiede nel lasso di tempo che intercorre tra la spedizione ed il rientro dei questionari auto-compilati. Se non particolarmente sensibilizzati, i rispondenti impiegano infatti molti giorni, per non dire settimane, per restituire le proprie risposte, cui si aggiungono gli inevitabili ritardi dovuti al sistema postale. La *web-survey* al contrario viene effettuata molto più rapidamente, con un netto risparmio oltre che di denaro, anche di tempo. Soprattutto se, come nel caso dell'indagine postale, anche in quella via web si costruiscono come vedremo panel di riferimento affidabili, nel giro di 3-4 giorni il processo dell'intervista potrà essere concluso con un numero di rispondenti significativamente elevato.

Concentriamo ora la nostra attenzione su quest'ultima modalità di raccolta delle informazioni (il Cawi), che è di fatto una sorta di incrocio tra l'intervista tramite Capi e la rilevazione postale: la gestione del questionario è demandata ad un apposito software (come nel Capi), ma il questionario è compilato autonomamente dall'intervistato stesso (come nel postale), senza la presenza quindi di un intervistatore. Gli intervistati accedono in modo autonomo alla Rete, dopo che è stato loro inviato l'indirizzo al quale collegarsi, e rispondono alle domande presenti nel questionario.

Questa modalità di rilevazione permette di ovviare anche ai problemi legati alla eccessiva libertà che potrebbe prendersi colui che compila un questionario cartaceo: la sequenza delle domande che compaiono sul video è infatti prefissata, e non c'è alcuna possibilità di mutarne l'ordine da parte dei rispondenti, garantendo in definitiva una adeguata *invarianza* degli stimoli cui essi sono sottoposti. Un ulteriore vantaggio è dato dal fatto che vengono definitivamente eliminati i tempi di registrazione su supporto informatico delle risposte fornite attraverso i questionari cartacei: nella *web-survey* le informazioni vengono infatti direttamente *salvate* in Rete nel momento della compilazione e sono immediatamente disponibili al ricercatore per le successive analisi dei risultati.

In generale, gli altri principali vantaggi del questionario auto-compilato sono:

- l'elevato risparmio sui costi, rispetto alle indagini dove è prevista la presenza di un intervistatore;
- la possibilità di raccogliere una quantità di informazioni nettamente superiore;
- una maggiore garanzia dell'anonimato, rispetto alle interviste personali, con una maggior facilità quindi di rispondere anche a domande più sensibili;
- la reale possibilità (almeno teorica) di copertura campionaria dell'intero territorio di riferimento.

Ma numerosi sono anche gli svantaggi, legati soprattutto all'altamente probabile auto-selezione dei rispondenti, che si articola in tre possibili componenti specifiche:

- una bassa percentuale di risposte;
- una elevata distorsione, rispetto all'universo, delle caratteristiche di chi risponde;
- un accesso privilegiato al mezzo utilizzato da parte della componente maggiormente istruita della popolazione.

Questi tre elementi analitici fanno parte di un unico problema di base che, eliminati o tenuti sotto controllo gli handicap metodologici, è probabilmente il vero limite della rilevazione demoscopica effettuata attraverso la raccolta online (simile peraltro alle antiche rilevazioni postali): l'effettiva impossibilità di gestire efficacemente il tasso, i tempi e la tipologia dei rispondenti. Dal momento che la buona riuscita del sondaggio è devoluta totalmente alla «buona volontà» di cooperare da parte degli individui campionati, il ricercatore può fare poco o nulla per eliminare i problemi che abbiamo elencato.

Generalmente, il tasso di risposta di un sondaggio online è giudicato soddisfacente se supera il 60-65% del campione; ma è ovvio che la quantità non è un elemento sufficiente, se non è accompagnato dalla qualità del sub-campione dei rispondenti, se cioè le distorsioni nelle caratteristiche campionarie risultano particolarmente elevate. In altri termini, sono da preferire tassi di risposta più ridotti ma equilibrati rispetto ad una situazione in cui, supponiamo, un *rientro* del 65% dei questionari sia dovuto quasi esclusivamente alla partecipazione al sondaggio della componente maschile del campione. D'altra parte, un tasso di risposta piuttosto basso è un'efficace spia di una possibile forte auto-selezione del campione di intervistati, causata o da difetti di impostazione del questionario (troppo lungo, complicato, noioso) ovvero da un disomogeneo interesse per le tematiche trattate (le indagini su temi esclusivamente politici, ad esempio, non scatenano in genere molto entusiasmo tra i rispondenti).

Nelle modalità di rilevazione in cui non è possibile tenere sotto controllo le caratteristiche campionarie, come è il caso dei questionari online, si fa generalmente ricorso alla tecnica della «ponderazione» statistica dei risultati. Essa consiste in un'operazione matematica mediante la quale si fanno pesare maggiormente gli individui appartenenti alle quote di popolazione «sotto-stimate» rispetto alle caratteristiche dell'universo e, viceversa, si fanno pesare meno quelli appartenenti alle quote «sovra-stimate».

