

La centralità dei dati per un welfare che cambia: una sfida per la ricerca sociale

Fabrizio Martire

RPS

Il cosiddetto welfare digitale dipende fortemente dai dati. Quale che sia la funzione loro attribuita – anticipare, stimare, localizzare fabbisogni o sostenere in modo più o meno automatizzato processi decisionali – la rinnovata centralità dei dati per il welfare, legata anche alle innovazioni tecnologiche che rendono possibili forme di rilevazione ed elaborazione delle informazioni inimmaginabili in passato, implica anche un'altra centralità: quella dei cittadini, principali stakeholders dei servizi di welfare e ora, nella cosiddetta «società digitale», anche fornitori insostituibili dei dati

e delle informazioni di cui il welfare si serve. Tutto ciò rende l'etica della ricerca un campo strategico di riflessione e di intervento, non solo di per sé, ma anche fattore che influenza la qualità dei dati raccolti. Nel saggio si sostiene la tesi secondo cui l'etica dei mezzi – approccio deontologico, rispetto della persona e della riservatezza dei suoi dati – debba essere più sistematicamente affiancata da un'etica dei fini – approccio consequenzialista, condivisione con i cittadini delle beneficialità pubbliche, estese e democratiche delle campagne di raccolta dati.

1. Società digitale e big data

La cosiddetta «società digitale», annunciata ormai da tempo (Lupton, 2015), si sta pienamente affermando anche come società delle informazioni e dei dati digitali. La sinergia tra le tecnologie della nostra vita quotidiana (computer, tablet, smartphone, ecc.) e quelle a disposizione dei grandi gestori di dati (istituzioni pubbliche e private, imprese, ecc.) implica, almeno in potenza, un salto di qualità nella rilevazione, costruzione, gestione e utilizzo dei dati. Questo nuovo scenario investe profondamente anche le pratiche di welfare, le quali – oltre che di sensibilità, etica, competenze, esperienze, modelli teorici di riferimento – vivono di dati.

Il welfare digitale è caratterizzato da profonde innovazioni potenziali nella progettazione e implementazione dei processi attraverso cui conoscere e anticipare i bisogni delle persone (vedi di seguito, par. 2), de-

cidere quali soluzioni di welfare adottare caso per caso e valutare l'efficacia degli interventi attuati (Pedersen e Wilkinson, 2018). La progressiva introduzione dell'intelligenza artificiale e degli algoritmi decisionali (vedi seguito, par. 3) implica una maggiore dipendenza delle pratiche di welfare dalla qualità dei dati su cui tali meccanismi si basano; come vedremo nella seconda parte del saggio, questo tema tipicamente metodologico – la qualità dei dati – implica una prospettiva di riflessione e azione più ampia, che metta al centro i cittadini nel loro doppio ruolo (fornitori di dati/informazioni e destinatari degli interventi di welfare), e guardi al welfare digitale come un sistema collaborativo in cui la reputazione delle istituzioni e la fiducia in esse da parte dei cittadini costituiscono due ingredienti decisivi.

L'espressione «società digitale» ne chiama, per immediata associazione di idee, un'altra ormai entrata nel *thesaurus* dei nostri giorni: «big data». L'uso ampio e diversificato di questa espressione è proporzionale alla difficoltà di darne una definizione univoca. A questo riguardo mi pare molto efficace la sintesi proposta da Salganik in un manuale di metodologia della ricerca sociale digitale di recentissima pubblicazione: «Una delle definizioni più comuni [di *big data*] pone l'attenzione sulle “3V”: *volume*, *varietà*, *velocità*. In sostanza esiste una grande quantità di *big data*, in una grande varietà di formati, in un processo di creazione continuo. I fan dei *big data* aggiungono anche altre “V”, come *veridicità* e *valore*, mentre i critici ne aggiungono di diverse, come *vaghezza* e *vacuità*» (Salganik, 2020, p. 30).

C'è un'altra questione metodologica che l'era digitale, o i *big data*, hanno posto, o meglio ri-attualizzato; e cioè la differenza tra dati/informazioni costruiti appositamente per gli scopi per i quali vengono usati e dati/informazioni che vengono usati per scopi indipendenti rispetto ai processi che ne hanno determinato la creazione. A questo proposito Salganik propone la coppia concettuale dati *ready-made* e *custom-made* (ivi, p. 24); nella prima categoria rientrano le informazioni che il ricercatore/utilizzatore di dati trova pronte (*ready*) e usa per i suoi obiettivi conoscitivi o di intervento; nella seconda le informazioni che ottiene grazie a un processo di rilevazione disegnato (*custom*) e implementato per le sue esigenze.

