

Ripensando ai dati in una società *data-driven*¹

Agnese Fabris²

Abstract

In un periodo storico eccezionale come quello che stiamo vivendo a causa della pandemia di Covid-19, in cui il mondo intero si è affidato ai dati per prendere decisioni che influenzano in maniera diretta e sostanziale la vita di tutti noi, si dimostra fondamentale promuovere un'alfabetizzazione ai dati della società che si caratterizzi per un nuovo modo di pensare ad essi. Il presente contributo offre una serie di spunti per sviluppare tale pensiero critico. Si inizierà riflettendo su come il processo di creazione dei dati sia intriso di decisioni umani, facendo sì che i numeri non siano né neutri né oggettivi. Si discuterà dei dati che non esistono e delle possibili ragioni dietro questi vuoti. Si esaminerà la questione della proprietà dei dati e il concetto di dati come bene comune. Infine, si concluderà sottolineando come la quantificazione sia solo un tipo di informazione, non necessariamente la migliore.

Parole chiave: Alfabetizzazione ai dati; Etica dei dati; Data feminism; Conteggio: aspetti sociali; Statistica: aspetti sociali

In an exceptional historical period such as the one we are experiencing due to the Covid-19 pandemic, in which governments all over the world have relied on data to take decisions that directly and substantially influence our lives, it is essential to promote data literacy in the society, together with a new way of thinking to data. This essay offers some starting points to develop such critical thinking. We will start by reflecting on how data creation processes are imbued with human decisions, ensuring that numbers are neither neutral nor objective. We will discuss data that do not exist and the possible reasons behind these gaps. We will analyse the question of data ownership and the concept of data as a common good. Finally, we will conclude by pointing out that quantification is only one type of information, not necessarily the best one.

Keywords: Data literacy; Data ethics; Data feminism; Counting: Social aspects; Statistics: Social aspects

¹ Saggio ricevuto in data 21/04/2021, accettato in data 24/05/2021, pubblicato in data 26/10/2021

² E-mail: agne.fab@gmail.com

1. Introduzione

Uno degli effetti dell'epidemia di Covid-19 in Italia è stato quello di catapultarci nel mondo dei dati. Improvvisamente dati, grafici e visualizzazioni hanno guadagnato protagonismo sulle pagine dei quotidiani italiani, così come nei discorsi e nelle decisioni politiche. Allo stesso tempo questa immersione nei numeri ha reso evidente la grande mancanza di una cultura dei dati, non solo nel grande pubblico, bensì nelle stesse istituzioni, che in questi mesi hanno pubblicato dati parziali, modificati in corso d'opera e il cui significato spesso si è fallito comunicare.

Partita dall'idea di scrivere un articolo di *data journalism* sull'impatto che la pandemia ha avuto in Italia sulla disuguaglianza di genere, e scontratami con la realtà dei tempi di pubblicazioni dei dati ufficiali e la scarsa ricchezza di dati disaggregati, ho deciso di utilizzare questo contributo per parlare più in generale di dati. Infatti, se da un lato sono una grande sostenitrice della cultura dell'argomentazione e dell'uso del metodo scientifico anche quando non si fa informazione scientifica e ritengo essenziale incentivare l'alfabetizzazione ai dati della società, dall'altro sono profondamente convinta che questa alfabetizzazione significhi, allo stesso tempo, insegnare a riconoscere i limiti dei dati e sfatare alcuni miti ad essi legati.

In tutto il mondo, l'emergenza pandemica ha sollevato domande importanti sull'apertura, la condivisione e l'uso dei dati, e ha messo in luce le sfide associate al loro utilizzo, risvegliando l'interesse del grande pubblico su temi e problematiche che già da alcuni anni venivano discussi tra i professionali della *data science*. Questo articolo si propone come una breve panoramica su queste tematiche. Per realizzarlo mi sono basata principalmente su tre lavori. Da una parte, il libro "*Data Feminism*"³ di Catherine D'Ignazio e Lauren F. Klein, pubblicato a marzo 2020, in cui le autrici esaminano in maniera approfondita le strutture di potere diseguali esistenti nel mondo dei dati, raccolgono e raccontano tentativi per rettificarle e propongono sette principi guida per tutti coloro che si occupano di dati o li utilizzano nel loro lavoro. Dall'altra parte, due articoli pubblicati a marzo 2021 su due riviste statunitensi. In primo luogo, l'articolo "*Why the Pandemic Experts Failed*"⁴ del *The Atlantic*, scritto da Robinson Meyer e Alexis C. Madrigal, giornalisti che si sono occupati per un anno del *COVID Tracking Project*⁵, una piattaforma interna al giornale per tenere traccia dei casi di Covid-19 e di altri dati relativi alla pandemia, e in cui gli autori raccontano i problemi incontrati durante il progetto. Infine, l'articolo

³ C. D'Ignazio, L.F. Klein, *Data Feminism*, The MIT Press, Cambridge (MA)-London 2020.

⁴ R. Meyer, A. C. Madrigal, *Why the Pandemic Experts Failed*, in «The Atlantic» (Internet); 15 marzo 2021. Consultato: aprile 2021. Disponibile all'indirizzo: <https://www.theatlantic.com/science/archive/2021/03/americas-coronavirus-catastrophe-began-with-data/618287/>

⁵ Il *COVID Tracking Project* è reperibile all'indirizzo: <https://covidtracking.com/>

“*What Data Can’t Do*”⁶ della matematica Hannah Fry pubblicato sulla rivista *The New Yorker*.

Il filo conduttore del presente contributo è un appello ad iniziare a pensare ai dati, allo loro raccolta, analisi e visualizzazione, con un approccio distinto, con un pensiero critico che non si faccia sedurre dalla loro apparente precisione e obiettività, e, pur riconoscendone il grande valore e potenziale, sia allo stesso tempo consapevole dei loro limiti.

