

# Dati, grandi dati, dati granulari e la nuova epistemologia del giurista\*

Vincenzo Zeno-Zencovich

## Abstract

Il contributo analizza alcune delle questioni centrali nella discussione sui “big data” e cioè la titolarità di tali dati, la intersezione con la disciplina dei dati personali, le responsabilità connesse all’uso dei dati, le decisioni pubbliche che su di esse si fondano, e gli effetti che questo nuovo contesto ha sul modo con cui il giurista conosce ed organizza il diritto.

The article analyses some of the most important issues surrounding the big data revolution: entitlement of data; the overlappings with data protection legislation; liability for damages arising from the use of data; public decision-making processes on the basis of data analytics; and the effects that this new scenario has on the ways through which jurists understand and organize the law.

## Sommario

1. Premessa. – 2. La “proprietà” dei dati. – 3. I dati personali. – 4. La responsabilità connessa all’uso dei dati. – 5. Decisioni pubbliche basate sui dati. – 6. La mutante epistemologia del giurista.

Keywords: Big data, Titolarità dei dati, Dati personali, Responsabilità per dati, Amministrazione digitale

---

## 1. Premessa

Esiste un mondo reale, fatto da questo pianeta; dagli spazi, infiniti, extraterrestri; da tutto ciò che la Terra contiene di minerale, vegetale, animale, e dai suoi sette miliardi di abitanti. Ogni momento questo mondo reale cambia, da una foglia che cade ad un essere umano che muore o che nasce.

Ed esiste un mondo digitale in cui praticamente tutto il mondo reale è riprodotto attraverso strumenti che raccolgono, conservano, elaborano dati. La differenza è che questo mondo digitale è caratterizzato dal fattore “T”. Mentre nel mondo reale esiste il passato, nel mondo digitale ogni elemento è sempre presente perché ne conserviamo una traccia, dalla pioggia che è caduta un giorno alla telefonata che una persona abbia fatto ad un’altra. Questo mondo – che possiamo chiamare “datasfera” – è al tempo stesso dinamico e immanente, nel senso che ne registriamo e conserviamo ogni mutamento (o non mutamento) nel tempo, che può essere sempre richiamato. Nel mondo reale ciò che non è presente lo possiamo ricreare attraverso il ricordo, il racconto, le statistiche. Nel mondo digitale possiamo in ogni momento ricostruire, anche visiva-

\* L'articolo è stato inviato su richiesta della direzione e pertanto non è stato sottoposto a referaggio.

mente, quanto era e accadeva un'ora, un giorno, un anno fa.

Lungi dall'essere fantascientifica questa rappresentazione ha una concretezza fatta di miliardi di miliardi (attualmente l'unità di misura è lo zettabyte ovvero  $10^7$ ) di dati, ma soprattutto ha una sua interazione sociale: noi siamo, oltre ad esseri in carne ed ossa, milioni di dati; noi compiamo scelte sulla base di dati (banalmente la scelta di un percorso stradale o di un locale pubblico); le imprese e le istituzioni pubbliche prendono decisioni (anche immediate e contestuali) sulla base di dati.

Si tratta di un fenomeno in costante ed esponenziale crescita: secondo per secondo dati si accumulano ad altri dati; unendosi a quelli precedenti generano ulteriori dati. Esso è destinato ad aumentare in modo ancor più vertiginoso con la diffusione di quel che viene gergalmente chiamato “*Internet of the things*” (IOT) ovvero “Internet delle cose” in cui saranno gli oggetti – anche più minuti – a registrare il proprio utilizzo e l'ambiente circostante (già lo fanno gli autoveicoli) trasmettendo altri miliardi di dati ogni istante.

Il giurista, che osserva la società e studia le regole che la governano, non può restare indifferente a tale evoluzione, come non poteva restarlo quando si scoprivano nuovi mondi.

Quali sono i profili che maggiormente lo interessano, sia per la loro rilevanza pratica che per quella teorica? Qui di seguito cerco di fornire un elenco, nella consapevolezza che è solo parziale e provvisorio, anche perché il giurista ha una passione nel creare questioni da discutere e approfondire, talvolta puramente ipotetiche, talaltra estremamente concrete.