Essendo questa una procedura che comunque altera di fatto i risultati, occorre avere a disposizione, per poterla utilizzare senza eccessive distorsioni, una numerosità minima anche delle quote sotto-stimate. È per questo motivo che, nelle *web-survey*, non è sufficiente che sia elevato il numero complessivo dei rispondenti, ma è necessario che sia elevato anche quello di ciascuna delle quote di ponderazione (che coincidono generalmente con i dati socio-anagrafici e, nel caso delle indagini politico-elettorali, il comportamento di voto alle ultime elezioni).

Come si è detto, le possibilità da parte del ricercatore di eliminare questi problemi sono minime, quasi inesistenti; non si può certamente obbligare chi non vuole rispondere a farlo, nemmeno con pressanti solleciti che corrono spesso il rischio di ottenere l'effetto opposto a quello desiderato. Per questo motivo, è buona norma aumentare in sede preliminare la numerosità campionaria.

L'utilizzo forse ottimale di questo strumento si accompagna alla costruzione di un «*panel*», cioè di una rete di intervistati che accettano di collaborare periodicamente con un istituto di ricerca o un dipartimento universitario, rispondendo a questionari, ad esempio, con cadenza trimestrale. Accanto ad alcuni svantaggi, legati alla possibile alterazione delle risposte in soggetti frequentemente interrogati, questa modalità di rilevazione offre indubbi vantaggi, tra cui: un tasso di risposta solitamente elevato (superiore spesso al 70%), il controllo costante del campione di intervistati, la maggiore *tranquillità* nel rispondere quando si è instaurato un buon rapporto fiduciario, la possibilità di controllare l'evoluzione nel tempo degli atteggiamenti o dei giudizi degli stessi individui.

Esistono già da anni numerosi software che permettono la realizzazione di *web-survey* gratuitamente (il più diffuso tra questi è «SurveyMonkey») con alcune limitazioni funzionali, sia nel numero di domande che nella numerosità campionaria, o a costi crescenti nei pacchetti più completi; se da un punto di vista tecnico il loro funzionamento non presenta particolari problemi, la reale criticità di questi strumenti risiede nella validità e nella generalizzabilità delle informazioni raccolte: il reclutamento dei rispondenti avviene molto spesso sulla base della disponibilità degli utenti a partecipare, dietro piccoli benefit, il che determina una ovvia distorsione dei risultati, che provengono da campioni

auto-selezionati (in generale più giovani e più scolarizzati) e pertanto non rappresentativi della popolazione di riferimento.

La diffusione delle *survey* online, su numerosi siti di informazione o di intrattenimento, risente delle limitazioni tipiche del tradizionale televoto; se a questi risultati ci si accosta in maniera corretta, prendendoli come una sorta di gioco virtuale, non occorrono specifiche contro-indicazioni. Diverso il discorso se l'obiettivo è invece quello di generalizzare i risultati ottenuti all'universo di riferimento: nessun tipo di campione auto-selezionato potrebbe qui esserci di aiuto. L'unica possibile strada è quella di effettuare una estrazione campionaria con metodi probabilistici, a partire da una popolazione nota a priori.

Ma anche in questo caso subentrano, ancor oggi in Italia, le possibili gravi distorsioni provocate dalla disomogenea possibilità di accesso al web, al contrario di quanto accadeva nelle indagini postali. La popolazione anziana, meno istruita, o residente in aree territoriali periferiche ha scarsa familiarità con la Rete, e i risultati di indagini di tipo «generalista», applicate quindi alla totalità della popolazione, rendono questo strumento relativamente attendibile per generalizzare le risposte ottenute tramite *web-survey* all'universo degli eleggibili. È pertanto opportuno integrare l'indagine su internet con una parallela indagine effettuata offline, e confrontare poi i differenti risultati.

Diverso il discorso nel caso, già oggi, la popolazione di riferimento sia formata da particolari categorie di individui che, per il loro lavoro o per specifici interessi, utilizzano frequentemente internet (come i docenti universitari, o gli abbonati a riviste online, o gli utenti di un sito ad accesso privato). In questi casi, la rilevazione Cawi è altamente consigliabile, rispetto ad altri metodi di raccolta delle informazioni, non presentando ulteriori svantaggi che non siano quelli derivanti da una programmazione non ottimale del questionario.

Il livello di informatizzazione del nostro paese è certo destinato a crescere velocemente; probabilmente il cosiddetto «digital divide» sarà superato quantomeno nel medio periodo, e la situazione potrà avvicinarsi a quella delle realtà più evolute dal punto di vista delle competenze digitali, come gli Stati Uniti o il Canada e, in Europa, come Francia o Uk. Come in quei paesi, anche da noi le indagini su internet diventeranno sicuramente il metodo più utilizzato, nel prossimo futuro.