2. I dati non esistono in natura

«I dati sono il modo in cui i nostri leader apprendono la realtà. In un certo senso, i dati sono la realtà del governo»⁷ affermano i giornalisti del *The Atlantic*, che osservano, ad esempio, come il processo di produzione dei dati sulla pandemia abbia finito modellando il modo in cui viene compresa la pandemia stessa.

Senza dubbio questa pandemia ha reso evidente l'importanza che i numeri possono arrivare a ricoprire nella nostra vita quotidiana. Decisioni politiche con effetti notevoli sui diritti fondamentali delle persone sono state, e continuano ad essere prese, sulla base dei dati, dimostrando quanto vulnerabile sia la società quando non si dispone di buone statistiche o quando si realizza una mala gestione o interpretazione di queste.

Oggi giorno, insomma, si dimostra ancora più impellente la necessità di intraprendere una riflessione critica sul mondo dei numeri e della quantificazione. Il primo passo per iniziare questa riflessione consiste, come affermano Meyer e Madrigal, nel comprendere che «tutti i dati sono creati, non esistono in natura»⁸.

Riconoscere che i dati non vengono semplicemente raccolti o analizzati, bensì esiste un processo di creazione, lungo e costoso, che inizia dal definire ciò che si vuole contare e/o classificare, e come si pretende farlo, significa al tempo stesso prendere coscienza delle numerose e importanti conseguenze che questo processo comporta.

La matematica Hannah Fry approfondisce questo tema sulle pagine del *The New Yorker*, offrendoci tra l'altro vari esempi tratti da due recenti libri che si occupano di dati e statistica: “*Counting: How We Use Numbers to Decide What Matters*”⁹ di Deborah Stone e “*The Data Detective*”¹⁰ di Tim Harford.

Nel complesso, possiamo identificare tre grandi conseguenze che il processo di creazione dei dati, volto ad ottenere una rappresentazione numerica della realtà, può generare.

In primo luogo, la conversione delle esperienze di vita in dati porterà sempre con sé una riduzione di tali esperienze e, sebbene i numeri possano fornirci utili *proxy* per un dato fenomeno, esisterà sempre un divario critico tra ciò che siamo in grado di misurare e ciò che ci interessa veramente. Esempi evidenti di questo sono

⁶ H. Fry, *What Data Can’t Do*, in «The New Yorker» (Internet), 22 marzo 2021. Consultato: aprile 2021. Disponibile all'indirizzo: <https://www.newyorker.com/magazine/2021/03/29/what-data-cant-do>

⁷ R. Meyer, A. C. Madrigal, *Why the Pandemic Experts Failed*, cit.

⁸ *Ibidem*.

⁹ D. Stone, *Counting: How We Use Numbers to Decide What Matters*, Liveright, New York 2020.

¹⁰ T. Harford, *The Data Detective: Ten Easy Rules to Make Sense of Statistics*, Riverhead Books, New York 2021.

l'uso del PIL come indicatore del benessere economico di un paese¹¹ o l'uso della performance degli studenti per misurare la qualità dell'educazione nelle scuole¹².

Un'altra conseguenza da considerare è l'impatto drastico che la definizione delle categorie, o delle stesse domande, utilizzate per raccogliere l'informazione può avere sui risultati. A questo proposito, Harford racconta la storia di una sconcertante discrepanza registrata nei tassi di mortalità infantile in Inghilterra, dove si osservava una mortalità notevolmente più alta nelle *Midlands* inglesi rispetto a Londra. Cosa stava succedendo? L'autore spiega come nel Regno Unito, qualsiasi gravidanza che termina dopo ventiquattro settimane viene legalmente considerata come una nascita, mentre una gravidanza che termina prima delle dodici settimane viene descritta come un aborto spontaneo. Tuttavia non esiste una terminologia legalmente stabilita per una gravidanza che si interrompa in un qualche momento tra questi due punti fissi. Quello che stava succedendo era che i medici dell'Inghilterra centrale avevano sviluppato l'abitudine di registrare che un bambino era morto; mentre i medici di Londra che si era verificato un aborto spontaneo. Una differente scelta di denominazione si traduceva in una sorprendente differenza nel tasso di mortalità infantile.

Infine, le semplificazioni, necessariamente introdotte per poter rappresentare il mondo attraverso i numeri, possono arrivare a produrre vere e proprie discriminazioni. Stone, ad esempio, nel suo libro racconta un episodio legato a un tentativo delle Nazioni Unite di sviluppare linee guida per misurare i livelli di violenza contro le donne. L'autrice spiega come rappresentanti di Europa, Nord America, Australia e Nuova Zelanda avanzarono idee sui tipi di violenza da includere negli indicatori sulla base di inchieste effettuate sulle vittime nei loro rispettivi paesi. Le categorie di violenza identificate includevano colpi, calci, morsi, schiaffi, spintoni, percosse e soffocamenti. Contemporaneamente, alcune donne del Bangladesh proposero di contare altre forme di violenza - atti che non sono rari nel subcontinente indiano - come bruciare donne, gettare acido su di loro, farle cadere dall'alto e costringerle a dormire in recinti per animali. Nessuno di questi atti fu incluso nella lista finale, con la conseguenza di rendere eventuali sondaggi basati sulle linee guida delle Nazioni Unite incapaci di rivelare granché sulle donne che subiscono queste forme di violenza.

Una volta preso atto dell'inevitabile ripercussione delle decisioni umane sui dati che raccogliamo, il passo seguente consiste nel considerare e nel dichiarare tale influenza quando si lavora con i dati, come si approfondisce nella sezione successiva.