Questa dicotomia concettuale e le sue implicazioni metodologiche non costituiscono un'assoluta novità. Anche nell'era pre-digitale i ricercatori potevano ricorrere a dati tipicamente *custom* – si pensi alle indagine campionarie con le quali il ricercatore «crea» informazioni, cioè le risposte delle persone alle domande di un questionario appositamente dise-

gnato, o *ready* – l'analisi secondaria di fonti, documenti, testi e altri tipi di artefatti umani creati per motivi diversi rispetto alle loro esigenze conoscitive. D'altra parte, i *big data* – e questo per alcuni costituisce il vero elemento di innovazione – hanno enormemente ampliato, almeno in potenza, l'accesso a flussi di dati *ready*, che possono essere sfruttati per esigenze conoscitive.

Tuttavia, l'entrata – o forse è meglio dire l'irruzione – del digitale nella ricerca sociale non significa solo potenziale accesso ad un nuovo mondo da osservare fatto di comunicazioni e interazioni sui *social networks*, tracce di transazioni finanziarie, dati di posizioni e spostamenti rilevati con sensori Gps (giusto per fare qualche esempio di tipici dati *ready-made* digitali). Per quanto differenziata per categorie di popolazione, la penetrazione della tecnologia nella vita quotidiana amplifica le possibilità e le modalità a disposizione del ricercatore per chiedere informazioni mirate (*custom-made*) ai loro target di riferimento. Si pensi ad esempio alle potenzialità offerte dalla tecnologia mobile; essa consente di «intercettare i rispondenti il più possibile a ridosso del comportamento che si vuole studiare, per limitare lo sforzo di cooperazione mnemonica e migliorare la precisione dei dati raccolti» (Schmitz, 2018, p. 24). I questionari somministrati tramite smartphone e app permettono, in linea di principio, di raccogliere informazioni in tempo reale, eliminando tutte quelle fonti di distorsione legate alla eventuale natura retrospettiva dei dati (che caratterizzano invece le survey tradizionali, in cui generalmente si chiedono agli individui informazioni sui loro comportamenti avvenuti anche molto tempo prima del momento della rilevazione).

Lungi dal considerare le innovazioni sopra sommariamente descritte come la panacea di tutti i mali; al contrario, e come vedremo nelle pagine che seguono, esse rinnovano problemi metodologici tradizionali e ne introducono di nuovi. Ciò non toglie che lo scenario descritto ha implicazioni rilevanti anche per l'uso dei dati nelle pratiche di welfare.

2. *Dati per il welfare*

Un ambito primario di rilevanza dei dati per il welfare riguarda la stima, e, quando possibile, l'anticipazione, di bisogni sociali in risposta ai quali progettare e implementare policy, interventi e prestazioni, tanto a livello individuale, quanto territoriale; parallelamente alcune delle principali innovazioni di processo e di strategia di ricerca sociale si sono in-

crociate, nel momento della loro applicazione, con il welfare e con le sue esigenze di conoscenza. A questo proposito va comunque sottolineata una rilevante differenza interna al mondo, ampio e diversificato, del welfare. Se in riferimento alla salute pubblica sono oggi a disposizione sistemi informativi adeguati a supporto della programmazione delle policy, riguardo agli altri ambiti principali del welfare i margini di miglioramento sono molto ampi, non potendo ancora contare su strategie di mappatura dei bisogni diffuse e sistematiche.

Il progetto *Google Flu Trends* (Ginsberg e al., 2009), ormai diventato un classico nella letteratura metodologica sui *big data*, costituisce un esempio significativo di uso dei dati digitali al servizio della salute pubblica¹. L'obiettivo del progetto era migliorare e rendere più sensibile il monitoraggio della diffusione dell'influenza stagionale. Tale misurazione negli Stati Uniti era affidata ai Centri di controllo e prevenzione delle malattie che raccoglievano ed elaboravano informazioni ricevute da presidi medici locali selezionati sul territorio. Nel progetto *Google Flu Trends* questi dati – ufficiali, controllati, ma giudicati lenti rispetto alla tempestività richiesta ad un sistema di monitoraggio epidemiologico – sono stati affiancati dall'analisi geo-localizzata delle ricerche che le persone fanno sui motori di ricerca digitando espressioni come «influenza», «rimedi per l'influenza», ecc. – cioè dati sicuramente meno precisi ma più veloci per ricostruire il passaggio dell'influenza stagionale in un territorio. Dal punto di vista degli ideatori di *Google Flu Trends*, l'analisi delle *queries* sui motori di ricerca poteva anche servire per «tracciare» i casi di influenza con sintomi lievi, cioè le situazioni in cui le persone non vanno dal medico, sfuggendo così al monitoraggio tradizionale.

