Big Data e Libertà nella dimensione digitale - Intervento di Antonello Soro, Presidente del Garante per la protezione dei dati personali

Big Data e Libertà nella dimensione digitale

Intervento di Antonello Soro, Presidente del Garante per la protezione dei dati personali
("GARRNEWS", 23 agosto 2018)
https://www.garrnews.it/rubriche-interne-18/speciale-18

Nel World Economic Forum di Davos, Angela Merkel ha affermato che il possesso dei Big Data segnerà le sorti della democrazia, della partecipazione e della prosperità economica.

1. Le 4 V

Quest'affermazione coglie indubbiamente uno dei dati più caratteristici della realtà attuale, così profondamente mutata dall'impatto che i Big Data hanno avuto sull'economia, sull'assetto politico e ordinamentale, sulla società, sul costume, sulla persona. Sono, forse, proprio queste ultime le implicazioni più sottovalutate della "rivoluzione" dei Big Data, che potrà avere effetti realmente positivi o negativi in ragione di quanto si porrà "al servizio dell'uomo", come recita il Regolamento protezione dati, con affermazione di valenza generale, in riferimento al trattamento dei dati personali.

Se, a proposito dei Big Data, si è parlato di quarta rivoluzione (pur con tutti i limiti che assumono questo tipo di periodizzazioni), è essenzialmente per gli effetti dirompenti determinati dalle loro caratteristiche addizionali di volume delle informazioni- di mole così rilevante da essere stoccate prevalentemente con il cloud- velocità dell'analisi cui sono sottoposti, eterogeneità di fonti, formato e struttura (potendo essere strutturati o meno), attendibilità.

È la regola delle 4 v: volume, velocità, varietà, veridicità, caratteristiche che ne generano una quinta: valore, profitto.

Ma l'innovazione principale dei Big Data consiste non solo nell'oggetto dell'analisi ma anche nel suo metodo: Big Data analytics e machine learning, che estraggono valore aggiunto superando i limiti computazionali cui eravamo abituati.

Queste caratteristiche hanno determinato un inimmaginabile progresso sulle tre dimensioni essenziali dell'estensione della realtà osservata, del tempo di analisi, della profondità della conoscenza.

Di qui la possibilità, dischiusa dai Big Data e straordinariamente innovativa, di sviluppare modelli interpretativi, analitici e anche predittivi di fenomeni e comportamenti umani, impensabili fino anche a solo pochi anni fa. Tali possibilità di utilizzo, a fini commerciali oltre che di utilità sociale, hanno conferito ai dati un valore ormai inestimabile, tali da renderli il petrolio dell'economia digitale, ma a differenza di questo suscettibili di rigenerazione continua e, anzi, auto-rinnovabili, in quanto tendenti ad autoalimentarsi con lo stesso uso che se ne faccia.

Le fonti di generazione di queste risorse preziose sono del resto molteplici: i social network in cui si proietta, più o meno scomposta, la nostra intera esistenza, le innumerevoli transazioni commerciali online, i flussi continui di dati alimentati dagli oggetti connessi dell'Internet of Things: dai giocattoli intelligenti ai vari dispositivi di domotica, dalle tecnologie indossabili agli apparati elettromedicali connessi al web.

La maggior parte delle informazioni utilizzate con la Big Data analytics sono cedute, dagli utenti della rete, con scarsa o nulla

consapevolezza degli effetti del loro atto dispositivo.

La quasi generalità dei servizi della società dell'informazione apparentemente gratuiti sono in realtà pagati da ciascun utente al prezzo – nient'affatto modico – dei propri dati personali, sfruttati dalle aziende per costruire profili di consumatori, indirizzarne le scelte, costruire bisogni del tutto indotti e plasmare così i comportamenti delle persone.

Torna, invertito, lo schema gramsciano dell'egemonia sovrastrutturale, che per il capitalismo del digitale risiede nella capacità di orientare scelte e comportamenti con la persuasione permanente.