¹¹ Al riguardo è opportuno menzionare che, sebbene si continui ad usare il PIL come indicatore del benessere economico delle nazioni, esiste una consapevolezza diffusa della sua limitatezza. Per una breve analisi sul tema si veda ad esempio: V. Lastrucci, M.J. Caldés Pinilla, *Il prodotto interno lordo: critiche e alternative per la misura del reale benessere delle nazioni*, in «Centro di Salute Globale» (Internet), 16 aprile 2015. Consultato: maggio 2021. Disponibile all'indirizzo: <http://www.centrosaluteglobale.eu/il-prodotto-interno-lordo-critiche-e-alternative-per-la-misura-del-reale-benessere-delle-nazioni/>

¹² Un esempio tra i tanti, *Eduscopio* ogni anno propone una mappa virtuale delle migliori scuole italiane stilata a partire dalle performance degli studenti al primo anno di università. Come si legge sul loro sito: «Medie dei voti e numero d'esami sono i criteri tenuti in considerazione dalla Fondazione Agnelli per la classifica Eduscopio 2020». Si veda: <https://eduscopio.it/>

3. I numeri non parlano da soli

«I dati si possono sempre ridurre a persone che contano cose» afferma la *data journalist* Meredith Broussard, nel suo libro “*Artificial Unintelligence*”¹³. Persone che, in base a una serie di decisioni, creano e raccolgono i dati. Prendere atto di ciò permette rendersi conto del fatto che i dati non sono neutri, bensì sono impregnati delle prospettive di chi sta decidendo cosa e come contare. Il fatto che i prodotti basati in dati sembrano oggettivi è semplicemente dovuto al fatto che «le prospettive di coloro che li producono – uomini bianchi d’élite e le istituzioni che essi controllano- sono prese come default»¹⁴ osservano le scrittrici di *Data Feminism*.

Seguendo le orme di filosofe femministe come Donna Haraway, Sandra Harding e Linda Alcoff, che offrono concetti alternativi a quello di obiettività universale considerato irraggiungibile, D’Ignazio e Klein sostengono che, invece di valorizzare l’ideale di neutralità e cercare di cancellare tutte le tracce umane dai prodotti basati sui dati, sia essenziale ricollegare i dati al contesto nel quale sono stati creati.

Infatti, come sostiene Haraway, tutte le forme di conoscenza sono *situate*¹⁵, nel senso che sono prodotte da persone specifiche in circostanze - culturali, storiche e geografiche- specifiche e per questo si rivela necessario impiegare quella che Haraway definisce *obiettività femminista*¹⁶ come strumento per svelare i differenti punti di vista o prospettive parziali e spiegare tale natura situata della conoscenza.

Anche e soprattutto quando si parla di dati il concetto di fondo è lo stesso: i numeri non parlano da soli. Il contesto nel quale si creano i dati importa e deve essere tenuto in considerazione in tutte le fasi di un progetto di dati, dalla raccolta, all’analisi, alla comunicazione dei risultati. Considerare il contesto significa comprendere la provenienza dei dati, ossia l’ambiente e le condizioni nel quale sono stati raccolti, svelare i differenziali di potere in azione, considerare chi ha finanziato la raccolta e indagare su eventuali conflitti di interesse.

In relazione a questo tema, ultimamente, nella comunità di *data scientists* si stanno sviluppando vari metodi e strumenti per mantenere il contesto attaccato ai dati. Un esempio è quello proposto dalla scienziata di dati Heather Krause con la cosiddetta *biografia del dato*¹⁷. Krause sostiene che chiunque si accinga a lavorare con un set di dati dovrà prima di tutto compilare la biografia di tali dati, rispondendo a quattro apparentemente semplici domande: «Da dove vengono i dati? Chi li ha raccolti? Come? E soprattutto perché?». Questo compito non è sempre così semplice come potrebbe sembrare a prima vista, ma portarlo a termine può aiutarci a rivelare lacune, pregiudizi, disinformazione o dettagli trascurati nel raccontare la nostra storia.

¹³ M. Broussard, *Artificial Unintelligence: How Computers Misunderstand the World*, The MIT Press, Cambridge (MA)-London 2018.

¹⁴ C. D’Ignazio, L. F. Klein, *Data Feminism*, cit., p.60.

¹⁵ D. Haraway, *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective*, In: «Feminist Studies», vol. 14, n. 3, Autumn 1988, pp. 575-599.

¹⁶ *Ibidem*.

¹⁷ H. Krause, *Data Biographies: how to get to know your data*, In: datassist.com, 17 marzo 2017. Consultato: marzo 2021. Disponibile all’indirizzo: <https://datassist.com/data-biographies-how-to-get-to-know-your-data/>

Riassumendo ancora una volta con le parole di D'Ignazio e Klein, lavorare con i dati attraverso un approccio femminista significa ricordare sempre che «dietro ai dati ci sono le persone, persone che offrono la loro esperienza per essere contattata e analizzata, persone che effettuano tale conteggio e analisi, persone che creano visualizzazioni per esporre e promuovere i risultati di un dato progetto e persone che usano il prodotto finale»¹⁸.

Questa considerazione ci conduce alla terza riflessione di questo articolo: i dati che non esistono. «Nel bene o nel male, infatti, ci saranno sempre persone che non vengono contattate»¹⁹.

4. Dati mancanti e contro-dati

Una delle conseguenze principali del fatto che la maggioranza dei lavori basati sui dati vengano creati dai gruppi dominanti è, indubbiamente, che alcuni dati non vengono proprio raccolti. Le ragioni dietro questa mancanza sono molteplici: dai pregiudizi, a una mancanza di volontà sociale e/o politica, a un vero e proprio disinteresse strutturale.

Quello che è chiaro è che ciò che viene contato esiste, sia nelle politiche pubbliche che nell'immaginario delle persone. «La quantificazione è rappresentazione» dice il giornalista Jonathan Stray²⁰. I numeri influenzano la nostra capacità di prenderci cura, condividere empatia e investire risorse. Sono la condizione di esistenza di un problema o di una data realtà sociale nell'agenda delle istituzioni.