Non è difficile immaginare la trasposizione dell'approccio *Google Flu Trends* ad altri domini del welfare. Vari tipi di tracce che le persone lasciano online (ad esempio *queries* di ricerca, interazione sui *social networks*, acquisti, richieste di informazioni su siti istituzionali) possono essere analizzate per cogliere nel loro stato embrionale, e cioè prima che si conclamino in forme codificate e ufficiali, situazioni di disagio

¹ Col senso di poi possiamo dire che il valore di questo progetto risiede più nelle prospettive da esso aperte che nella sua efficacia intrinseca. I suoi risultati si sono rivelati nel tempo non proprio soddisfacenti a causa di alcuni problemi di funzionamento che, per motivi di spazio, qui non posso descrivere, ma che sono ben illustrati in (Lazer e al., 2014). Al di là di questi limiti specifici, l'approccio di *Google Flu Trends* nei suoi connotati generali è stato applicato nel tempo in ambiti di ricerca anche molto diversi tra loro.

familiare, economico-lavorativo, psichico, ecc., con l'obiettivo di individuare target di popolazione o territoriali potenzialmente a rischio rispetto ai quali prepararsi ad interventi di welfare mirati (Cao, 2012; Pedersen e Wilkinson, 2018).

Riprendendo la dicotomia concettuale illustrata nel primo paragrafo, quella che distingue tra dati *ready-made* e *custom-made*, possiamo dire che lo scenario sopra descritto si riferisce alle potenzialità del primo tipo di dati per il welfare. Anche sull'altro fronte, quello del *custom-made*, si stanno affermando, soprattutto nel contesto internazionale, strategie innovative di ricerca che aprono scenari nuovi per le politiche di welfare, con particolare riferimento al tema della *ageing society*. Mi sto riferendo alle cosiddette *biosocial surveys*, cioè indagini campionarie – solitamente di tipo longitudinale e svolte su campioni molto grandi – in cui, oltre alle variabili e agli aspetti tipicamente studiati con le survey (caratteristiche sociodemografiche, opinioni, atteggiamenti, stili di vita quotidiana, ecc.), si rilevano anche alcuni *biomarkers*, cioè dati fisici e biologici (altezza, peso, caratteristiche dell'apparato respiratorio, pressione arteriosa, colesterolo, e così via) attraverso i quali valutare e monitorare nel tempo le condizioni di salute degli intervistati (Sala, 2013). Nello specifico, lo scopo principale delle *biosocial surveys* è studiare le relazioni a livello individuale tra tipi di variabili molto diverse tra loro (parametri dello stato di salute effettivo, condizione di salute percepita, stili di vita e di lavoro, ecc.) per analizzare le influenze reciproche tra gli aspetti fisiologici, psicologici e sociali nella vita quotidiana.

In anni recenti, mappando il Dna degli intervistati da campioni di saliva o sangue, l'ambito di rilevazione delle *biosocial surveys* si è esteso alla genetica, campo di studi controverso, soprattutto quando viene messo in relazione – come nel caso delle *biosocial surveys* – con aspetti psicologici, stili e qualità della vita. Ferme restando le implicazioni etiche legate all'apertura alla genetica, nelle *biosocial surveys* «da combinazione tra due approcci consolidati – la survey da un lato e la misurazione di caratteristiche tipicamente associate allo stato di salute dall'altro – crea un ibrido originale funzionale soprattutto allo studio su larga scala delle relazioni tra predisposizioni individuali (anche di natura genetica), stili di vita, condizioni di salute e benessere» (Martire e Pitrone, 2018, p. 61). Come intuibile già da questa breve descrizione, le *biosocial surveys* implicano uno sforzo progettuale e di implementazione molto gravoso; la loro natura longitudinale e la rilevazione congiunta di aspetti psicologici, sociali e *biomarkers* – che richiede il ricorso a competenze mediche in fase di raccolta dei dati oltre a quelle solitamente richieste agli inter-

vistatori nei sondaggi – rendono le *biosocial surveys* molto costose, e generalmente dipendenti da investimenti pubblici significativi in ricerca e sviluppo. A livello internazionale si segnalano attualmente pochissimi progetti di *biosocial surveys* attivi e due di questi sono strettamente legati alle politiche di welfare, con particolare riferimento, come accennato in precedenza, al problema dell'invecchiamento della popolazione.

Negli Stati Uniti il *National Institute on Aging* coordina – con il sostegno, anche finanziario, di altri enti governativi – il progetto di ricerca longitudinale *The Health and Retirement Study* (Hrs); lo scopo del progetto è fornire dati e analisi utili ai responsabili delle politiche sulla salute, sul lavoro e sul welfare, e più in generale alimentare un dibattito pubblico sull'invecchiamento della popolazione che sia informato da un approccio integrato in cui gli aspetti sanitari e di benessere sociale siano trattati nelle loro relazioni reciproche. Dal 2012 il progetto Hrs include anche la rilevazione di informazione genetiche (Sonnega e al., 2014). Nel 2001 prende il via il progetto *Survey on Health, Ageing and Retirement in Europe* (Share)², con l'esplicita intenzione di adattare il progetto Hrs al contesto europeo, e contribuire così alla costruzione di un sistema informativo trans-nazionale sui temi dell'invecchiamento³.