4. I dati virtuali nella ricerca qualitativa: focus group online e blog

Diverse sono anche le novità introdotte dall'uso del web nella ricerca qualitativa, in particolare sui *focus group* e sul cosiddetto «blog».

Per quanto riguarda la prima tecnica, che consiste in una discussione di gruppo moderata generalmente da un ricercatore-psicologo o sociologo, l'utilizzo della Rete – tramite un collegamento audio-visivo con tutti i partecipanti – appare da un certo punto di vista più complesso, poiché le dinamiche interne al focus non possono venir controllate in maniera così efficace come quando tutti i soggetti sono riuniti intorno ad un tavolo; ma dal punto di vista logistico, il collegamento via web facilita in maniera decisiva la possibilità di reclutare persone anche a grande distanza, semplificando le modalità di accesso al focus e consentendo una maggiore numerosità dei gruppi che si possono prendere in considerazione per l'analisi delle tematiche oggetto di studio. Anche in questo settore, da anni esistono software che gestiscono la possibilità di effettuare focus online con costi contenuti.

Mentre il focus online ci fornisce i classici «dati virtuali», parzialmente differenti sono quelli che provengono dall'utilizzo dell'altra tecnica, il blog, una sorta di forum svolto in Rete che non ha una specifica derivazione dai tradizionali metodi offline. In questo caso i dati ottenuti si avvicinano a quelli «digitali», poiché nascono nel web ma, secondo la definizione che abbiamo utilizzato, differiscono da quelli realmente nativi in quanto non nascono spontaneamente, bensì vengono *provocati* dallo stimolo di un moderatore, in modo molto simile a quanto accade negli esperimenti di laboratorio propri della psicologia sociale.

Benché il termine blog sia ormai entrato in uso nel linguaggio del web per indicare una sorta di «diario in Rete», nell'approccio metodologico per blog si intende invece una tecnica che prevede una sorta di incontro virtuale tra 15-30 persone, selezionate preventivamente sulla base delle caratteristiche richieste dalla ricerca, che discutono di un tema dato per più giorni (di solito da 2 a 5). Il ruolo del moderatore è di porre le domande e di stimolare la discussione su aspetti specifici oggetto della ricerca. Il blog differisce dal forum online (che produce al contrario veri e propri «dati digitali» non provocati) per la funzione chiave esercitata dal moderatore, che redige e invia ai partecipanti in Rete gli spunti di discussione e li orienta verso le tematiche che interessano il committente.

I vantaggi del blog sono molteplici: consente di coinvolgere simultaneamente soggetti anche fisicamente molto distanti tra loro, il che rappresenta un risparmio, rispetto ai focus tradizionali, dal punto di vista organizzativo ed economico; inoltre, per il ricercatore è interessante e utile verificare come la stessa domanda-stimolo possa produrre risposte diverse in relazione, per esempio, alla collocazione geografica o all'ampiezza del luogo di residenza.

È particolarmente adatto quando l'oggetto della ricerca è complesso e presenta più aspetti che richiedono tempo per essere trattati: si possono ad

esempio raccogliere impressioni sulla linea politica di un partito, sui suoi leader, sullo stile di comunicazione, sui rapporti con altre forze politiche ecc.

È efficace quando i partecipanti sono esperti di una materia o hanno in comune tratti specifici, come l'appartenenza a un partito o un'associazione: in questo caso tutti conoscono l'argomento in discussione e hanno un linguaggio comune.

Infine, gli intervistati hanno tempo per riflettere sugli argomenti trattati e possono approfondire di più e meglio di quanto farebbero in un *focus group*, in cui le dinamiche relazionali possono inibire i più timidi a vantaggio dei più aggressivi.

Numerosi sono peraltro anche i limiti intrinseci a questa tecnica: chi scrive può talvolta autocensurarsi e, per dare di sé e del suo pensiero un'immagine accettabile, evitare di esprimersi come farebbe in una discussione diretta; va anche detto però che la frequentazione dei social (chi partecipa a un blog è ovviamente familiarizzato con la comunicazione online), con il ricorso a un linguaggio più espressivo/emotivo che analitico/riflessivo, limita questo rischio e non è infrequente che si risponda di getto a una domanda, come si farebbe a voce.

A proposito dell'affinità tra blog e social, accade spesso che l'interazione tra i partecipanti al forum sia scarsa: molti tendono ad affermare il loro punto di vista, trascurando la relazione con gli altri partecipanti. Da questo punto di vista il blog non sempre riesce a rappresentare realmente una discussione collettiva, configurandosi al contrario come la somma di più voci isolate.

Quando l'argomento proposto dal moderatore non è noto, si ha la possibilità di documentarsi prima di rispondere, falsando in qualche modo la qualità delle risposte (se non si conosce ad esempio un uomo politico, si cerca in internet la sua biografia e ci si attiene a quella piuttosto che dichiarare la propria ignoranza).