## 2. La “proprietà” dei dati

Miliardi di dati; ma “appartengono” a qualcuno? Il giurista formatosi nella tradizione romanistica è portato spontaneamente a farsi questa domanda perché dalla qualificazione della forma di appartenenza di un bene dipende la utilità – giuridicamente tutelata – che se ne può trarre.

Questi dati, dunque, “appartengono” a chi li genera? (*e.g.* il conducente del veicolo?) Al proprietario dell'oggetto che li genera (*e.g.* il proprietario del veicolo)? Al titolare della apparecchiatura che li registra (*e.g.* il produttore della “scatola nera” sul veicolo)? Oppure a chi li raccoglie e li elabora (*e.g.* la casa automobilistica verso la quale confluiscono i dati di tutti i veicoli da essa costruiti)? Gli esempi – tratti da una accesa contesa già in atto – evidenzia la molteplicità di possibili soluzioni. Ma, poi, come possiamo qualificare la forma di appartenenza? “Proprietà”, nel senso romanistico del termine, appare difficile per ragioni concettuali e comparatistiche. Da un lato sono ben note le impervietà teoriche quando si cerchi di applicare la disciplina della proprietà ed entità non materiali (*ex multis*, la “proprietà del credito”) con tutte le aporie riguardanti il modo di acquisto, di godimento, di trasmissione ed i relativi rimedi. Nel contempo, considerato che i “dati”, per la loro immaterialità e il loro intrinseco legame con le reti di telecomunicazione, non possono essere radicati nel territorio (e dunque nel diritto) di un solo Stato, va tenuto presente che in taluni ordinamenti (primo fra tutti quello tedesco) il

concetto di bene è normativamente legato alla corporeità dell'oggetto.

Più plausibile la prospettazione di una generica "titolarità" la quale attribuisce al soggetto – sulla scia della consolidata tradizione dei beni immateriali – una serie di diritti, facoltà e correlativi limiti e rimedi. Ma proprio l'esperienza dei beni immateriali ci ammonisce quanto l'affermazione di pretese giuridiche su di essi dipende da espressi riconoscimenti normativi, che nel caso dei dati (e con la esclusione di quelli "personali" e delle basi di dati di cui alla Direttiva 9/1996) ancora difetta. I dati godono di una qualche tutela ma indirettamente attraverso la disciplina della concorrenza sleale, dei segreti d'impresa, della responsabilità civile, del patrimonio dell'azienda.

Peraltro anche la prospettiva della "titolarità sui dati" presente una serie di problematiche.

In primo luogo va ricordato che i dati presentano le caratteristiche che gli economisti definiscono come "beni pubblici" ovvero sia la loro non consumabilità (il loro valore in generale è indipendente dall'uso che se ne fa: la data della scoperta dell'America può essere fruita da uno o da tutti, ma ciò non ne diminuisce il valore); la non-rivalità (la circostanza che uno sia "titolare" di un dato non impedisce – come invece per i beni corporali – che qualcun altro allo stesso tempo lo sia. Tali elementi ovviamente incidono sulla conformazione dei diritti che si vogliono affermare sul bene in funzione alle utilità che se ne intendono trarre. La esclusività, dunque, presenta profili peculiari. In secondo luogo di fronte ai miliardi e miliardi di dati, i quali rappresentano l'intero mondo, e lo duplicano digitalmente, la pretesa appropriativa rischia di apparire presuntuosa e utopistica, come quella di chi volesse appropriarsi delle gocce che compongono il mare o delle particelle che compongono l'aria. In altri termini, di fronte a grandezze di difficile commensurabilità e ad entità la cui circolazione e duplicazione è naturale e praticamente senza costo schemi generali di titolarità risultano di scarsa utilità ai fini pratici, ovvero sia quello di assicurare ad un soggetto – e solo a quel soggetto – il diritto di trarre un vantaggio economico dalla sua posizione.