Del resto, quando l'offerta è senza corrispettivo, il prezzo, o meglio il prodotto venduto sei tu: così Andrew Lewis descrive efficacemente la dinamica dell'economia digitale, ove i dati – molto più del bitcoin - sono divenuti la valuta con cui si acquistano beni e servizi al prezzo di frammenti più o meno importanti della nostra libertà.

2. I Big Data e le categorie della protezione dati

La dinamica di gestione dei Big Data ha, del resto, caratteristiche talmente innovative da scardinare le coordinate principali del diritto applicabile ai dati personali (analoghe considerazioni potrebbero farsi sul diritto d'autore e la proprietà industriale).

La nozione di titolarità del trattamento mostra, infatti, tutti i suoi limiti rispetto alla moltiplicazione dei gestori dei dati che caratterizza il processo di utilizzazione dei Big Data, lungo catene dagli anelli infiniti. I principi di minimizzazione, limitazione della finalità e conservazione per il solo tempo indispensabile alla realizzazione del trattamento non si attagliano a raccolte così massive di dati, acquisiti spesso non per esigenze attuali ma in vista di future, eventuali necessità e riutilizzati per fini ulteriori non sempre compatibili con quelli originari.

Sfuma, poi, come sottolinea il Parlamento europeo nella risoluzione del 14 marzo 2017, la distinzione tra dati sensibili e non, potendo i primi essere estratti combinando tra loro dati comuni.

La stessa nozione di dato anonimo (quale limite esterno delle garanzie accordate dalla disciplina di protezione dati) subisce una contrazione speculare all'estensione del concetto di dato personale, in funzione ampliativa della tutela.

Il GDPR, in particolare, valorizza la dimensione dinamica del dato personale, nella consapevolezza di come le potenzialità della Big Data analytics di estrarre informazioni che ci riguardano anche da semplici frammenti privi di correlazioni tra loro, aumenti a dismisura le possibilità di reidentificazione anche di dati in apparenza anonimi.

Ciò che conta nella realtà dei Big Data è, del resto, la possibilità di ricondurre un dato non tanto e non solo a una persona nominativamente identificata, quanto piuttosto a un profilo tale da determinare effetti significativi e, spesso, anche potenzialmente discriminatorii, in capo agli interessati.

In questo senso si muovono anche le proposte di revisione della Convenzione 108/81 del Consiglio d'Europa, nel cui ambito si precisa come la nozione di identificabilità non si riferisca esclusivamente all'identificazione in senso giuridico, ma a tutto ciò che consente di individuare e trattare un soggetto diversamente dagli altri (il single out).

Come osservato rispetto al confine tra dato personale e dato anonimo, la disciplina di protezione dati è uno tra i pochissimi settori dell'ordinamento a tentare di normare alcuni aspetti di questo fenomeno tanto dirompente quanto, altrimenti, sfuggente.

Pur informandosi al principio di neutralità tecnologica – per evitare di cristallizzare le norme in un determinato contesto tecnico, suscettibile di veloce obsolescenza - il Gdpr contiene, infatti, alcune norme e garanzie di particolare interesse per i trattamenti su larga scala quali quelli realizzati su Big Data.

Anzitutto, il criterio di applicabilità del Regolamento stesso anche a trattamenti svolti da imprese situate all'estero ma i cui servizi siano destinati a (o profilino) persone che si trovino nell'Unione europea.

Si tratta di un'innovazione importante, che consente di attrarre nella giurisdizione europea i big player dell'economia digitale, situati prevalentemente oltre-oceano e che accentrano nelle proprie mani la pressoché totalità dei Big Data, con la limitazione, che necessariamente ne consegue, delle garanzie dei cittadini rispetto all'uso dei loro dati e con gli squilibri e le asimmetrie nei rapporti

di forza che inevitabilmente ne derivano, sul piano geopolitico. Del resto, in una realtà, quale quella digitale, per sua stessa natura refrattaria ai confini di leggi e giurisdizioni, non possiamo più consentire forum shopping e dumping digitale: la tutela dei cittadini rispetto a un diritto, quale quello alla protezione dati, non può che essere uniforme e ugualmente garantita a prescindere da stabilimenti più o meno di comodo del titolare.