Un ultimo aspetto negativo, al contrario di quanto accade nei *focus group* online, riguarda il fatto che il moderatore non ha la possibilità di controllare gli aspetti non verbali della comunicazione e, di conseguenza, alcune risposte possono risultare neutre o scarsamente significative, dal momento che non vengono accompagnate dall'espressività del corpo e della voce.

5. Lo studio dei dati digitali

Benché alcuni dei principali strumenti per l'analisi dei dati digitali provengano dalle ricerche offline, il loro utilizzo in Rete prevede un tipo di approccio completamente differente, proprio per le caratteristiche dei dati stessi.

Qui di seguito una breve rassegna dei più rilevanti strumenti propri dello studio dei dati digitali (Natale e Airoidi 2017; Rogers 2009).

L'analisi del contenuto è una tecnica già utilizzata in passato per i testi (in particolare i giornali e le trasmissioni Tv), ma che con l'avvento di internet ha vissuto una nuova epifania, per due ordini di motivi: da una parte la sempre maggiore diffusione dei social media, che ha reso disponibile al ricercatore una mole infinita di nuovi contenuti testuali; dall'altra il fatto che ad ogni post è associabile una misurazione delle interazioni (*engagement*) che gli utenti hanno avuto con lo stesso (numero di like, commenti, condivisioni, retweet...). In questo modo, i contenuti disponibili sulle piattaforme social possono venir non soltanto classificati, ma anche messi in relazione con le relative metriche di engagement, analizzando dunque, oltre il puro contenuto testuale di un post/tweet, anche il tipo ed il livello di interazione contenuto-utenti (Schwartz e Ungar 2015).

La *sentiment analysis* è una delle più note tecniche che permettono al ricercatore di analizzare – in maniera più o meno automatizzata – collezioni di testi di grandi dimensioni: si tratta di un insieme di metodi computer-assisted volti all'estrazione del sentiment di contenuti testuali. L'obiettivo di questa tecnica è quello di capire, attraverso strumenti matematici e statistici, l'orientamento delle persone a partire dai testi da loro prodotti. In estrema sintesi, si può dire che la *sentiment analysis* ci permette di identificare la polarità dell'opinione (o emozione) espressa in un singolo testo – sia questa positiva, negativa o neutra – e, a livello aggregato, ci fornisce la possibilità di comprendere l'esistenza o meno di veri e propri trend di opinione (Mäntylä *et al.* 2017).

Anche l'analisi delle reti di interazione sociale (la *social network*) è antecedente all'avvento di internet, ma ha assunto nel linguaggio odierno un doppio significato: esso indica, da un lato, l'insieme di inter-relazioni tra attori sociali e, dall'altro, le piattaforme digitali attraverso cui queste reti sociali vengono gestite o ampliate (Boyd 2011). Un'ambiguità che si risolve nell'applicazione, sempre più diffusa, della tecnica della *network analysis* allo studio delle reti sociali online, che giunge alla possibile valutazione della struttura dei gruppi informali che si creano, identificandone gli attori chiave ed esplorandone le interconnessioni.

Pure nel caso dell'etnografia digitale appaiono evidenti i rimandi all'approccio maturato offline, dall'antropologia all'osservazione partecipante della ricerca etnografica (Boellstorff *et al.* 2012; Caliandro 2014). Applicando tale approccio alla specificità della Rete, le tecniche utilizzate divengono plurime, e si adattano ad una vasta gamma di analisi, sia qualitative che quantitative. L'etnografia digitale si propone dunque non come una specifica e definita tecnica di analisi, ma come un multiforme insieme di metodi che consentono di

studiare le dinamiche spontanee del web con un elevato livello di flessibilità, avvalendosi di un approccio *bottom-up*, in grado di conferire risalto non solo al contenuto delle conversazioni presenti online, ma anche alla struttura delle interazioni tra gli utenti. L'apparato teorico e metodologico dell'etnografia digitale prende solamente spunto dall'antropologia tradizionale, non limitandosi ad adattarla, procedendo invece a una sua integrazione con tecniche informatiche, dando così vita a metodi nati e pensati a partire dalla Rete stessa, dalla sua conformazione naturale e dalle sue dinamiche.

Il «*Machine learning*», tradotto in italiano come «apprendimento automatico» è infine un insieme di tecniche che, analizzando una mole complessa e non strutturata di dati, ha come obiettivo quello di fare previsioni, stime e classificazioni sui fenomeni naturali e sociali (Flach 2012). Più nello specifico, Machine Learning si preoccupa di definire dei modelli in grado di aumentare la precisione di predizione al crescere del numero di informazioni disponibili. Al contrario della ricerca socio-politica classica, questa procedura non ha come primario interesse l'identificazione dei rapporti causa-effetto dei fenomeni studiati, quanto la mera identificazione di modelli funzionanti nella previsione dei fenomeni stessi (Murphy 2012).