Infine, e a tale ultimo riguardo, ci si può porre la domanda – ricorrente nel settore – se il valore non sia nei dati in sé, bensì sugli strumenti informatici (i c.d. *data analytics*) utilizzati per analizzare tali moli così grandi e variegate. In teoria tutti potrebbero avere accesso a tali dati, ma solo chi dispone degli algoritmi e dei programmi idonei. E' su questi ultimi che si afferma la privativa industriale.

### 3. I dati personali

Un indice del regime giuridico dei dati può essere dato dalla elefantia normativa europea, il Regolamento generale sulla protezione dei dati personali (679/16) composto di appena un paio di centinaia di "considerando" e di un centinaio di articoli, cui si aggiungono una corona di direttive (680/16 e 681/16), per altre centinaia di articoli, sul trattamento dei dati nelle indagini penali, dei dati nelle prenotazioni aeree e (in preparazione) sulle reti di telecomunicazioni. Da questo complesso è comunque chiaro che sui dati non si vanta un diritto dominicale. Senza scomodare Ulpiano (*dominus nemo membrorum suorum videtur*) il rapporto giuridico che si instaura è piuttosto di controllo

sull'uso che terzi possono fare di tali dati, assistito da uno *ius arvendi* che si concretizza nell'ordine di cancellazione dei dati illegittimamente trattati.

Peraltro proprio l'analisi della normativa e della prassi in materia di dati personali fa comprendere quanto la pretesa da parte del titolare di esercitare i diritti attribuitigli è veramente illusorio, giacché quei dati, una volta acquisiti sono utilizzati, trasferiti, elaborati, ceduti, duplicati infinite volte a sua insaputa, spesso in un'altra parte del globo, grazie alle ubique tecnologie c.d. *cloud*.

Un simulacro di titolarità, dunque, che è ancor più evidente con riferimento a dati non qualificabili soggettivamente o oggettivamente come "personali" per via o dell'ente cui si riferiscono ovvero di quanto in essi rappresentato.

Il rispetto, dunque, dei diritti individuali è rimesso non a istituti privatistici bensì a ben più penetrati controlli pubblicistici effettuati dai "Garanti per la protezione dei dati personali" i quali dispongono altresì delle armi della sanzione penale. Quello dei dati diventa uno, fra i tanti, "mercati regolamentati" cui è preposta una autorità e che opera secondo logiche di sistema. D'altronde, considerata la dimensione della "datasfera", è difficile immaginare che possa avvenire diversamente: la similitudine va al rapporto fra tutela privatistica della proprietà delle immissioni e disciplina pubblicistica dell'ambiente e di contrasto all'inquinamento.

#### **4. La responsabilità connessa all'uso dei dati**

Da quasi 40 anni ci si interroga sui profili di responsabilità civile connessi all'uso di dati: l'esempio archetipale era quello di un "sistema esperto" (si chiamavano così all'epoca) di assistenza al traffico aereo che, a causa di un errore di elaborazione, provocava un incidente. L'esperienza di questi decenni non fornisce elementi che ci consentano di affermare che tali sistemi (a partire dal più banale programma informatico, che assai semplicemente elabora i dati che immettiamo nel sistema, anche il più semplice, come quello del tasto "invio") siano privi di difetti e non provochino danni. Anzi, ripetutamente gli organi di informazione riportano, non si sa con quanta attendibilità, siffatti incidenti. Quel che però si constata è che, setacciando la giurisprudenza al di qua e al di là dell'Atlantico, e senza trascurare il Pacifico, il numero di casi nei quali l'universale istituto della responsabilità civile è stato chiamato in causa per raddrizzare i torti informatici è insignificante in confronto con altri settori. Il fenomeno va considerato attentamente: solitamente l'avvento di nuove tecnologie porta con sé il verificarsi di incidenti (si pensi ai treni prima e agli autoveicoli poi). In questo caso, invece, nonostante la onnipresenza degli strumenti informatici e dell'uso dei dati, questo non è successo. Le ragioni possono essere legate alla dimensione delle imprese, alla generalizzata copertura assicurativa ma soprattutto, e più verosimilmente, alla immaterialità dei processi produttivi del danno, alla difficoltà di ricondurla ad una errata o insufficiente elaborazione dei dati, e alla circostanza che il danno sovente consiste nella interruzione di attività (come nei classici casi di interruzione della energia elettrica).