3. Dati di welfare

I dati, le strategie per la loro raccolta, elaborazione ed interpretazione, non sono centrali solo nella fase di anticipazione e stima di bisogni sociali. Oggi più che mai un intervento di welfare può essere rappresentato come un sistema decisionale complesso di allocazione di risorse che si alimenta continuamente di dati per il suo funzionamento.

Nella letteratura specializzata, per rappresentare ed analizzare i processi decisionali implicati dagli interventi di welfare si fa spesso riferimento al concetto di *Decision Support System* (Dss) (Power, 2008). L'idea di base consiste nell'introdurre in un processo decisionale (ad es., stabilire se una famiglia debba o meno essere oggetto di una misura di welfare, oppure individuare modalità e intensità di tale misura, ecc.) una qualche forma standardizzata di raccolta ed elaborazione di informazioni, la cui logica prescinde dalle specificità dei singoli casi oggetto di deci-

² <http://www.share-project.org/home0.html>.

³ Per ulteriori testimonianze di progetti di biosocial survey vedi: Benzeval e al., 2014; <https://www.understandingsociety.ac.uk/>; <http://www.cuore.iss.it/>.

sione, in questo senso standardizzata, e che assume un ruolo più o meno vincolante per chi ha la responsabilità della decisione.

Nell'ambito del welfare i Dss possono assumere forme molto diverse, generalmente collocabili su due dimensioni: il grado di strutturazione delle informazioni raccolte e, come accennato sopra, il grado di automatismo previsto dal sistema decisionale complessivo. La strategia di raccolta dei dati può essere altamente strutturata (questionari standardizzati, schede di rilevazione che prevedono a monte tutte le casistiche ammissibili), oppure prevedere una qualche combinazione tra dati strutturati e non, contemplando la possibilità di rilevare informazioni più biografiche e situazionali, che possono variare caso per caso (ad es. integrare questionari e schede compilate da una famiglia con le valutazioni spontanee e non standardizzate che può fare un operatore quando entra in contatto con una specifica situazione di disagio effettivo o potenziale).

In un Dss gli output di un sistema informativo possono variare in funzione del ruolo che essi assumono rispetto al processo decisionale. A grandi linee si può dire che si va da un grado zero di automatismo in cui le informazioni raccolte ed elaborate hanno l'obiettivo di informare al meglio il decisore, la cui autonomia è pienamente preservata rispetto ad esse; a gradi intermedi in cui i dati elaborati individuano un range limitato di alternative decisionali possibili, lasciando comunque la responsabilità della scelta finale al decisore; fino a Dss normativi, in cui l'output del sistema informativo è la decisione stessa. In quest'ultimo modello, altamente standardizzato, la responsabilità dei gestori dell'intervento di welfare non si esercita a valle del processo di raccolta ed elaborazione delle informazioni, e cioè nel loro uso caso per caso, quanto piuttosto a monte, nella progettazione e implementazione degli algoritmi che definiscono una volta per tutte le regole di trasformazione di una data combinazione di informazioni in una decisione (Keen, 1987; Schneider e Seelmeyer, 2019).

Quale che sia la funzione attribuita ai dati – anticipare, stimare, localizzare fabbisogni o sostenere in modo più o meno automatizzato processi decisionali – la loro rinnovata centralità per il welfare, legata anche alle innovazioni tecnologiche che rendono possibili forme di rilevazione ed elaborazione delle informazioni inimmaginabili in passato, implica anche un'altra centralità: quella dei cittadini, principali stakeholder dei servizi di welfare e ora, nella cosiddetta «società digitale», anche fornitori insostituibili dei dati e delle informazioni di cui il welfare si serve. Alle principali implicazioni di questa considerazione saranno dedicate le pagine che seguono.

4. La ricerca sociale come partecipazione

Le specificità della ricerca sociale rispetto ai modi dell'indagine empirica propri delle cosiddette «scienze dure» (fisica, biologia, chimica, ecc.) sono molteplici e riguardano, tra l'altro, caratteristiche dell'oggetto di studio, logiche e strumenti di rilevazione, potenzialità della formalizzazione e della matematizzazione dei processi e delle conoscenze acquisite. Tali differenze appaiono così significative e irriducibili da mettere in crisi qualsiasi tesi monista in ambito epistemologico, e cioè l'idea per cui si possa e si debba arrivare alla definizione e applicazione di una logica e metodologia dell'indagine scientifica comune a tutte le discipline (Marradi, 2016).