Rispetto all'attività svolta dalle multinazionali, poi, da un lato il criterio del one-stop-shop - con l'attribuzione, all'Autorità di protezione dati capofila, della competenza prevalente sul trattamento - consente una maggiore effettività dei controlli a fronte di minori oneri burocratici. Dall'altro lato, tuttavia, il temperamento determinato dal principio di prossimità consente ai cittadini di non doversi rivolgere, per ottenere tutela, all'Autorità del luogo di stabilimento del titolare, potendo invece adire l'Autorità di protezione dati o l'autorità giudiziaria del proprio Stato.

Innovative sono poi le garanzie adottate – tanto dal Regolamento quanto dalla direttiva 680 per i settori di polizia e giustizia penale - rispetto ai processi decisionali automatizzati sui quali si basa l'economia dei Big Data, assicurandone la contestabilità e la trasparenza della logica, dei criteri e delle sue conseguenze ed esigendo, almeno in ultima istanza, il filtro dell'uomo, per contrastare la delega incondizionata al cieco determinismo dell'algoritmo.

Rilevantissime, inoltre, le misure di privacy by design e by default, che mirano a inscrivere direttamente nei sistemi e nei dispositivi le tutele per i diritti dell'interessato, consentendo in tal modo di minimizzare i rischi del trattamento in ragione delle stesse caratteristiche organizzative e funzionali della tecnologia utilizzata.

Il che non significa, ovviamente, ridurre la protezione dati allo strumento, pena il rischio di discriminazioni per censo (beneficia della crittografia solo chi può acquistare l'iPhone), ma promuovere l'estensione generalizzata di tecniche di tutela che incorporino esse stesse garanzie adeguate per gli interessati.

Infine, il GDPR realizza un ragionevole equilibrio tra le esigenze di riutilizzo di dati su larga scala – in particolare di Big Data – per fini di utilità sociale con il diritto degli interessati alla protezione delle informazioni che li riguardano.

In tal senso, ad esempio, anche qualora non possa svolgersi su dati del tutto anonimi, l'archiviazione di dati per fini di ricerca, anche scientifica, è possibile previa adozione di misure, quali tra le altre la pseudonimizzazione, idonee a minimizzare l'impatto del trattamento sugli interessati.

A questa cornice sembra potersi ricondurre la norma della legge europea 2017 che ha legittimato il riutilizzo, per fini di ricerca scientifica o statistici, dei dati anche sensibili (purché non genetici), con l'adozione di tecniche di anonimizzazione o minimizzazione ritenute idonee a tutela degli interessati e previa autorizzazione del Garante.

Infine, la riscrittura del sistema sanzionatorio realizzata con il GDPR tenta anche di correggere, sia pur in parte, lo squilibrio determinato dall'ingresso dei big tech nel rapporto tra Stato e privati. Rispetto a imprese i cui fatturati sono spesso pari al Pil di molte nazioni, persino le più elevate sanzioni pecuniarie previste dagli ordinamenti nazionali si sarebbero rivelate del tutto inefficaci: di qui l'adozione del criterio della proporzionalità della sanzione pecuniaria al fatturato, che consente di modulare la misura sanzionatoria in ragione della capacità patrimoniale del titolare.

3. La dimensione collettiva dei rischi

La disciplina di protezione dati assicura, quindi, garanzie importanti ai diritti degli interessati nel contesto di trattamenti così invasivi quali quelli condotti sui Big Data. E tuttavia, gli "effetti collaterali" di questa particolare categoria di trattamenti non si esauriscono sul piano individuale, ma toccano aspetti più profondi delle dinamiche sociali, sui quali è bene riflettere.

L'accentramento della disponibilità dei Big Data nelle mani di poche aziende, già solo per questo oligopoliste, accentua lo iato tra valorizzazione economica e utilità sociale, con una serie di rischi che vanno dalla discriminazione sociale fondata sulla discriminazione algoritmica, alla eccessiva e irragionevole marginalizzazione del fattore uomo nei processi decisionali, per effetto della delega all'algoritmo di valutazioni che sono e devono restare squisitamente umane.