Negli Stati Uniti, l'artista e designer Mimi Onuoha ha creato il progetto “*The library of missing datasets*”²¹, ovvero “la biblioteca dei dataset mancanti”, che consiste in una raccolta di serie di dati che ci si potrebbe aspettare esistere nel mondo, ma che in realtà non sono mai state create. Il progetto esiste sia come pagina web che fisicamente, come un oggetto artistico consistente nel tipico armadietto a cassette usato per i cataloghi delle biblioteche, pieno di schede che presentano titoli del tipo: «Persone escluse dalle case popolari a causa di precedenti penali», «Immigrati privi di documenti attualmente incarcerati e/o sottopagati», «Quanto Spotify paga ciascuno dei suoi artisti per riproduzione di un brano», etc. I visitatori possono scorrere con le dita tra le schede e, quando ne tirino fuori una che chiami loro l'attenzione, scopriranno che è vuota. Come tutte le altre. Quei dati non esistono.

Onuoha sostiene che «ciò che ignoriamo rivela più di ciò a cui diamo la nostra attenzione» e, «in un mondo altresì saturo di dati», «i punti che abbiamo lasciato in bianco rivelano i nostri pregiudizi sociali e indifferenze nascoste»²². L'artista identifica varie possibili motivazioni alla radice di questi vuoti. A volte, sostiene, c'è un vero e proprio interesse da parte di chi è al potere a tenere nascoste certe statistiche. Altre volte, le istituzioni semplicemente ritengono che

¹⁸ C. D'Ignazio, L. F. Klein, *Data Feminism*, cit., p.10.

¹⁹ *Ibidem*.

²⁰ J. Stray, *The Curious Journalist's Guide to Data*, in «Columbia Journalism Review» (Internet), 24 marzo 2016. Consultato: aprile 2021. Disponibile all'indirizzo: https://www.cjr.org/tow_center_reports/the_curious_journalists_guide_to_data.php

²¹ M. Onuoha, *On Missing Data Sets*, GitHub (Internet) (2018). Consultato: marzo 2021. Disponibile all'indirizzo: <https://github.com/MimiOnuoha/missing-datasets>

²² *Ibidem*.

l'investimento necessario per raccogliere certi dati non superi un'analisi costi-benefici. In altri casi ancora, sono le stesse persone coinvolte nel conteggio a percepire come troppo costoso emotivamente il processo di raccolta delle loro esperienze, come può essere il caso degli abusi sessuali, la cui occorrenza viene generalmente sottostimata. Una delle ragioni alla base di questo errore sistematico è che in molti casi l'atto stesso di denunciare un'aggressione sessuale si rivela un processo troppo intenso, doloroso e difficile, al punto da non superare il vantaggio del denunciare. Ci sono poi casi in cui i dati da raccogliere resistono alla semplice quantificazione, infatti, non tutte le cose sono facilmente misurabili e il mondo non si adatta sempre a rigide categorie. Possiamo pensare ad esempio agli episodi di razzismo istituzionale che possono essere ambigui o confutabili, o alle emozioni che proviamo e che molto spesso sono difficili da tradurre in numeri o categorie.

Una strategia per combattere il fenomeno dei dati mancanti, fenomeno comune a tutte le società caratterizzate da relazioni di potere diseguali, è l'azione collettiva volta a raccogliere *counter-data*²³, ovvero “contro-dati”, dati raccolti dal basso per apportare una visione alternativa rispetto a quella dei dati ufficiali.

In Italia esistono varie iniziative di tipo *counter-data*. Un esempio è la campagna “*Che Aria Tira?*”²⁴ del comitato *Torino Respira*, che a partire dal 2019 misura in maniera capillare i livelli di inquinamento atmosferico nel territorio torinese, utilizzando dei campionatori finanziati dai cittadini stessi e installati in punti prossimi ai luoghi di vita, di scuola e di lavoro dei partecipanti. O ancora, *L'Osservatorio CiviO sulla casa e la residenza*²⁵ di Venezia avviato nel 2018 per analizzare, attraverso il recupero e lo studio di dati, la questione abitativa nella Venezia insulare, e che offre una piattaforma di tipo collaborativo che consente di inserire dati, materiali e strumenti utili per approfondire temi come: acquisti, affitti turistici, seconde case, flussi migratori tra città storica e terraferma, etc. Infine, un'altra iniziativa che merita la pena citare è il progetto femminista *Obiezione Respinta*²⁶, che si propone di combattere la scarsità di dati ufficiali relativi agli obiettori di coscienza nelle strutture sanitarie pubbliche. Il progetto mette a disposizione una piattaforma dove chiunque può segnalare la sua esperienza con ospedali e farmacie, con l'obiettivo di costruire in maniera collettiva una base di dati su medici e farmacisti che ostacolano il diritto all'aborto.

5. Invisibilità o visibilità giusta?

Tantissimi sono gli esempi, in Italia e nel mondo, di iniziative basate su raccolte dati partite dal basso, dalle stesse comunità che soffrono discriminazioni o oppressioni e sono interessate a far capire a coloro che stanno al potere il carattere e la dimensione di problemi troppo distanti dalla loro realtà.

Ciononostante, non bisogna dimenticare che in certi casi possono esistere dei vantaggi nel non essere contattati o tracciabili. Un dataset mancante può funzionare, per un gruppo vulnerabile, come una forma di protezione, uno scudo che può rendere invisibili ad un'attenzione non richiesta, che troppo spesso assomiglia più a

²³ Si veda in proposito: M. Currie *et al.*, *The Conundrum of Police Officer-Involved Homicides: Counter-Data in Los Angeles County*, in «Big Data & Society», vol. III, n. 2, 2016, pp. 1-14.