Una di queste differenze irriducibili è a mio avviso ben rappresentata dall'idea secondo cui la ricerca sociale, quando si basa su dati individuali, è sempre e comunque partecipazione. I referenti empirici delle scienze dure (atomi, molecole, strati geologici, corpi celesti, ecc.) possono essere propriamente definiti «oggetti» di rilevazione, nel senso che sono inerti rispetto agli obiettivi conoscitivi e alle procedure messe in atto dal ricercatore. Lo stesso non si può dire per le persone, le quali, quando sono oggetto di rilevazione, attivano comunque complessi processi di attribuzione di senso che influenzano significativamente l'esito della rilevazione stessa. Che si tratti di compilare un questionario semplice e strutturato, oppure di ricostruire in modo approfondito esperienze, atteggiamenti o opinioni, le persone rispondono alle richieste dei ricercatori dando, anche implicitamente, molteplici significati all'interazione col ricercatore, anche in funzione delle intenzioni che gli attribuiscono e degli usi che prefigurano possano essere fatti delle informazioni che vengono raccolte (Pitrone, 2009; Cataldi, 2012).

Riguardo a questo aspetto può essere a mio avviso utile mettere a fuoco una differenza di fondo tra la raccolta di dati che le Università, gli istituti di ricerca pubblici e privati svolgono per scopi acquisizione e miglioramento della conoscenza scientifica di un dato fenomeno sociale e le rilevazioni, come quelle che si fanno nell'ambito del welfare, che producono dati per i quali è previsto un uso diretto, più o meno immediato o differito nel tempo, più o meno mirato su un target di popolazione specifico o generalizzato.

Nel primo caso – ricerca per scopi prevalentemente conoscitivi – le persone che partecipano all'indagine percepiscono il proprio ruolo soprattutto in termini di contributo volontario ad un'attività il cui output (i risultati della ricerca) non avrà conseguenze immediate e dirette sulla

propria vita quotidiana. In sostanza possiamo dire che in questi casi i soggetti – nella loro rappresentazione della situazione – concedono dati, risposte, informazioni ad un committente che persegue scopi conoscitivi suoi propri. La metafora della partecipazione come concessione è implicitamente confermata dalle principali strategie che, nella letteratura metodologica sulla survey, vengono proposte e discusse al fine di gestire al meglio la fase di contatto tra ricercatore/intervistatore e intervistato.

Generalmente si fa riferimento alla dimensione del questionario e alla durata dell'intervista (che devono essere mantenute sotto una certa soglia), alla possibilità da parte dell'intervistato, entro i limiti permessi dal disegno di indagine complessivo, di scegliere tempi e modalità dell'intervista, al rispetto della riservatezza e anonimato delle informazioni raccolte (Palmieri e Liani, 2016). Il patto implicito tra ricercatore e intervistato che queste ed altre strategie più specifiche sembrano sottendere colloca quest'ultimo nel ruolo di chi concede il suo tempo, il suo impegno cognitivo, le informazioni che lo riguardano al primo; un patto asimmetrico in cui è l'intervistato a dare qualcosa al ricercatore, e non viceversa. Nel frame di questo patto l'azione corrispettiva e reciproca del ricercatore consiste nel ridurre il più possibile i costi dell'aiuto che chiede all'intervistato.

Non a caso tra gli istituti di ricerca pubblici e privati si è diffuso ormai da tempo il ricorso a veri e propri incentivi materiali (buoni spesa, denaro, benefit) per agevolare la partecipazione degli intervistati (Church, 1993; Sanchez-Fernandez e al., 2010). L'asimmetria del patto tra ricercatore e intervistato è plasticamente confermata: l'indagine e i suoi risultati interessano al primo e non anche al secondo; l'intervista, piuttosto che una collaborazione, è quindi una prestazione, e in quanto tale va pagata.

Quando, come nel caso delle indagini sociali al servizio del welfare, una campagna di raccolta dati avviene per scopi di uso applicativo diretto e non solo per sviluppare conoscenza, la metafora della concessione sopra descritta non tiene più. Nelle rappresentazioni di chi partecipa alla rilevazione, il rapporto con il ricercatore, e con l'indagine in generale, non si esaurisce nel momento della concessione di informazioni, e quindi la logica degli incentivi e quella della riduzione dei costi sopra descritta appaiono riduttive. Verosimilmente in questi casi, chi concede i dati si sente, rispetto all'uso che può esserne fatto, portatore di interessi tanto quanto chi li riceve. Di conseguenza, la natura del patto implicito tra ricercatore e soggetti studiati si arricchisce: entrano forte-

mente in gioco aspetti come la fiducia, la reputazione e la credibilità, sia rispetto agli attori in campo (chi coordina la raccolta delle informazioni e chi ha la responsabilità dell'uso dei dati), sia rispetto alle procedure messe in atto, non solo in fase di rilevazione, ma anche di utilizzo. Evidentemente, queste tre dimensioni – fiducia, reputazione e credibilità – non sono del tutto ininfluenti nei contesti di ricerca di tipo più conoscitivo, ma il sempre crescente ricorso alla rilevazione di dati massivi come fase integrante dei processi decisionali istituzionali – nel welfare, ma non solo – ne aumenta il valore strategico, e pone sfide importanti per la ricerca sociale.