Poco di nuovo dunque? L'elemento che va oggi considerato e sul quale vi è un notevole dibattito è rappresentato dagli algoritmi – che riescono a scavare, selezionare e orga-

nizzare milioni e milioni di dati – sulla loro genesi e costruzione. Soprattutto quando essi siano utilizzati non per conoscere il passato, ma piuttosto per orientare scelte future e/o prevedere con ragionevole approssimazione comportamenti futuri, costituisce dato acquisito che essi riflettono scelte valoriali del soggetto che li realizza ovvero li commissiona. La prima esigenza che si pone è quella della trasparenza di tali opzioni, onde evitare discriminazioni digitalmente orientate, tipicamente quelle di genere, di età o di razza.

La seconda, connessa, è quella della individuazione di parametri professionali nella elaborazione degli algoritmi. Attualmente essi sono assenti e dunque difetta uno “stato dell’arte” su cui fondare un giudizio di diligenza/negligenza. In particolare va considerato che essi sono solitamente inclusi nell’interno di programmi più articolati, la cui realizzazione dipende da molteplici produttori e assemblatori (ad es. “app” che orientano le scelte del visitatore in una certa località; sistemi di ricerca di soluzioni di viaggio; navigatori per il traffico stradale). Anche in questi casi l’emersione di profili di responsabilità appare estremamente difficile in primo luogo perché esso deriverebbe da scelte volontarie – ancorché indirizzate – dell’utente; ed in secondo luogo dalla sfuggente prova del danno. Di qui l’esigenza di interventi di natura preventiva idonei a formulare giudizi di responsabilità oggettiva.

### **5. Decisioni pubbliche basate sui dati**

L’ultima considerazione svolta nel paragrafo precedente porta ad analizzare il rilievo giuridico che hanno decisioni di soggetti pubblici effettuate sulla base di analisi di dati. Tali decisioni sono del genere più svariato e possono riguardare soggetti singoli, come gruppi di persone o comunità.

Nel primo caso – e si parla solitamente di “dati granulari” – l’esempio più ovvio sono le indagini informatiche per identificare soggetti che possono avere connessioni con il terrorismo o con altri gruppi criminali. Ma se ne trovano molteplici nel campo fiscale (di cui, per rimanere in Italia, gli “studi di settore” sono un esempio egregio). In taluni casi la analisi, anche predittiva, dei dati è regolamentata (Direttiva 680/16), in altri casi è coperta da una coltre di incertezza regolamentare, soprattutto laddove il soggetto pubblico, nell’esercizio della sua – doverosa – attività ispettiva, operi a campione.

Un altro settore di grandissima importanza pratica è quello della previdenza sociale nel quale piani contributivi e pensionistici vengono scelti, da lavoratori ma anche da imprese, sulla base di dati relativi al singolo individuo o alla singola impresa, ma anche dall’universo dei soggetti secondo previsioni fortemente influenzate dalla demografia e da politiche di spesa pubblica.

In tutti questi casi la verificabilità del processo di analisi appare indispensabile a tutela dell’indagato, del contribuente, del lavoratore, i quali possono subire gravissimi ed ingiustificati danni da qualche impostazione discrezionale che lo fa oggetto di indagine/ispezione (il soggetto cui viene negato l’imbarco in aeroporto perché rientrante in una categoria ipoteticamente rischiosa; l’impresa perfettamente in regola che viene tartasata da ispezioni fiscali o contributive).

Vi sono poi scelte pubbliche che riguardano interi gruppi (ad es. soggetti affetti da una certa patologia) o comunità (in particolare territoriale). Si pensi alla pianificazione urbanistica, alla rete dei trasporti, alla dislocazione dei servizi (scuole, ospedali): in una prospettiva di decisione informata è naturale che il soggetto pubblico si avvalga di sistemi di elaborazione di grandi moli di dati. Sarebbe, anzi, un vizio procedurale se non lo facesse. Ma è necessario comprendere quali dati e quali metodiche sono stati utilizzati. E' noto da quasi un secolo che le statistiche subiscono la influenza soggettiva di chi le raccoglie ed elabora. Lo stesso può dirsi della analisi dei dati i quali, ancora più facilmente, possono essere piegati per giustificare talune soluzioni.