Come si è sottolineato, ciò che cambia in maniera strutturale, in tutti questi tipi di analisi dei dati digitali, è l'approccio di fondo, un approccio non riconducibile, e che anzi se ne distanzia in maniera evidente, alle due principali strade della ricerca, ai due tradizionali paradigmi che per decenni si sono prima *combattuti* per poi tentare una difficile *collaborazione*: il paradigma positivista e quello interpretativista. I cardini del loro approccio epistemologico e metodologico, nell'analisi di ciò che accade sul web, debbono almeno in parte venir dimenticati, in favore di una strumentazione del tutto particolare, e di logiche di studio parecchio differenti.

Vediamo brevemente come, e perché.

6. I due tradizionali paradigmi della ricerca

Nella storia della ricerca sociale, come noto due sono state prevalentemente le metodologie di analisi utilizzate, in diretta relazione alle due diverse (in parte contrapposte) modalità di conoscenza della realtà: il metodo quantitativo e quello qualitativo. Il primo ha come costante punto di riferimento il cosiddetto paradigma positivista, che ha come capostipite Emile Durkheim, il secondo il paradigma interpretativo, o della sociologia comprendente, legato a Max Weber. In estrema sintesi, la storica contrapposizione tra questi due para-

digmi ha a che vedere con un dibattito sempre presente nella ricerca teorica ed empirica, sul metodo cioè più appropriato per conoscere la realtà e i fenomeni sociali.

Nell'approccio quantitativo, le caratteristiche di fondo specifiche della scuola positivista (o post-positivista) possono essere così riassunte:

- esiste una realtà oggettiva, esterna all'interpretazione che ne danno gli individui;
- questa realtà è conoscibile attraverso una serie di procedure e metodi;
- lo studioso deve rimanere il più possibile distaccato dal suo oggetto di studio, per evitare manipolazioni o interferenze che modifichino l'oggetto stesso;
- l'obiettivo del ricercatore è quello di trovare delle leggi, sia pur provvisorie e soggette ad una possibile falsificazione, che governano i fenomeni sociali;
- il metodo utilizzato, derivato da quello scientifico delle scienze fisico-naturali, è caratterizzato da una decisa formalizzazione logico-matematica delle procedure e dalla standardizzazione del linguaggio, attraverso la costruzione del cosiddetto «linguaggio delle variabili» (Lazarsfeld 1967);
- l'analisi prevalente è quella «per variabili».

I punti salienti del filone che fa capo alla ricerca qualitativa sono invece i seguenti:

- non esiste una realtà oggettiva, ma tante realtà legate all'interpretazione che ne danno gli individui o i gruppi di individui;
- queste realtà sono conoscibili soltanto attraverso la comprensione dei significati che ne dà ogni individuo o gruppo;
- lo studioso, ai fini di una migliore comprensione, deve interagire costantemente con il suo oggetto di studio, per giungere ad una sorta di empatia con lo studiato;
- l'obiettivo del ricercatore è quello di comprendere a fondo il comportamento individuale o dei gruppi che sta studiando, senza bisogno di trovare leggi perenni, ma soltanto «enunciati di possibilità» o «tipi ideali» (Weber 1958);
- il metodo utilizzato, derivato dagli studi psicanalitici e psicologici, è caratterizzato da una costante interazione con il soggetto, senza alcuna specifica formalizzazione pre-costituita, ma facendo uso degli strumenti che volta per volta si prestano maggiormente alla scoperta di nuove realtà;
- l'analisi prevalente è quella «per soggetti».

Dal punto di vista della ricerca operativa, la differenza più radicale tra le due scuole di pensiero consiste proprio nella distinzione tra analisi «per variabili» e analisi «per soggetti», due approcci totalmente inapplicabili nell'analisi dei dati digitali. Cosa significa e quali sono le conseguenze empiriche di questa distinzione?

Con analisi per variabili intendiamo esattamente il procedimento utilizzato nell'approccio ai risultati di una *survey*, sia offline che online (con i dati virtuali): ogni individuo intervistato non viene preso in considerazione nella sua interezza, ma soltanto per la risposta che egli ha fornito ad ogni singola domanda (la «variabile») che gli è stata posta. Come se il singolo soggetto venisse sezionato, scomposto analiticamente sulla base di una serie di attributi, proprietà, atteggiamenti, comportamenti, ecc., e a questi ricondotto.

Totalmente differente è l'approccio della cosiddetta analisi per soggetti, caratteristico delle tecniche qualitative: in questo caso l'obiettivo principale del ricercatore è proprio quello di cercare di rilevare l'interezza (o almeno quella parte più funzionale all'obiettivo dell'indagine) degli individui analizzati. Allo scopo di pervenire alla comprensione del significato attribuito da ogni singolo soggetto alla propria azione dotata di senso, al proprio atteggiamento, alla propria scala dei valori.

