

EDITORIALE

DI STATISTICA PUBBLICA, OPEN E BIG DATA

Pau Origoni

Ufficio di statistica (Ustat)

L'offerta di dati non è mai stata così ampia come in questo momento: complici la sempre maggiore diffusione delle tecnologie dell'informazione e della comunicazione, si dispone oggi di una quantità di dati enorme che, in pochi clic, può essere sfruttata a fini statistici, dagli statistici ma non solo. In questo contesto sono andati affermandosi due nuovi concetti di cui si discute con una certa frequenza nel campo della statistica pubblica negli ultimi anni: *open data* e *big data*.

Gli *open data* fanno riferimento alla filosofia dell'*open government*, secondo la quale la pubblica amministrazione dovrebbe essere "aperta" e trasparente nei confronti dei cittadini, favorendo il più possibile la partecipazione (diretta) al processo decisionale. Un elemento chiave di questa visione sono i dati inerenti all'attività amministrativa, che vengono messi a disposizione del pubblico in un formato aperto e libero (nel rispetto della sfera privata), così da favorirne la condivisione e l'analisi. La Confederazione crede in questo concetto e proprio nel 2016 ha ufficialmente lanciato uno specifico portale, accessibile all'indirizzo <https://opendata.swiss/it>.

Il secondo concetto, quello dei *big data*, si riferisce al fatto che disponiamo oggi di grandi masse di dati che sono un prodotto (secondario) delle tecnologie digitali. Si pensi per esempio ai dati generati dai navigatori satellitari, dall'uso dei *social network* o della telefonia mobile. Si tratta di dati talmente corposi da rendere necessario mettere a punto strumenti innovativi per analizzarli. È un tema d'attualità anche per Ustat, visto che in questo numero della rivista ci sono due articoli che toccano questo tema: il primo, di Arnaboldi e colleghi, presenta un progetto incentrato su *big data*, mentre il secondo ne discute l'interesse, ma piuttosto in una prospettiva critica (v. Losa).

La domanda che si pone lo statista a questo punto è: l'affermarsi degli *open data* e dei *big data* rende il ruolo dello statista pubblico desueto? Possiamo considerare la nostra missione compiuta? Nei prossimi anni i dati saranno sempre più autoprodotti e condivisi (e analizzati) da tutti? Per il momento, la risposta è negativa; le vie indicate appaiono certo molto promettenti, ma ci sono una serie di aspetti critici da ritenere.

Per quanto concerne gli *open data*, gli elementi positivi sono numerosi: oltre ai risvolti etici relativi alla maggiore trasparenza dell'attività dello Stato, si tratta di una strategia che dovrebbe favorire lo sfruttamento di informazioni che – per motivi diversi – potrebbero rimanere inutilizzate, favorendo pure l'affermarsi di approcci analitici originali e innovativi. E questo avrebbe sicuramente delle ricadute positive per la statistica (oltre che per la democrazia). D'altra parte però, l'impressione è che questi discorsi, nella nostra realtà, siano oggi

avanguardistici. Affinché alla semplice messa a disposizione di *open data* faccia seguito uno sfruttamento consapevole e di qualità, va promossa una maggiore diffusione della cultura statistica, non intesa tanto come insieme di capacità tecniche di analisi, quanto piuttosto come forma mentale di estremo rigore nella valutazione corretta di cosa rappresentano i dati e cosa è possibile trarne. È una competenza ancora scarsamente diffusa nella realtà odierna, e gli esempi – anche recenti – non mancano. Si tratta di un lavoro ancora da fare, nel quale gli statistici potrebbero avere un ruolo importante, e che dovrebbe cominciare nelle scuole, con effetti per domani o, più probabilmente, per dopodomani. Questo lavoro costituisce un presupposto fondamentale per fare in modo che la messa a disposizione di dati *open* finisca per alimentare un circolo virtuoso e non l'inverso.

Anche per quanto riguarda i *big data* gli elementi positivi non mancano: questi dati possono portare alla produzione di informazioni nuove e originali, a basso costo, che riducono il carico sui rispondenti e i tempi di produzione. D'altra parte, i *big data* comportano anche qualche risvolto negativo. Si tratta in primo luogo di dati che – come quelli di natura amministrativa – non nascono con un obiettivo statistico. Questo ha delle implicazioni importanti e, prima di poterli usare, è necessario svolgere una serie di investigazioni volte a capire meglio la natura di ogni fonte e, conseguentemente, l'uso che se ne può fare. Un altro aspetto da considerare attentamente è poi quello della rappresentatività (un aspetto che viene invece toccato dall'articolo di Losa), in particolare quando si usano dati prodotti nell'ambito di iniziative che hanno anche una componente commerciale (dati dai *social network*, per esempio). Quest'ultimo aspetto può infatti introdurre delle distorsioni profonde in termini di gruppi toccati, rendendo di fatto le informazioni ottenute parziali o addirittura distorte. In questo senso, i *big data* hanno molti risvolti interessanti, sono estremamente utili, specialmente in alcuni ambiti tematici, ma hanno un ruolo complementare a quello della statistica pubblica. Anche in questo campo (innovativo) risulta dunque fondamentale il (tradizionale) rigore degli statistici pubblici per il metadato.

Sembrerebbe dunque che la statistica e gli statistici pubblici abbiano un futuro. Le avanguardie andranno sempre più considerate, perché foriere di miglioramenti sul piano dell'offerta e delle analisi, ma sono ancora oggi da considerare più come estensioni della statistica più che concorrenti. I progetti in questo senso non mancano, e torneremo presto a parlarne con la rivista.

Buona lettura a tutte e tutti!