

# EPIDEMIOLOGIA DIGITALE

Molto di quello che facciamo e diciamo lascia una traccia digitale, dalle ricerche in rete, ai tweet, ai “mi piace” su Facebook. L’epidemiologia digitale è la scienza che permette di condurre studi epidemiologici (che tradizionalmente sono definiti come studi della distribuzione e dei determinanti delle situazioni o degli eventi collegati alla salute in una specifica popolazione<sup>1</sup>) e di applicare questi studi al controllo dei problemi, utilizzando dati provenienti dalle tracce che noi lasciamo quotidianamente interagendo su internet attraverso i social media e/o le App scaricabili sui nostri cellulari. Le fonti digitali dei dati, se ben sfruttate, possono fornire informazioni tempestive di grande interesse epidemiologico.

Il lavoro dell’epidemiologo è un po’ come quello dell’investigatore e quindi, con l’emergere di informazioni digitali sempre più disponibili a livello sia locale che globale, sono state affinate le tecniche di analisi e di utilizzo di questa grande mole di dati non strutturati disponibile in rete. Essere bloccati a letto da una malattia, aspettando di stare meglio, è abbastanza spiacevole e noioso. Cercare i sintomi online e leggere le voci specifiche su Wikipedia e WebMD, o inviare messaggi su Facebook e Twitter per raccontare come si stia male, può aiutare molti pazienti a trascorrere il tempo. Le conseguenze di queste azioni sono notevoli soprattutto se gli utenti descrivono i propri sintomi in tempo reale, fornendo anche informazioni su dove si trovano nel caso in cui abbiano attivata sui propri dispositivi la funzione di geolocalizzazione.

## Le tracce digitali della nostra salute

Questa è l’epidemiologia digitale: l’idea che la salute di una popolazione può essere misurata attraverso le tracce digitali che lascia su internet in tempo reale. Non si sostituisce all’epidemiologia tradizionale, ma ne rimodella i confini. In pratica, rappresenta un sistema di monitoraggio della salute collettiva che integra i sistemi di analisi dell’epidemiologia tradizionale con i dati digitali globali per costruire un quadro della diffusione e dei determinanti di una malattia in un momento specifico e fare previsioni di impatto a lungo termine.

Una delle applicazioni più popolari dell’epidemiologia digitale, soprattutto negli Stati Uniti, è stata Google Flu Trends. Lanciata nel 2008 per aiutare a prevedere epidemie influenzali,

si basava su un’idea molto semplice: quando le persone hanno l’influenza, spesso si rivolgono a internet per trovare informazioni sui loro sintomi<sup>2</sup>. Per molti anni Google Flu Trends ha rappresentato un esempio virtuoso di epidemiologia digitale e ha senza dubbio diffuso l’idea di utilizzare i dati digitali per ottenere informazioni epidemiologiche, ma ha anche dimostrato che questo non è un compito facile. Durante la stagione influenzale 2012-2013 Google Flu Trends ha infatti sovrastimato la prevalenza dell’influenza fino al 100% rispetto ai dati ufficiali della sorveglianza epidemiologica pubblicati dai Centers for Disease Control and Prevention (CDC)<sup>3,4</sup>. In seguito a questo errore, Google ha deciso di sospendere l’applicazione, lasciando disponibili solo i dati storici generati nei diversi Paesi<sup>5</sup>. Una critica importante mossa a Google come fonte di acquisizione di dati per le elaborazioni epidemiologiche, a prescindere dall’esperienza di Google Flu Trends, è stata che le sue stime non potevano essere riprodotte facilmente perché sia i dati grezzi sia l’algoritmo utilizzato per la selezione delle parole chiave e per l’estrazione non erano pubblici e scaricabili da altri ricercatori.

## Nuove fonti di dati e nuovi campi di applicazione

Il controllo di Google sull’accesso ai dati della propria piattaforma ha indotto i ricercatori interessati all’epidemiologia digitale a individuare altre fonti di dati, che fossero pubblici, accessibili e scaricabili da tutti. I dati generati da Twitter e Wikipedia hanno queste caratteristiche. Da quando l’epidemiologia digitale ha iniziato a utilizzarli, ha generato una serie di applicazioni in moltissimi campi. Si va dal monitoraggio e dalla sorveglianza delle malattie infettive per avere informazioni sulle epidemie e sulla diffusione delle malattie a livello globale, utilizzando le tracce digitali lasciate dai pazienti in ogni istante sulla rete<sup>6</sup>, ad applicazioni per capire se la diffusione di un certo comportamento, come smettere di fumare, può avere un impatto positivo sulla salute della persona. Questi dati consentono anche di sondare se la popolazione è favorevole o no alle vaccinazioni (il cosiddetto *sentiment*)<sup>7</sup> oppure di monitorare i fattori di rischio comportamentali per patologie legate all’obesità, al fumo, all’abuso di sostanze alcoliche, alla mancata attività fisica<sup>8-10</sup>.