7. La ricerca sul web: una terza via

Se questi sono dunque i due tradizionali approcci alla ricerca in generale, e a quella socio-politica in particolare, è evidente come quella sul web si situi in un ambito non riconducibile né al primo né al secondo dei paradigmi qui brevemente descritti.

Quando si analizzano i «dati digitali» abbiamo ben presente che gli utenti della Rete sono individui che non rappresentano i consueti «oggetti» della ricerca sociale o politica di stampo positivista, non vengono interrogati per sapere cosa pensano, come votano, in cosa credono, come accade invece nelle indagini demoscopiche standard. Da questo punto di vista, non rappresentano nemmeno campioni significativi di popolazione, per cui i risultati provenienti dall'analisi della Rete non possono in alcun modo venir paragonati alle informazioni provenienti, ad esempio, da una *survey* effettuata con tecniche di estrazione probabilistiche. Le loro caratteristiche individuali non sono pertanto riconducibili alle cosiddette «variabili» della ricerca quantitativa.

Nemmeno l'altro approccio, quello qualitativo, può essere paragonato allo studio della Rete benché, in qualche modo, sia a questo maggiormente

prossimo, essendo centrato non sulle variabili ma sugli individui analizzati. Il profilo dell'utente, una sorta di riassunto di tutte le tracce lasciate dall'utilizzatore di internet, si avvicina infatti per certi versi all'idea olistica che demarca il paradigma interpretativo: il soggetto può essere compreso soltanto nella sua completezza, e non è riconducibile alla somma delle sue parti.

Ma è difficile cogliere l'interesse di un individuo attraverso le tracce lasciate in Rete, prima di tutto perché le sue informazioni, le sue impressioni, le sue opinioni sono certamente parziali e frammentate, riguardando unicamente i temi che di volta in volta divengono motivo di discussione; in secondo luogo, e soprattutto, perché risentono costantemente della volontà dell'attore sociale di *presentarsi* agli altri secondo modalità scelte da lui stesso, per interagire con loro, al fine di assumere – consciamente o inconsciamente – una sua specifica personalità, un suo personale ruolo, all'interno del frame di internet: il contestatore, il moderato, il pacifista, il cosiddetto *hater*, colui che insulta costantemente tutto e tutti, e così via.

Un comportamento che rimanda al noto tema della «desiderabilità sociale», la distorsione nelle proprie dichiarazioni provocata dalla volontà dei soggetti di apparire a volte diversi da ciò che si è, maggiormente conformi alle aspettative di ruolo normate dalla società. Nel caso della Rete, il problema può diventare ancora più complesso, poiché si possono sviluppare, e di fatto si sviluppano, attitudini sia «conformiste» (tipiche delle ricerche offline) che «anti-conformiste», più osservabili proprio negli interventi su internet, laddove gli utenti, protetti da un certo grado di anonimato, accentuano la propria alterità nei confronti del pensiero comune e del cosiddetto *politically correct*, lasciandosi andare a volte a pesanti dosi di insulti che sarebbero impensabili nel caso di interviste telefoniche, o *face-to-face*, o nei *focus group*.

Ma ciò che porta alla conclusione di essere in presenza di una vera e propria «terza via» della ricerca è la genesi dei dati che si hanno a disposizione, che non sono il frutto né del classico cortocircuito stimolo-risposta, caratteristico delle ricerche quantitative, né dell'interazione osservatore-osservato tipica delle ricerche qualitative. I dati digitali sono auto-prodotti dagli attori sociali, in maniera del tutto *naturalistica*, senza alcun livello di intrusività degli strumenti utilizzati per studiarli, secondo modalità molto simili appunto alla classica osservazione naturalistica, dove non c'è alcun contatto tra ricercatore e oggetto dell'analisi, e dove l'osservato ignora di essere oggetto di osservazione (come nel caso dell'antropologia del punto-vendita e della cosiddetta *shopping experience* applicata alla ricerca sociale).

A differenza di quest'ultima tecnica di analisi, però, gli attori sociali non agiscono in maniera realmente *naturale* sulle piattaforme di internet, poiché potrebbero tendere a mostrare solamente una faccia di sé, quella forse giudicata

più interessante da condividere con la propria popolazione di riferimento, i propri amici o i propri *followers*. È ovvio che nessuno di noi è esente, nella vita quotidiana, dai condizionamenti del frame in cui volta per volta si trova inserito, ma nell'analisi caratteristica della ricerca qualitativa, compito del ricercatore è proprio quello di ricostruire l'interezza dell'individuo a partire dalla sua *innaturale* naturalezza. Gli utenti dei *social network* non sono invece osservabili, e non potremo mai conoscere il grado di naturalezza dei loro interventi.