5. *L'etica della ricerca, una dimensione portante della qualità*

I due casi citati in precedenza – *Google flu trends* e le *biosocial surveys* – esemplificano due tipi di campagna di raccolta dati molto diversi tra loro, ascrivibile all'approccio *ready-made* il primo, e *custom-made* il secondo (sul punto vedi sopra, par. 2). Eppure, condividono un obiettivo innovativo: entrambi costituiscono due tentativi di ampliare la base di dati di riferimento per la progettazione di interventi di welfare. L'ambizione che li accomuna è valicare i confini dei presidi istituzionali tradizionali e raccogliere dati massivi riferibili ad ampie popolazioni di riferimento, e non solo a *target* specifici, cioè le persone che entrano in contatto con tali presidi. I *biomarkers*, cioè uno dei tipi di dati rilevati con le *biosocial surveys*, non sono una novità in sé; l'innovazione sta piuttosto nel tentativo di costruire e usare *databases* di *biomarkers* riferiti alla popolazione nel suo complesso, non solo quella ospedaliera. *Mutatis mutandis*, lo stesso vale per l'esperienza *Google flu trends*, nell'ambito però dell'analisi secondaria, cioè della raccolta e del trattamento di informazioni già esistenti (*ready-made*) e non rilevate ad hoc.

Proprio in questo spostamento su larga scala stanno sia le potenzialità che la fragilità dei sistemi di raccolta di dati massivi resi possibili dalle nuove tecnologie. Qualsiasi «macchina» di rilevazione di informazioni, per quanto sofisticata e affidabile sul piano tecnologico, poggia le sue fondamenta sulla qualità dei dati di input, e quindi, nel caso di rilevazioni di massa, sul contributo diffuso e granulare di tanti singoli individui sparsi sul territorio, fuori da quei presidi istituzionali che, in linea di principio, consentono un maggior controllo delle informazioni raccolte.

Questa prospettiva di larga scala rende le tre dimensioni sopra citate – fiducia, reputazione e credibilità di chi svolge la ricerca e di chi userà i

dati – strategiche per la qualità del materiale empirico rilevato; costituiscono cioè prerequisito irrinunciabile affinché ci si possa aspettare dai cittadini accuratezza nel fornire dati. Messa in questi termini, un'etica della ricerca improntata ai valori della fiducia, della reputazione e della credibilità sembra un'esigenza rilevante solo per la qualità dei sistemi informativi *custom-made*. In realtà anche quando si raccolgono e analizzano dati già esistenti – come le tracce che le persone lasciano online – queste tre dimensioni favoriscono l'accettabilità sociale di tali pratiche di ricerca, agevolandone il funzionamento.

Nella fase post-emergenza Covid-19 che stiamo vivendo in Italia, due casi mi paiono emblematici e rappresentativi del discorso sopra avviato. Mi riferisco all'app Immuni e all'indagine campionaria su vasta scala con test sierologici coordinata dall'Istat e dalla Croce rossa, cioè due importanti iniziative prese dal Governo italiano in ambito epidemiologico. Entrambe si configurano come due progetti di rilevazione massiva di informazioni, idealmente rivolte alla popolazione nel suo complesso e non a target specifici ospedalizzati. Una delle due (app Immuni) si configura come un sistema di monitoraggio esteso e capillare che sfrutta la diffusione degli smartphone in molte fasce della popolazione. Mentre sto scrivendo queste due iniziative sono ancora in corso⁴; un bilancio ed un'analisi accurata del loro andamento e della loro efficacia è quindi prematuro. Tuttavia, gli indicatori di partecipazione dei cittadini a queste due iniziative (tasso di consenso al prelievo sierologico da parte delle persone incluse nel campione Istat, percentuale di persone che hanno scaricato l'app Immuni, ecc.) hanno da subito dato segnali preoccupanti.