Per chiudere il cerchio rispetto all'epidemiologia tradizionale, si stanno studiando alcune tecniche per mettere in atto misure di prevenzione e controllo delle malattie estrapolando i comportamenti a rischio degli utenti iscritti ai social media ed eventualmente inviando loro messaggi specifici che possano influire sul loro stile di vita, determinandone un cambiamento o un miglioramento<sup>11</sup>. È però molto complesso valutare in quale misura gli utenti seguano effettivamente le raccomandazioni contenute nei messaggi inviati.

Un'altra applicazione di notevole impatto è rappresentata dalla sorveglianza sugli eventi avversi ai farmaci<sup>12, 13</sup>. L'Agenzia europea dei medicinali (Ema) ha infatti suggerito la possibilità di utilizzare i dati provenienti dai social media, e in generale dai media, per individuare dati in grado di contribuire alla farmacovigilanza routinaria, di solito effettuata attraverso le segnalazioni dei medici o della popolazione<sup>13</sup>. Questi dati sono uno strumento potenzialmente molto importante perché consentono di catturare tutto ciò che viene avvertito dai cittadini come evento avverso in maniera automatica e tempestiva.

### Il limite del campione

I dati che l'epidemiologia digitale utilizza non provengono da un campione appositamente predisposto per la raccolta di informazioni a scopo epidemiologico. Non si ha quindi chiaramente a che fare con una coorte di pazienti selezionati in modo rappresentativo e ben controllata. I limiti di questo approccio sono in parte legati al cosiddetto *selection bias*, cioè al problema di selezionare adeguatamente il campione da analizzare. Da questo punto di vista la problematica da

affrontare è duplice: da un lato il segnale rilevabile dalle fonti di dati alle quali attinge l'epidemiologia digitale non è pulito, dall'altro nei Paesi in via di sviluppo l'accesso al web è limitato o comunque non in tutti i Paesi i social network sono utilizzati nello stesso modo. Per tale ragione, per esempio, i dati ricavati da Twitter funzionano bene per la sorveglianza dell'influenza in quei Paesi in cui questo tipo di social network è molto diffuso, ma non altrettanto bene in Paesi come l'Italia dove Twitter è ancora sottoutilizzato.

### L'integrazione con l'epidemiologia tradizionale

I vantaggi dell'epidemiologia digitale sono tali che il limite rappresentato dalle caratteristiche delle fonti di dati utilizzate è superabile. Diversi articoli scientifici hanno verificato, per esempio, che il dato sull'incidenza dell'influenza misurato contando il numero di volte che gli utenti menzionano su Twitter parole legate alle sindromi influenzali è strettamente correlato all'incidenza rilevata con la sorveglianza nazionale, effettuata dai medici sentinella che riportano settimanalmente il numero di casi di influenza diagnosticati tra i loro pazienti. Quindi è vero che il segnale di Twitter è rumoroso e non rappresentativo, ma l'abbondanza dei dati è tale per cui – aggregandolo a livello di Paese – la curva che si ottiene è strettamente simile a quella ottenuta attraverso la sorveglianza basata sui medici sentinella. E questo ci conforta perché – nonostante tutti i problemi legati al tipo di dati – il segnale che si ottiene è comunque robusto e affidabile. Lo stesso tipo di analisi è stata condotta anche per altri sistemi. Alcuni ricercatori statunitensi hanno dimostrato, per esempio, che negli USA il conteggio del numero di 'click' degli utenti di

Società Italiana di Farmacia Ospedaliera e dei Servizi Farmaceutici delle Aziende Sanitarie



**XXXVIII CONGRESSO NAZIONALE SIFOCARE**  
Il farmacista nel futuro del sistema salute  
ROMA, 23 – 26 NOVEMBRE 2017 HOTEL ROME CAVALIERI  
info: [www.congressosifo.it](http://www.congressosifo.it)



Wikipedia su articoli il cui titolo conteneva parole collegate ai sintomi influenzali generava una stima di incidenza molto simile a quella rilevata con la sorveglianza tradizionale dai CDC<sup>14</sup>.

Ma l'epidemiologia digitale non utilizza solo dati pubblicamente accessibili. Dati anonimi, provenienti dai telefoni cellulari, hanno permesso di costruire modelli di analisi per comprendere come il movimento delle persone influenzi la dinamica delle malattie infettive. È stato possibile misurare in che modo i modelli di viaggio umani diffondano la malaria e valutare rapidamente i movimenti di popolazione durante i disastri e le epidemie, come durante il terremoto e la successiva epidemia di colera ad Haiti nel 2010<sup>15</sup>.