Questa incertezza di fondo potrebbe rendere i risultati delle analisi dei dati digitali difficilmente generalizzabili, non solo (ovviamente) alla popolazione in senso lato, ma nemmeno alla più ristretta popolazione di riferimento, quella degli utenti internet. O meglio, nelle tecniche di analisi potremo correttamente riferirci all'ambiente web, a livello aggregato, ma non alle persone che lo popolano, che potrebbero essere differenti da come appaiono nelle loro manifestazioni.

Nella ricerca il cui oggetto di studio sono i dati digitali nel web, le caratteristiche di riferimento non sono dunque accomunabili, se non in maniera molto parziale, a quelle precedentemente elencate per i due paradigmi tradizionali, e possono essere così riassunte:

- esiste sia una realtà oggettiva (i dati digitali, le tracce presenti in Rete) che una realtà soggettiva, poiché i dati sono generati a partire dalle scelte degli utenti, dalla loro volontà di partecipare attivamente alla vita virtuale;
- questa realtà è analizzabile attraverso una serie di procedure e metodi, che non possono però prescindere dalla comprensione dei significati dati da ogni individuo o gruppo;
- il rapporto tra studioso e oggetto di studio è inesistente, non esiste alcuna manipolazione o interferenza che possa modificare l'oggetto stesso;
- il rischio derivante dalla «desiderabilità sociale» è costantemente presente, e non può in alcun modo essere tenuto sotto controllo dal ricercatore;
- il grado di generalizzabilità dei risultati alla totalità della popolazione, ma anche alla popolazione web, non è mai controllabile, se non attraverso misurazioni esterne o per alcuni casi specifici, come la misurazione delle interazioni dei singoli utenti (il cosiddetto *engagement*);
- l'analisi prevalente è «per contesti» (i dati aggregati), prossima a quella «per soggetti» della ricerca qualitativa, ma con valenze olistiche (l'interezza del sé) quasi inesistenti.

TAB. 1. *Le tre vie per la ricerca sul web. Una proposta*

	Ricerca quantitativa: post-positivismo	Ricerca qualitativa: interpretativismo	Ricerca sul web: la terza via
Ontologia (essenza) Discorso sull'essere	Realtà oggettiva	Realtà soggettiva	Realtà oggettiva (dati digitali) e realtà soggettiva (scelte degli utenti)
Epistemologia (conoscenza) Discorso sulla conoscenza	Procedure e metodi	Comprensione dei significati degli attori sociali	Procedure e metodi, e comprensione
	Distacco	Interazione	Rapporto inesistente
Metodologia (metodo) Discorso sul metodo	Leggi di medio raggio	Tipi ideali	Generalizzabilità: ?
	Formalizzazione (linguaggio delle variabili)	Interazione e interpretazione	Formalizzazione (rischio desiderabilità sociale)
	Tecniche quantitative	Tecniche qualitative	Tecniche miste
	Analisi per variabili	Analisi per soggetti	Analisi per contesti

Fonte: elaborazione dell'autore.

La pratica di ricerca sui dati digitali, per tutto quanto finora sottolineato, si configura realmente come una sorta di «terza via» alla ricerca, e necessita pertanto di un approccio e di tecniche nuove, mai sperimentate prima, che hanno quindi bisogno di costanti aggiornamenti e miglioramenti man mano che si affinano gli algoritmi e i metodi di analisi.

Accanto al paradigma positivista e quello della sociologia comprendente, che hanno dato forma e contenuto alla ricerca quantitativa e a quella qualitativa, in questo ultimo ventennio una terza via si è aperta nel campo della ricerca, lo studio dei dati digitali, che non è riconducibile ai due tradizionali modi di accostarsi allo studio della società, ma che a quelli si affianca, per darci informazioni e risultati diversi dai precedenti, ma non certo meno significativi e meno interessanti.

Certo, è soltanto un adolescente, rispetto ai giganti del passato, che sono stati oggetto di costanti verifiche e confronti per validare i risultati provenienti da quegli studi. È un adolescente ancora incerto, con capacità euristiche ancora limitate, che procede per tentativi, errori e modifiche del proprio percorso. Ma è un adolescente straordinariamente ricco di informazioni e di dati, un archi-

vio praticamente illimitato su cui svolgere le proprie analisi (Calise e Musella 2019). Dove peraltro il dato oggettivo, se vogliamo paragonarlo alla ricerca quantitativa, è ancora troppo legato alle modalità di esecuzione e alla discrezionalità del ricercatore; e la capacità di comprensione dei soggetti che si studiano, se paragonata alla ricerca qualitativa, è soltanto parzialmente olistica, non riesce ad arrivare all'interezza degli individui, ne riflette forse solamente alcuni frammenti, quelli presenti sulla Rete.