Giustamente, in ottemperanza di uno dei principi cardine dell'etica della ricerca, e cioè il rispetto della persona e della sua autonomia (Manson e O'Neill, 2007), le due campagne di raccolta di dati garantiscono, nel decidere se partecipare o no, libertà di scelta agli individui. Addirittura, l'app Immuni prevede quattro livelli di volontarismo: scaricamento dell'app, attivazione del bluetooth che consente la registrazione di eventuali contatti, disponibilità a sottoporsi al tampone in caso di segnalazione da parte dell'app, consenso allo sbloccaggio dei dati registrati dal proprio smartphone in caso di esito positivo del tampone.

A differenza delle applicazioni di tracciamento sviluppate in altri paesi, Immuni è stata progettata per una tutela massima della riservatezza e

⁴ A inizio agosto 2020 sono stati resi pubblici i primi risultati dell'indagine Istat-Croce rossa.

della privacy del singolo utente; tuttavia questo non è bastato a promuovere adeguatamente il suo utilizzo, il quale, da un punto di vista quantitativo, rimane sotto le attese minime che, secondo gli esperti, devono essere raggiunte perché Immuni possa essere efficace come strumento di monitoraggio diffuso e su larga scala. Come anticipato, qualsiasi analisi di bilancio rigorosa è prematura, tuttavia le testimonianze e le valutazioni degli esperti sui media nazionali hanno sottolineato il ruolo giocato dalle perplessità e dai dubbi dei cittadini sull'effettivo utilizzo dei dati raccolti da Immuni; perplessità e dubbi che i portavoce delle istituzioni non sono riusciti a dissipare.

Se, come credo, i problemi che sta incontrando Immuni non sono contingenti, ma potenzialmente strutturali di qualsiasi sistema di raccolta di informazioni massive per il welfare, gli *stakeholders* istituzionali e scientifici che vogliono investire nella centralità dei dati si trovano di fronte ad una sfida importante.

6. Considerazioni conclusive

Nello spazio che mi resta a disposizione cercherò di delineare due dimensioni, due ambiti di riflessione e di intervento, a mio avviso strategici per una ridefinizione più funzionale del patto tra istituzioni e cittadini nell'ambito delle campagne pubbliche di raccolta di dati.

Maggior equilibrio tra componente deontologica e consequenzialista nell'etica della ricerca. Nella letteratura specializzata si individuano due principi diversi che possono orientare l'etica della ricerca. Da una parte la deontologia, che mette in primo piano i modi con cui si conducono le indagini e chiede ai ricercatori il rispetto di certi limiti; rientrano in questo principio le procedure attraverso cui viene garantita l'autonomia dei partecipanti e il rispetto della riservatezza dei loro dati. Dall'altra il consequenzialismo, che invece valorizza i fini della ricerca, che deve essere orientata all'interesse pubblico e ad una beneficialità il più possibile estesa e democratica (Sinnott-Armstrong, 2014; Alexander e Moore, 2015).

Sia in riferimento alla sua progettazione/implementazione, sia per come è stato presentato al pubblico, Immuni, a mio avviso, rappresenta prevalentemente un approccio deontologico all'etica della ricerca.

Come si diceva in precedenza, molto si è infatti insistito sulla sua poca invasività, sull'autonomia lasciata al singolo utilizzatore e sulla non rintracciabilità delle identità personali, aspetto quest'ultimo che secondo alcuni esperti riduce le sue potenzialità come strumento di monitorag-

gio. Tutto ciò comunque non ha dissipato diffidenze e preoccupazioni da parte dei cittadini.

Verosimilmente, quando si tratta di campagne di raccolta dati che prevedono un loro uso applicativo diretto, e non organizzate per scopi prevalentemente se non esclusivamente conoscitivi, l'approccio deontologico non è sufficiente. In ottica consequenzialista, si potrebbero potenziare forme di comunicazione istituzionale che valorizzino chiaramente e dettagliatamente i benefici – per il singolo e per la comunità – dell'indagine che si intende svolgere; benefici che in qualche modo possano compensare e superare i rischi che le persone inevitabilmente associano ad una loro eventuale partecipazione ad essa. Si tratta cioè di affiancare ad una protezione rispetto ai mezzi (ottica deontologica) la condivisione dei fini (ottica consequenzialista). È inoltre importante che tale condivisione non assuma la forma di una comunicazione *top-down* dalle istituzioni al cittadino, ma venga invece costruita in modo partecipato, coinvolgendo la società civile nella varietà delle sue espressioni e dei suoi rappresentanti.

L'etica della ricerca come «ambiente» istituzionale. Per quanto una singola campagna di raccolta dati possa essere progettata e svolta in modo più o meno etico, credibilità, fiducia e reputazione rappresentano più propriamente tre dimensioni di base di un contesto più ampio – culturale, sociale, istituzionale – nel quale le ricerche vengono svolte. Una singola iniziativa, per quanto caratterizzata da una spiccata responsabilità etica, può fare poco per contrastare gli effetti di un senso generalizzato e diffuso di sfiducia verso l'uso che le istituzioni possono fare dei dati.