²⁴ <https://www.torinorespira.it/cheariatira2021/>

²⁵ <https://ocio-venezia.it/>

²⁶ Il sito della progetto è disponibile all'indirizzo: <https://obiezionerespinta.info/>

controllo e sorveglianza, e può portare con sé rischi importanti per la salute o la sicurezza delle persone appartenenti a tale comunità. Dai senz'altro ai detenuti, dai migranti alle prostitute, molti sono i gruppi che generalmente preferiscono rimanere invisibili piuttosto che affrontare la repressione, lo stigma o la deportazione.

Su questo tema l'emergenza pandemica ci obbliga, ancora una volta, ad affrontare nuove sfide, costringendoci a riflettere su possibili soluzioni. A tal proposito, nell'articolo "*Making migrants visible to COVID-19 counting: the dilemma*", pubblicato ad aprile 2020 su openDemocracy, i ricercatori Stefania Milan, Annalisa Pelizza e Yoren Lausberg scrivono:

La pandemia COVID-19 ci pone di fronte a un dilemma per quanto riguarda le popolazioni invisibili, e i migranti in particolare, che ha a che fare con preoccupazioni sia sociali che tecnologiche. Da una parte, le lacune di visibilità sono un aspetto sistemico nella gestione della popolazione, che potrebbe essere accolto favorevolmente sia dai responsabili politici che dalle popolazioni stesse. [...] Dall'altra, un aumento della visibilità delle popolazioni migranti potrebbe aiutare a frenare il contagio ed evitare focolai massivi tra le popolazioni vulnerabili. Essere invisibili, infatti, si traduce nell'incapacità di accedere a servizi cruciali al tempo della pandemia, primo fra tutti l'assistenza sanitaria. L'accesso a test e cure richiede un'assicurazione e l'assicurazione richiede essere contattati. Anche quando i costi dell'assicurazione possono essere assunti dalla collettività, venire contattati resta una condizione fondamentale di accesso.²⁷

Questo è un chiaro esempio di quello che le autrici di *Data Feminism* chiamano *paradosso dell'esposizione*, ovvero uno scenario in cui coloro che potrebbero guadagnare in modo significativo dall'essere contattati vengono posti in grande pericolo da quello stesso atto di conteggio (o classificazione)²⁸.

Nel loro articolo, Milan, Pelizza e Lausberg sostengono che la situazione attuale imponga riconsiderare la relazione tra dati, popolazioni e (in)visibilità e si chiedono in quali condizioni includere popolazioni invisibili nel conteggio generale creato per la gestione dell'emergenza possa rivelarsi una soluzione giusta. L'approccio che suggeriscono è un approccio su più fronti, che oltre a valutare come effettuare il conteggio, si concentri anche ad esaminare quale infrastruttura digitale usare a tale scopo.

I tre ricercatori propongono che si sviluppi un nuovo sistema di conteggio, che garantisca quella che chiamano *visibilità giusta*, ovvero una visibilità che comporti una forma de facto di inclusione civile delle popolazioni migranti, garantendo un accesso permanente all'assistenza sanitaria, al welfare e ai diritti civili. I principi su cui dovrebbe basarsi questa visibilità giusta includono prima di tutto la minimizzazione dei dati, ovvero la riduzione dell'informazione raccolta allo stretto necessario, e la limitazione delle finalità, ovvero il definire scopi specifici e legittimi ed esplicitarli. In aggiunta, il processo di conteggio dovrebbe trattare i dati personali in modo trasparente per gli interessati, sottoponendosi inoltre al controllo democratico e all'obbligo di rendiconto. Si dovrebbe garantire il diritto all'oblio, ovvero il diritto di essere cancellati da qualsiasi base di dati e di non essere tracciati

²⁷ S. Milan, A. Pelizza, Y. Lausberg, *Making migrants visible to COVID-19: the dilemma*, in «openDemocracy» (Internet); 28 aprile 2020. Consultato: aprile 2021. Disponibile all'indirizzo: <https://www.opendemocracy.net/en/can-europe-make-it/making-migrants-visible-covid-19-counting-dilemma/>

²⁸ C. D'Ignazio, L. F. Klein, *Data Feminism*, cit., p.105.

oltre gli obiettivi originari. Allo stesso tempo gli autori sostengono l'importanza di considerare adeguatamente la dimensione infrastrutturale, soprattutto per quanto riguarda la proprietà e il controllo, pubblico rispetto a privato, della infrastruttura digitale scelta. Infine, sottolineano l'importanza di includere tra le entità responsabili di effettuare il conteggio organizzazioni guidate da migranti, centri di accoglienza, istituzioni sanitarie, sindacati e organizzazioni che lavorano sul campo.

In conclusione, le iniziative di conteggio e di classificazione, specialmente per quanto riguarda i gruppi minoritari, dovrebbero sempre bilanciare danni e benefici. Infatti, raccogliere dati su persone reali e sulle loro vite comporterà sempre una serie di rischi che vanno dall'esposizione indesiderata alla violenza. Per questo si dimostra di fondamentale importanza ottenere il consenso degli individui ad essere contati o classificati e fornire loro il diritto di rifiutare tali atti. Contare è sempre complicato. Ma, se intrapreso deliberatamente, adattato a obiettivi specifici, e tenendo in considerazione la questione della privacy e dei potenziali danni, il conteggio può contribuire a creare una visibilità importante e desiderata.

6. I dati come bene comune

Nelle sezioni precedenti si è riflettuto sui processi di creazione e raccolta dei dati, mettendo in risalto l'inevitabile influenza delle decisioni umane su tali processi, decisioni che riguardano cosa misurare esattamente (e cosa no), quando e dove farlo e con quali metodi. Il prossimo aspetto da analizzare è la questione della proprietà dei dati.