Nonostante le criticità, le potenzialità che l'epidemiologia digitale porta a quella tradizionale e alla salute pubblica sono numerose e rilevanti. Anzitutto la disponibilità di dati in tempo reale consente di monitorare costantemente l'evolvere, per esempio, di un'epidemia o, nell'ambito della farmacovigilanza, di identificare segnali relativamente a eventi avversi a farmaci. L'idea, oltre che la sfida, per l'epidemiologia tradizionale è di integrare la sorveglianza routinaria con il digitale e con un approccio partecipativo da parte degli utenti per avere informazioni epidemiologiche molto accurate.

### Le sfide da affrontare

Grazie alla disponibilità di dataset amministrativi e di dati provenienti dal web e allo sviluppo di sistemi innovativi per il caricamento, la conservazione e l'analisi di grandi quantità di dati, buona parte dei quali non strutturata, l'epidemiologia si trova ad affrontare un cambiamento critico: riuscire ad integrare dati da sorgenti tradizionali e nuove per garantire una comprensione più chiara e accurata del mondo.

La sanità pubblica è tradizionalmente basata su dati raccolti da medici e operatori sanitari che, a loro volta, li ottengono da pazienti malati. Questo meccanismo include però solo quelle popolazioni che hanno facile accesso all'assistenza sanitaria o che decidono di rivolgersi al medico. E per lo più comprendono informazioni solo sulle malattie soggette a notifica. Un'ultima considerazione, non meno importante, riguarda il fatto che i sistemi tradizionali non permettono di raccogliere dati su comportamenti, sentimenti e opinioni sulla salute in tempo reale. In questo senso l'epidemiologia digitale può aggiungere nuove informazioni e riempire alcuni degli spazi vuoti. Naturalmente l'epidemiologia digitale

non cattura l'intera popolazione (obiettivo peraltro non raggiunto nemmeno dai metodi tradizionali per la raccolta dei dati epidemiologici), ma permette di rilevare e rispondere precocemente alle minacce di salute, costruendo sistemi di allerta digitale destinati a fermare le malattie emergenti che potrebbero diffondersi al livello globale<sup>16</sup>.

L'epidemiologia digitale, infine, pone anche importanti sfide etiche in materia di sorveglianza e privacy, che devono essere affrontate seriamente in quanto la popolazione è sempre più preoccupata per la mancanza di riservatezza e il possibile utilizzo improprio dei dati personali postati sul web.

**Caterina Rizzo**

*Dipartimento Malattie Infettive - Istituto Superiore di Sanità*

### BIBLIOGRAFIA

1. Last JM, A dictionary of epidemiology, Oxford University Press, 1988.
2. <http://static.googleusercontent.com/media/research.google.com/en/us/archive/papers/detecting-influenza-epidemics.pdf>
3. Butler D, When Google got flu wrong. *Nature* 2013; 494: 155-156.
4. [https://dash.harvard.edu/bitstream/handle/1/12016836/The%20Parable%20of%20Google%20Flu%20\(WP-Final\).pdf](https://dash.harvard.edu/bitstream/handle/1/12016836/The%20Parable%20of%20Google%20Flu%20(WP-Final).pdf)
5. <https://www.google.org/flutrends/about/>
6. <http://www.healthmap.org/en/>
7. Salathé M, Khandelwal S, Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol* 2011; 7 (10): e1002199.
8. Gregg R, Patel A, Patel S, O'Connor L, Public reaction to the UK government strategy on childhood obesity in England: A qualitative and quantitative summary of online reaction to media reports. *Health Policy* 2017; 121: 450-457.
9. Lipsky LM, Nansel TR, Haynie DL, Liu D, Li K, Pratt CA et al, Diet quality of US adolescents during the transition to adulthood: changes and predictors. *Am J Clin Nutr* 2017; Apr 26. doi: 10.3945/ajcn.116.150029
10. Bou-Karroum L, El-Jardali F, Hemadi N, Faraj Y, Ojha U, Shahrouh M et al, Using media to impact health policy-making: an integrative systematic review. *Implement Sci* 2017; 12 (1): 52.
11. Clar C, Dyakova M, Curtis K, Dawson C, Donnelly P, Knifton L, Clarke A, Just telling and selling: current limitations in the use of digital media in public health: a scoping review. *Public Health* 2014; 128 (12): 1066-1075.
12. European Medicine Agency, Farmacovigilance. [http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general\\_content\\_000258.jsp](http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_000258.jsp)
13. AIFA, Sanità digitale: AIFA e i big data del Servizio Sanitario Nazionale. <http://www.agenziafarmaco.gov.it/content/sanit%C3%A0-digitale-aifa-e-i-big-data-del-servizio-sanitario-nazionale>
14. McIver DJ, Brownstein J, Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Comput Biol* 2014; 10(4): e1003581.
15. Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K et al, Using mobile phone data to predict the spatial spread of cholera. *Sci Rep* 2015; doi:10.1038/srep08923
16. Brownstein JS, Freifeld CC, Madoff LC, Digital disease detection harnessing the web for public health surveillance. *N Engl J Med* 2009; 360: 2153-2157.