Criticità, queste, che caratterizzano anche l'uso dei Big Data nel settore pubblico.

Se, ad esempio, le attività di contrasto e di prevenzione si avvalgono sempre più della social media intelligence e del web scraping

(letteralmente: raschiare la rete, per rinvenirvi ovviamente qualche informazione utile in termini investigativi), si rischia non solo di ingenerare negli utenti un atteggiamento talmente difensivo da comprimere irragionevolmente la libertà di espressione, ma anche di affidare scelte investigative essenziali all'algoritmo.

Per evitare, dunque, la totale marginalizzazione dell'intervento umano in processi decisionali suscettibili di incidere poi sulla libertà personale e, per converso, sulla sicurezza pubblica e nazionale, è stato suggerito da più parti di coniugare la social media intelligence (socmint) e la human intelligence (humint) nella digital humint.

Per altro verso, le infinite possibilità dischiuse dall'editing genomico, se vanno promosse al fine di impedire lo sviluppo di patologie altrimenti inevitabili con la riscrittura di sezioni di genoma responsabili delle degenerazioni, esigono tuttavia una cornice etica e giuridica rigorosa di riferimento, perché un'attività di primaria utilità sociale non degradi ad eugenetica.

Ma soprattutto, si è dimostrato che gli algoritmi non sono matematica pura (come tale infallibile e neutra) ma piuttosto opinioni umane strutturate in forma matematica e riflettono quindi spesso, in misura più o meno rilevante, le precomprensioni di chi li progetta o le serie storiche assunte a riferimento.

Con il rischio, dunque, non soltanto di cristallizzare il futuro nel passato, leggendo sempre il primo con gli schemi del secondo, ma anche di assumere le correlazioni (quasi sempre contingenti) delle serie storiche considerate, come relazioni necessariamente causali.

Un algoritmo utilizzato negli Usa per il calcolo del rischio di recidiva penale si è dimostrato, ad esempio, incline ad assegnare – in assenza di ragioni criminologiche - un tasso maggiore ai neri rispetto ai bianchi, solo sulla base delle correlazioni desunte da una determinata serie storica assunta a riferimento.

Il risultato che si trae dall'impiego di tecnologie che dovrebbero assicurare la massima terzietà rischia dunque di essere, paradossalmente, più discriminatorio, lombrosiano o anche solo antistorico di quanto possa essere la pur fallibile razionalità dell'uomo.

Sono, questi, alcuni soltanto dei rischi - sul piano sociale, politico, etico - che un uso poco accorto dei Big Data può determinare.

Le linee guida del Consiglio d'Europa del gennaio 2017 colgono questo aspetto, integrando le valutazioni di protezione dati con alcuni standard etici minimi, secondo il modello Pesia (privacy ethical and social impact assessment). In particolare, si afferma che la valutazione del rischio da condurre su tali trattamenti deve assumere, quali parametri, non solo la protezione dati ma anche l'impatto etico e sociale, considerando dunque anche la dimensione collettiva del rischio cui i Big Data ci espongono. È questa, sicuramente, la direzione da seguire per rendere l'uso dei Big Data uno strumento di promozione dei diritti e del progresso sociale.

È una partita di importanza cruciale: in gioco vi sono i limiti che la "libertà e la dignità umana" impongono all'iniziativa economica privata (art. 41 Cost.) e il senso stesso che attribuiamo al rapporto tra individuo e mercato.

Ma vi è anche l'idea della democrazia in cui vogliamo riconoscerci, in quel difficile e sempre mutevole equilibrio tra libertà e sicurezza, individuo e collettività, che misura il grado di civiltà di un Paese.

Riprendendo le parole di Angela Merkel citate in apertura, allora, ciò che segnerà le sorti della democrazia sarà probabilmente non solo il possesso dei Big Data, ma la loro gestione nel rispetto dei diritti e delle libertà.