Prima di addentrarsi in tale questione è fondamentale considerare che raccogliere, analizzare e immagazzinare dati è un processo molto costoso che richiede grandi risorse. Pertanto solo istituzioni potenti, come grandi aziende, governi e grandi istituti di ricerca o università, dispongono dei mezzi per lavorare con i dati su grande scala. Questo fatto comporta che i dati vengono utilizzati in maniera consistente principalmente a tre fini: scienza, lucro e sorveglianza.

Negli ultimi anni, sempre più spesso abbiamo sentito parlare dei dati come del *nuovo petrolio*²⁹. In un certo senso questa è un'analogia accurata. I dati sono una sorta di risorsa naturale inesplorata che, una volta elaborata e raffinata, si trasforma in conoscenza preziosa, in profitto e, in ultima analisi, in potere. Inoltre, come succede con l'industria petrolifera, la *data economy* comporta un costo ecologico importante. Infatti, lo stoccaggio di volumi sempre più grandi di dati richiede spazi fisici sempre più ampi, che necessitano potenti impianti di raffreddamento, tanto che l'impatto ambientale di questi *data centers* sembra possa arrivare ad avere un effetto acceleratore sul cambiamento climatico globale³⁰.

Nondimeno, esiste una differenza importante rispetto al petrolio. Mentre quest'ultimo viene estratto dalla terra e poi venduto alle persone, i dati vengono sia estratti dalle persone che rivenduti a queste, mettendo in atto quella che alcuni

²⁹ Si veda in proposito l'articolo: *The world's most valuable resource is no longer oil, but data* in «The Economist» (Internet), 6 maggio 2017. Consultato: aprile 2021. Disponibile all'indirizzo: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>

³⁰ Si veda in proposito l'articolo di N. Jones, *How to stop data centres from gobbling up the world's electricity*, in «Nature», 561, 2018, pp. 163-166. Disponibile online all'indirizzo: <https://www.nature.com/articles/d41586-018-06610-y>

studiosi definiscono una nuova forma di colonialismo, il *colonialismo dei dati*³¹, che, sebbene con modalità, scale e intensità diverse dal colonialismo storico, condivide con quest'ultimo la funzione storica di espropriare.³² Si crea così una profonda asimmetria tra chi raccoglie, immagazzina e analizza i dati e coloro i cui dati vengono raccolti, archiviati e analizzati.

A seguito dell'attuale trasformazione digitale della società e dell'economia, una fetta sempre più grande delle nostre vite si svolge online, e ciò comporta che un gruppo molto ristretto di grandi imprese tecnologiche sia proprietario della grande maggioranza dei dati da noi prodotti. Infatti, sfruttando il fatto che gli individui singoli non hanno un grande potere contrattuale, le aziende chiedono, ed ottengono, alle persone di rinunciare a ogni controllo sui loro dati personali.

Sulla base di queste considerazioni, già da alcuni anni, nel settore si stanno proponendo e studiando nuovi sistemi di *data governance*. Ad esempio, si discute sulla possibilità, e le eventuali modalità, di implementare una qualche forma di nazionalizzazione dei dati³³, estendendo al mondo digitale l'idea, né nuova né assurda, che il controllo delle risorse di un dato territorio spetti al popolo che abita tale territorio.

Un'idea che sta prendendo piede è quella dei cosiddetti *data trust*, ovvero l'idea di affidare i dati, personali o commerciali, a un organismo indipendente e democraticamente responsabile, che conservi le informazioni adeguatamente anonimizzate e le renda disponibili per la pubblica utilità. Ad esempio, il sociologo e scrittore Evgeny Morozov sostiene che, per potere sfruttare il grande potenziale che deriva dal combinare molti set di dati, è importante che questi appartengano ad un'unica entità che però non dovrebbe essere una grande azienda tecnologica, ma piuttosto un fondo dati nazionale di proprietà di tutti i cittadini³⁴. In un'intervista del 2018 Morozov afferma:

Negli Stati Uniti si parla molto di imporre una divisione dei server - il cosiddetto movimento "Breakup Google". Si pensa che ciò che ha funzionato all'inizio del ventesimo secolo per bloccare il monopolio del petrolio ora funzionerà con i dati. Ma quando divisi i dati perdono gran parte del loro valore. E l'obiettivo non deve essere quello di distruggere l'intelligenza artificiale. Non vogliamo rinunciare a usarla per sconfiggere il cancro o ottimizzare l'energia, piuttosto garantire che stati

³¹ Si veda in proposito: N. Couldry, U. A. Mejias, *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*, in «Television & New Media», vol. 20, n. 4, 2019, pp. 336–349.

³² *Ibidem*

³³ Si veda ad esempio: B. Tarnoff, *Big data for the people: it's time to take it back from our tech overlords*, in «The Guardian» (Internet), 14 marzo 2018. Consultato: aprile 2021. Disponibile all'indirizzo: <https://www.theguardian.com/technology/2018/mar/14/tech-big-data-capitalism-give-wealth-back-to-people>. Si veda anche: *Who owns the web's data?*, in «The Economist» (Internet), 24 ottobre 2020. Consultato: maggio 2021. Disponibile all'indirizzo: <https://www.economist.com/business/2020/10/22/who-owns-the-webs-data>. O ancora: U. A. Mejias, *Why the Global South should nationalise its data*, in «Aljazeera» (Internet), 14 dicembre 2019. Consultato: marzo 2021. Disponibile all'indirizzo: <https://www.aljazeera.com/opinions/2019/12/14/why-the-global-south-should-nationalise-its-data>