Ma, d'altra parte, è proprio la Rete l'oggetto principale di studio di questa terza via. Sono proprio i frammenti, le tracce che gli utenti lasciano in maniera spontanea sul web, senza essere stimolati né interrogati, a costruire il corpo di indagine della ricerca politica e sociale del nuovo millennio. Un corpo che prima non esisteva, così come non esisteva sessant'anni fa il pubblico televisivo. Ed è un corpo che deve essere studiato, nelle forme e nei modi inediti, proprio perché sta diventando essenziale, per la ricerca politica, comprenderne i confini, gli elementi salienti, le interazioni che si creano a diversi livelli della vita quotidiana. *Tweet, post, like, engagement, big data*: tutte espressioni *native* della quotidianità contemporanea, materiali empirici e approcci analitici che ci permettono di comprendere cose che altrimenti non sapremmo mai, interazioni che avvengono sulla Rete, e vengono replicate solamente in parte nella vita *offline* – sempre che questa distinzione continui ad avere senso.

La vita politica, la vita sociale, la vita relazionale hanno avuto un deciso incremento di salienza attraverso l'attività svolta sul web; e gli strumenti della ricerca devono adeguarsi a questo cambiamento epocale, affinando le tecniche – oggi ancora certamente aleatorie, in molti casi – per poter produrre conoscenze inedite sui fenomeni politici e sociali contemporanei. La strada intrapresa è ardua, i risultati ottenuti sono forse tuttora incerti, ma per molti versi è proprio questo il futuro della ricerca.

Riferimenti bibliografici

- AIROLDI, M. (2017a), *I metodi digitali. Fare ricerca tra online e offline, qualitativo e quantitativo*, in M. CARDANO e F. ORTALDA (a cura di), *Metodologia della ricerca psicosociale*, Torino, UTET Università, pp. 329-343.
- AIROLDI, M. (2017b), *Oltre il virtuale. Studiare la società attraverso i dati digitali*, in P. NATALE e M. AIROLDI (a cura di), *Web e social media*, Santarcangelo di Romagna, Maggioli Editore, pp. 1-20.
- BOELLSTORFF, T., NARDI, B., PEARCE, C. e TAYLOR, T.L. (2012), *Ethnography and Virtual Worlds: A Handbook of Method*, Princeton, Princeton University Press.

- BOYD, D. (2011), *Social Networking Sites as Networked Publics: Affordances, Dynamics and Implication*, in Z. PAPACHARISSI (a cura di), *A Networked Self: Identity Community and Culture on Social Network Sites* London, Routledge, pp. 39–58.
- CALIANDRO, A. (2014), *Ethnography in Digital Spaces: Ethnography of Virtual Worlds, Netnography, & Digital Ethnography*, in R. DENNY e P. SUNDERLAND (a cura di), *Handbook of Anthropology in Business*, Walnut Creek, Left Coast Press, pp. 738-761.
- CALISE, M. e MUSELLA, F. (2019), *Il principe digitale*, Roma-Bari, Laterza.
- FLACH, P. (2012), *Machine Learning. The Art and Science of Algorithms that Make Sense of Data*, Cambridge, Cambridge University Press.
- FLORIDI, L. (2014), *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*, Oxford, Oxford University Press.
- LAZARSFELD, P. (1967), *Metodologia e ricerca sociologica*, Bologna, Il Mulino.
- MAYER-SCHOENBERGER, V. e KENNETH, C. (2013), *A Revolution that will Transform How we Live, Work and Think*, Boston-New York, Houghton Mifflin Harcourt.
- MÄNTYLÄ, M.V., GRAZIOTIN, D. e KUUTILA, M. (2017), *The Evolution of Sentiment Analysis. A Review of Research Topics, Venues, and Top Cited Papers*, in «Computer Science Review», 27, pp. 16-32.
- MURPHY, K.P. (2012), *Machine Learning. A Probabilistic Perspective*, Cambridge, MIT Press.
- NATALE, P. (2004), *Il sondaggio*, Roma-Bari, Laterza.
- NATALE, P. (2009), *Attenti al sondaggio!*, Roma-Bari, Laterza.
- NATALE, P. e AIROLDI, M. (2017) (a cura di), *Web e social media*, Santarcangelo di Romagna, Maggioli.
- NEGROPONTE, N. (1995), *Being Digital*, New York, Alfred A. Knopf.
- ROGERS, R. (2009), *The End of the Virtual: Digital Methods*, Amsterdam, Amsterdam University Press.
- ROGERS, R. (2015). *Digital Methods for Web Research*, in R.A. SCOTT e S.M. KOSSELYN (a cura di), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, New York, John Wiley & Sons, <https://doi.org/10.1002/9781118900772.etrds0076>.
- SCHWARTZ, H.A., e UNGAR, L.H. (2015), *Data-driven Content Analysis of Social Media: A Systematic Overview of Automated Methods*, in «The Annals of the American Academy of Political and Social Science», 659(1), pp. 78-94.
- WEBER, M. (1958), *Il metodo delle scienze storico-sociali*, Torino, Einaudi.