La responsabilità dei soggetti pubblici si estende dunque ad una “*data accountability*” che richiede nuove forme di verifica ed accertamento. Di certo, in società democratiche evolute, non ci si può limitare – come una volta – a ritenere che la sanzione di scelte pubbliche errate sia esclusivamente politica, nel voto. Il procedimento partecipativo alla formazione delle decisioni e la loro sindacabilità passa, ora, inevitabilmente attraverso una verifica – che potrebbe anche consistere nell’opporre altri dati, altri algoritmi – digitale.

## **6. La mutante epistemologia del giurista**

L’ultimo riferimento ci porta ad un tema di grande rilievo anche se di teoria generale del diritto e di lunga portata.

La circostanza che il giurista operi in un mondo ed in contesto dominato dai dati cambia il suo modo di conoscere il diritto, di crearlo, di interpretarlo.

L’ipotesi che si sottopone alla discussione che nel mondo dei Big Data cambia l’epistemologia del giurista, ovvero sia il modo in cui egli organizza le regole, soprattutto nella tradizione dell’Europa continentale in cui è abituato a ragionare per categorie astratte e per norme di portata generale.

Il cambio di passo mentale consiste nel passaggio da una logica essenzialmente causale, in cui l’argomentazione – e dunque la concatenazione delle regole – segue un ragionamento consequenziale, ad una logica di tipo inferenziale.

Nel mondo dei grandi dati interessano non certezze bensì collegamenti probabilistici fra elementi fra di loro ontologicamente diversi. L’esempio più ovvio è quando scegliamo di acquistare un bene o un servizio sulla rete Internet ed immediatamente ci viene proposto l’acquisto di un bene o servizio che l’algoritmo governante inferisce poter essere di nostro interesse (“Coloro i quali hanno acquistato X hanno acquistato anche Y”). Come si applica questo procedimento al mondo del diritto? In primo luogo nella fase di produzione delle norme in cui si sostituisce ad un dato certo e apparentemente statico (le statistiche, appunto) la elaborazione dinamica di dati, anche in funzione predittiva, nella valutazione di impatto della nuova norma. Inferiamo – perché non siamo, né possiamo, essere certi – che una determinata norma porterà al conseguimento degli obiettivi che essa si prefigge, anche perché, sempre inferenzialmente, assumiamo che la condotta dei suoi destinatari sarà di un certo genere.

Ben inteso in questo non vi è nulla di intrinsecamente errato ed è anzi ben possibile che



il risultato sia ben più coerente rispetto a quello conseguito da una norma basata solo su assunzioni valoriali se non ideologici, anche non espressi.

Quel che va sottolineato è il diverso procedimento logico seguito che punta alla creazione di una norma che, probabilisticamente e predittivamente, coprirà il maggior numero di casi. Senza fare eccessivi sforzi di fantasia si pensi alla applicazione di questi criteri nel campo delle norme tributarie in cui il gettito è legato al grado di spontanea adesione dei contribuenti e alla difficoltà per questi di adottare strategie elusive.

Ma quel che colpisce di più è che la logica inferenziale non è utilizzata solo per comprendere ciò che è avvenuto, ma anche e soprattutto in funzione predittiva. La c.d. *predictive analytics* diventa dunque lo strumento attraverso il quale vengono prese decisioni che riguardano la collettività (legislative, regolamentari) o l'individuo (orientamento al lavoro, indagini penali). Gli esempi più evidenti sono l'uso che già ora viene fatto di tali strumenti nelle procedure di scelta di assunzione di un lavoratore o di una lavoratrice. Oppure nella stipula di un contratto di assicurazione e a quali condizioni. Con il che alcuni principi che sono fra i capisaldi del diritto contemporaneo, quello di eguaglianza e di non discriminazione, sono messi in crisi, e si torna ad un conflitto, le cui radici vanno ben oltre la filosofia e affondano nella religione, fra determinismo e libero arbitrio.