³⁴ E. Morozov, *To tackle Google's power, regulators have to go after its ownership of data*, in «The Guardian» (Internet), 2 luglio 2017. Consultato: maggio 2021. Disponibile all'indirizzo: <https://www.theguardian.com/technology/2017/jul/01/google-european-commission-fine-search-engines>

e città non debbano dipendere da enti privati per farlo. Invece di scindere i giganti della tecnologia, è meglio costringerli a riconoscere che i dati non appartengono a loro e obbligarli a renderli anonimi e accessibili alla comunità. Il loro utilizzo dovrebbe poi essere regolamentato (ad esempio, gratuito per chi progetta servizi pubblici, a pagamento per chi ha finalità commerciali). Ma tutto questo dovrebbe essere accompagnato da una strategia industriale robusta e proattiva che permetta a chi è rimasto finora solo a guardare (l'Europa) di sfruttare il talento che esiste in abbondanza nelle sue università per costruire sistemi di intelligenza artificiale a beneficio della collettività.³⁵

La realtà è che, sebbene siamo ancora molto lontani dalla nazionalizzazione dei dati, stiamo assistendo a una sorta di risveglio globale su questi temi, e i cittadini iniziano a mobilitarsi per reclamare i loro diritti e a proporre e sviluppare sistemi alternativi di proprietà dei dati. Ad esempio, stanno progressivamente nascendo “cooperative di dati”, ovvero gruppi di persone che mettono in comune i propri dati o impronte digitali per creare una risorsa preziosa che si possa sfruttare sotto condizioni stabilite collettivamente. Un esempio tra i tanti è la cooperativa MIDATA³⁶, fondata in Svizzera nel 2015. MIDATA offre una piattaforma digitale dove gli utenti-membri possono caricare copie dei propri dati medici, così come informazioni supplementari come dieta, esercizio fisico o metriche del sonno, informazioni oggi giorno facilmente accessibili grazie alle sempre più diffuse applicazioni mobili. Come si legge sul loro sito: «I titolari di account dati MIDATA possono contribuire attivamente alla ricerca medica e agli studi clinici fornendo un accesso selettivo ai propri dati personali. Inoltre, possono diventare soci della cooperativa e controllarla in quanto tali». Trasformandosi in un importante archivio di dati sanitari MIDATA si propone di incentivare e agevolare la realizzazione di progetti di ricerca in campo medico, dimostrando come i dati possano essere utilizzati per il bene comune, ma salvaguardando al contempo il diritto dei cittadini al controllo dei propri dati personali.

Il cambiamento di mentalità che porta a considerare i dati un bene comune da usare a beneficio della società, piuttosto che una semplice risorsa commerciale con la quale fare soldi, si sta diffondendo sempre di più anche nell'opinione pubblica. A radice dell'emergenza pandemica, in Italia, ad esempio, è nata l'iniziativa *#DatiBeneComune*³⁷, lanciata a novembre 2020 da alcune tra le più importanti organizzazioni italiane attive sui temi della trasparenza e dell'integrità, così come da giornalisti e da cittadini privati, per chiedere al governo di aprire i dati sulla gestione dell'emergenza Covid-19. I sostenitori della campagna, che ha già superato le cinquantamila firme, chiedono al Governo di rendere i dati pubblici, disaggregati, continuamente aggiornati, ben documentati e facilmente accessibili a ricercatori, decisori, media e cittadini. Nella lettera al Governo, i promotori scrivono:

La trasparenza dei dati deve essere garantita ed essere riconosciuta come un principio cardine perché costituisce uno straordinario presidio a difesa dell'imparzialità delle scelte dei decisori pubblici, affinché non prevalgano interessi

³⁵ E. Morozov, *Why Big Data should belong to people*, L. Traldi, in «Medium» (Internet), 28 settembre 2018. Consultato: maggio 2021. Disponibile all'indirizzo: <https://medium.com/@lauratraldi/evgeny-morozov-why-big-data-should-belong-to-the-people-81e4e38408fc>

³⁶ Il sito della cooperativa è disponibile all'indirizzo: <https://www.midata.coop/it/homepage/>

³⁷ Il sito dell'iniziativa è disponibile all'indirizzo: <https://www.datibenecomune.it/>

particolari sull'interesse generale. In questo momento storico più che mai rivendichiamo il ruolo attivo della cittadinanza e il suo diritto ad avere accesso a dati aperti relativi all'avanzamento di un piano strategico che si deve comporre di obiettivi e risultati attesi monitorabili attraverso indicatori specifici³⁸.

La strada è ancora lunga, però, il fatto che si inizi a chiedere alle amministrazioni che i dati pubblici siano forniti in un formato aperto³⁹ rappresenta, senza dubbio, un primo passo importante.

7. I dati non sono sempre la risposta

Abbiamo parlato di dati, del loro potere, dei loro limiti e del loro valore, per concludere questa panoramica rimane ancora una riflessione da fare, per la quale ricorrerò nuovamente ai giornalisti responsabili del *COVID Tracking Project* che sulle pagine del *The Atlantic* affermano: «i dati sono solo un altro tipo di informazione»⁴⁰. Un tipo alternativo, non necessariamente un tipo migliore, come invece spesso ci ritroviamo a pensare a causa di quel pregiudizio inconscio che ci porta a valutare di più il misurabile dell'incommensurabile e che la etnografa Tricia Wang denomina *quantification bias*⁴¹.

Wang sostiene che, affinché aziende, governi o organizzazioni possano ottenere un quadro completo di una determinata situazione oggetto di studio, sia fondamentale integrare i dati quantitativi con quelli che lei definisce *thick data*⁴², ovvero dati densi. Tali dati consistono nell'informazione che emerge utilizzando metodi di ricerca qualitativa ed etnografica, la quale è in grado di rivelare le emozioni e le storie delle persone, apportando così spessore e profondità ai fenomeni osservati su larga scala con lenti quantitative.