Per questo l'ottica consequenzialista sopra descritta dovrebbe servire alla costruzione di una cultura dell'informazione e della centralità del dato che in parte prescinde da singole iniziative di ricerca e che rispetto ad esse si configurerebbe come ambiente ideale di funzionamento. In questa prospettiva la collaborazione tra istituzioni e comunità scientifica dovrebbe avere un carattere più strutturale, piuttosto che attivarsi caso per caso, problema per problema.

Riferimenti bibliografici

Alexander L. e Moore M., 2015, *Deontological Ethics*, in Zalta E.N., *The Stanford Encyclopedia of Philosophy*, Stanford University, Stanford.

Benzeval M. e al., 2014, *Understanding Society: The UK Household Longitudinal Study. Biomarker User Guide and Glossary*, Institute for Social and Economic Research, Essex.

Cao L., 2012, *Social Security and Social Welfare Data Mining: An Overview*, «Ieee Transactions on Systems, Man, and Cybernetics», vol. 42, n. 6, pp. 837-853.

Cataldi S., 2012, *La ricerca sociale come partecipazione. Il rapporto tra ricercatore e attore sociale nell'indagine sociologica*, FrancoAngeli, Milano.

Church A.H., 1993, *Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis*, «Public Opinion Quarterly», vol. 57, n. 1, pp. 62-79.

Ginsberg J. e al., 2009, *Detecting Influenza Epidemics Using Search Engine Query Data*, «Nature», n. 457, pp. 1012-1014.

Keen P.W., 1987, *Decision Support Systems: The Next Decade*, «Decision Support Systems», vol. 3, n. 3, pp. 253-265.

Lazer D., Kennedy R., King G. e Vespignani A., 2014, *The Parable of Google Flu: Traps in Big Data Analysis*, «Science», n. 343, pp. 1203-1205.

Lupton D., 2015, *Digital Sociology*, Routledge, Londra.

Manson N.C. e O'Neill O., 2007, *Rethinking Informed Consent in Bioethics*, Cambridge University Press, Cambridge.

Marradi A. (a cura di), 2016, *Oltre il complesso d'inferiorità. Un'epistemologia per le scienze sociali*, FrancoAngeli, Milano.

Martire F. e Pitrone M.C., 2018, *L'integrazione tra le scienze sociali e le scienze biologiche nelle biosocial surveys*, «Sociologia e ricerca sociale», n. 116, pp. 61-74.

Palmieri M. e Liani S., 2016, *Il contatto nelle interviste telefoniche: rifiuti e strategie di persuasione*, «Rassegna italiana di sociologia», n. 1, pp. 5-30.

Pedersen J.S. e Wilkinson A., 2018, *The Digital Society and Provision of Welfare Services*, «International Journal of Sociology and Social Policy», vol. 38, n. 3/4, pp. 194-209.

Pitrone M.C., 2009, *Sondaggi e interviste. Lo studio dell'opinione pubblica nella ricerca sociale*, FrancoAngeli, Milano.

Power D.J., 2008, *Decision Support Systems: A Historical Overview*, in Burstein F. e Holsapple C., *International Handbooks Information System. Handbook on Decision Support Systems 1: Basic Themes*, Springer, Berlino.

Sala E., 2013, *Le indagini campionarie del nuovo millennio. Quali novità all'orizzonte?*, «Quaderni di sociologia», n. 62, pp. 77-90.

Salganik M.J., 2018, *Bit By Bit. Social Research in the Digital Age*, Princeton University Press, Princeton; trad. it. 2020, *Bit By Bit. La ricerca sociale nell'era digitale*, il Mulino, Bologna.

Sanchez-Fernandez J. e al., 2010, *An Analysis of the Effect of Pre-Incentives and Post-Incentives Based on Draws on Response to Web Surveys*, «Quality and Quantity», n. 44, pp. 357-373.

Schmitz N.H., 2018, *Rimanere connessi: la trasformazione delle indagini campionarie nell'era digitale*, «Sociologia e ricerca sociale», n. 116, pp. 20-32.

Schneider D. e Seelbeyer U., 2019, *Challenges in Using Big Data to Develop Decision Support Systems for Social Work in Germany*, «Journal of Technology in Human Services», vol. 37, n. 2/3, pp. 113-128.

Sinnott-Armstrong W., 2014, *Consequentialism*, in Zalta E.N., *The Stanford Encyclopedia of Philosophy*, Stanford University, Stanford.

Sonnega A. e al., 2014, *Cohort Profile: The Health and Retirement Study (Hrs)*, «International Journal of Epidemiology», n. 43, pp. 576-585.