A tal riguardo, la ricercatrice racconta un episodio emblematico. Nel 2009 lavorava per Nokia, al tempo una delle maggiori imprese di cellulari del mondo, quando scoprì qualcosa che credeva avrebbe potuto mettere in discussione l'intero modello commerciale dell'azienda. Grazie ad anni di ricerca etnografica in Cina, vivendo tra i migranti, lavorando come venditrice ambulante e frequentando assiduamente internet caffè, Wang era arrivata alla conclusione che i consumatori a basso reddito di questo paese erano disposti a pagare per smartphone più cari. Decise quindi riportare i risultati della sua ricerca all'azienda, suggerendo un cambio di strategia di sviluppo, consistente nel togliere centralità alla produzione di smartphone cari destinati ad utenti d'élite per darla alla produzione di smartphone

³⁸ Lettera aperta al Presidente Mario Draghi, 17 febbraio 2021. Disponibile all'indirizzo: <https://www.datibenecomune.it/lettere/>

³⁹ I dati aperti sono un sottoinsieme dei dati pubblici che devono rispettare tre requisiti: 1. sono accessibili gratuitamente o a costi marginali; 2. sono provvisti di una licenza che ne consente il massimo riutilizzo, anche commerciale; 3. sono disaggregati, ben documentati e forniti in un formato aperto, ovvero neutro da un punto di vista tecnologico, e *machine readable*.

⁴⁰ R. Meyer, A. C. Madrigal, *Why the Pandemic Experts Failed*, cit.

⁴¹ Si veda in proposito: T. Wang, *The human insights missing from big data*, in «TED: Ideas Worth Spreading» (Internet), settembre 2016. Disponibile all'indirizzo: https://www.ted.com/talks/tricia_wang_the_human_insights_missing_from_big_data

⁴² T. Wang, *Why Big Data Needs Thick Data*, in «Medium» (Internet), 20 gennaio 2016. Consultato: marzo 2021. Disponibile all'indirizzo: <https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7#.xhtm64pe>

economici per utenti a basso reddito. Le sue raccomandazioni tuttavia non vennero ben accolte. Nokia rispose che la sua ricerca era debole e la dimensione del suo campione troppo piccola rispetto al loro campione costituito da milioni di punti. Inoltre sostennero che nei loro dataset non esistessero segni di quanto concluso da Wang. Per contro, la ricercatrice non considerava sorprendente tale mancanza di indizi, ben consapevole che la nozione di domanda dell'azienda si basasse su un modello quantitativo rigido incapace di rappresentare un funzionamento della domanda influenzato da parametri culturali. Oggi tutti conosciamo il declino subito da Nokia nell'ultima decada. Secondo Wang, tra le molte ragioni all'origine di tale crollo, un ruolo fondamentale fu giocato dal fatto che l'azienda abbia fatto troppo affidamento sui numeri, confondendo misurabilità e valore⁴³.

Questo monito a non considerare i dati l'informazione per eccellenza si dimostra di particolare importanza in un periodo storico come quello che stiamo vivendo, come ci ricordano i giornalisti responsabili del *COVID Tracking Project*:

I dati sembrano avere una pretesa preminente sulla verità. I responsabili politici si vantano del processo decisionale basato sui dati e promettono di “seguire la scienza”. Ma abbiamo trascorso un anno immersi nei dati. Fidati di noi: i dati non sono davvero niente di speciale. I dati sono solo un mucchio di conclusioni qualitative organizzate in modo numerabile. Il pensiero basato sui dati non è necessariamente più accurato di altre forme di ragionamento e, se non capisci come vengono creati i dati, le loro cuciture e cicatrici, potrebbero anche finire per fuorviarti. [...] Governare durante una pandemia, o qualsiasi emergenza, significa prendere le decisioni meno peggiori con le migliori informazioni disponibili. Queste informazioni possono assumere molte forme; non devono necessariamente essere dati. Ma se si decide di considerare i dati, allora è fondamentale capire come è stato creato ogni punto, ogni cella; in caso contrario, esiste una grande probabilità di essere indotti in errore.⁴⁴

8. Conclusioni

Come sostengono le autrici di *Data Feminism*, lavorare con i dati con una nuova prospettiva, una prospettiva femminista, richiede la conoscenza e il riconoscimento della storia imperfetta esistente dietro ai dati. Del fatto che troppo spesso vengano usati da coloro che detengono il potere per consolidare il loro controllo su luoghi, cose e persone, con l'obiettivo di preservare uno status quo ineguale⁴⁵. Ma prendere atto di «questa storia imperfetta non significa cedere il controllo del futuro ai poteri del passato. I dati fanno parte del problema, certo. Ma fanno anche parte della soluzione»⁴⁶ rimarcano D'Ignazio e Klein. Utilizzando i dati, infatti, si possono sfidare e riequilibrare distribuzioni diseguali di potere, i dati rappresentano uno strumento potente per rivendicare storie trascurate, per denunciare ingiustizie, per distinguere tra esperienza personale e oppressione collettiva⁴⁷, per emancipare, per costruire collettività e solidarietà. Sempre che non ci si dimentichi che, come

⁴³ *Ibidem*.

⁴⁴ R. Meyer, A. C. Madrigal, *Why the Pandemic Experts Failed*, cit.

⁴⁵ C. D'Ignazio, L. F. Klein, *Data Feminism*, cit., p.17.

⁴⁶ *Ibidem*.

⁴⁷ Si veda in proposito: A. Oakley, *Paradigm Wars: Some thoughts on a Personal and Public Trajectory*, in «International Journal of Social Research Methodology», vol. II, n. 3, 1999, pp. 247-254.

afferma Deborah Stone, «per contare bene, serve umiltà per sapere cosa non può o non deve essere conteggiato»⁴⁸.

⁴⁸ D. Stone, *Counting: How We Use Numbers to Decide What Matters*, cit.