Modello abilitante l'elaborazione del dato all'interno del Repository Sanitario Nazionale

Elaborazione dei Big Data alla luce del GDPR

Tra i principi enunciati nell'art. 5 del Regolamento ci sono specifici dettati normativi che richiedono una riflessione nell'ambito del tipo di elaborazione del dato che viene effettuato con sistemi di Intelligenza Artificiale.

Il Principio di esattezza ed aggiornamento dati

Secondo il Regolamento, art. 5 lettera d), i dati devono essere *esatti* e, se necessario, *aggiornati* a testimonianza dell'attenzione che il Regolatore pone alla qualità del dato. Non ci sono dubbi infatti sulla rilevanza della qualità del dato nella creazione di sistemi di Intelligenza Artificiale e soprattutto, nella fattispecie, quando essi abbiano un'applicazione nell'ambito della salute. Il dettato normativo diviene però particolarmente impegnativo dal momento che richiede al titolare di adottare tutte le misure ragionevoli per cancellare o rettificare tempestivamente i dati inesatti *rispetto alle finalità per le quali sono trattati.* In sostanza questo principio comporta il divieto di usare tecniche di Big Data Analysis come se di facesse una sorta di "pesca a strascico".

La possibilità di trattare ed interrogare i dati anche in tempi diversi senza aver chiare le finalità specifiche a priori al momento della raccolta è una delle potenzialità più attrattive del Big Data Analysis e del Repository Sanitario Nazionale. A questo proposito si rimanda alla proposta di lettura abilitante al paragrafo "Diritto al controllo dei dati".

Il Principio di minimizzazione dei dati

Questo principio presente alla lettera c dell'articolo 5 del Regolamento prevede che i dati siano adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati («minimizzazione dei dati»). Anche l'applicazione di questo principio è dunque contro la "pesca a strascico" in quanto, se i dati devono essere minimizzati, è necessario che il titolare del trattamento abbia chiare le finalità.

A questo proposito oltre a rimandare nuovamente alla proposta di lettura del Regolamento al paragrafo "Diritto al controllo dei dati", si richiama anche al paragrafo 1 dell'art. 89 in materia di garanzie da adottare per i dati trattati ai fini di ricerca scientifica: "tali misure possono includere la pseudonimizzazione, purché le finalità in questione possano essere conseguite in tal modo."

Misure di sicurezza e data breaches

La sicurezza rientra tra i doveri specifici che il Regolamento pone in carico al titolare e le relative disposizioni sono contenute negli art. 32 "Sicurezza del trattamento", art. 33 "Notifica di una violazione dei dati personali all'autorità di controllo" e art. 34 "Comunicazione di una violazione dei dati personali all'interessato". Nella logica del GDPR la sicurezza dei dati è prima di tutto una condizione di legittimità del trattamento e una componente essenziale della responsabilità del titolare. La sicurezza assume anche un carattere funzionale a garantire la qualità dei dati, la loro affidabilità e il diritto alla interoperabilità. La pseudonimizzazione e la cifratura sono le misure metodologicamente da adottare specificatamente indicate dalla lettera a) dell'art. 32. Si tratta di metodologie da seguire piuttosto che a tecniche specifiche da utilizzare che sono rimesse al titolare che tiene conto delle modalità di trattamento e del livello di rischio che esse presentano. Il rischio va valutato sia complessivamente sia tenendo conto di ciascun anello della catena di trattamenti tipica dei processi di AI.

Sono necessarie poche parole per descrivere la necessità di un livello di sicurezza adeguato al trattamento dei dati sanitari raccolti all'interno del Repository Sanitario Nazionale. A questo proposito si richiama

quanto indicato dal paragrafo 3 dell'art. 32 nella parte in cui prevede che la conformità delle misure di sicurezza adottate e dei sistemi utilizzati, e la relativa costante verifica, possono essere oggetto anche di Codici di condotta e certificazioni approvati dalle Autorità di controllo o dai certificatori nominati da queste ultime.

Secondo quanto affermato dal Garante della Privacy nel suo discorso di presentazione della Relazione al Parlamento nel 2018, in ambito sanitario, l'incremento degli attacchi informatici ha toccato l'acme del 99% rispetto all'anno precedente, con effetti più gravi rispetto ad altri settori in quanto l'alterazione dei dati sanitari può determinare errori diagnostici o terapeutici[11]. La protezione dei dati è un fattore determinante l'integrità del dato e con essa l'efficienza sanitaria che si basa sulla correttezza del processo analitico fondato su big data.

Una possibile soluzione potrebbe essere quella di sfruttare le potenzialità della tecnologia Blockchain che, oltre a garantire una maggiore sicurezza dei dati clinici, consentirebbe di poter accedere a livello nazionale a tutti i dati di un paziente, e anche la gestione da parte del paziente stesso[12].

Codici di condotta e certificazioni, marchi e sigilli

I costruttori e i venditori di sistemi basati su algoritmi intelligenti sono chiamati a garantire il rispetto dei principi etici da parte dell'algoritmo e la conformità del trattamento dei dati al GDPR. Il Regolamento contiene due strumenti specifici a questo scopo che sono i Codici di condotta i quali definiscono, con l'approvazione delle Autorità di controllo e degli organismi da queste ultime designati, le *modalità* con le quali si intende applicare il Regolamento e le Certificazioni che consentono di ottenere dalle Autorità stesse e dai certificatori da queste riconosciuti, dichiarazioni che attestino che chi ne è in possesso opera in *conformità* al GDPR. Questi due strumenti condividono lo scopo di consentire alle associazioni e organizzazioni di categoria di definire le modalità specifiche che intendono adottare e le misure di garanzia e sicurezza che prevedono di utilizzare, non soltanto tenendo conto del livello di rischio e dei diritti degli interessati, ma anche delle specificità dei trattamenti posti in essere. L'art. 40 e 42 in materia di codici di condotta e certificazioni rispettivamente, stabiliscono che gli

Stati membri, le Autorità di controllo e la Commissione incoraggiano l'elaborazione di Codici di condotta e Certificazioni.

A questo proposito si auspica che il presente position paper possa essere un primo passo per instaurare un rapporto proattivo tra le imprese e operatori, che nei settori della Sanità e Ricerca Biomedica hanno il trattamento dei dati sanitari come elemento essenziale della loro attività, e le Autorità di controllo.
Lo stesso Considerando 99 richiede che *"nell'elaborare un codice di condotta, o nel modificare o prorogare tale codice, le associazioni e gli altri organismi rappresentanti le categorie di titolari del trattamento o di responsabili del trattamento dovrebbero consultare le parti interessate pertinenti, compresi, quando possibile, gli interessati, e tener conto delle osservazioni ricevute e delle opinioni espresse in riscontro a tali consultazioni".*
L'ambizione specifica è definire ed ottenere, con l'evoluzione del presente lavoro, l'approvazione delle modalità di trattamento del dato sanitario per le attività di sanità e ricerca biomedica che prevedono strumenti di AI sia sotto il profilo della corretta applicazione del GDPR (Codici di condotta) sia sotto il profilo delle garanzie adeguate da assicurare a tutti gli stakeholder (Certificazioni).

Inoltre è doveroso sottolineare che sia i Codici di condotta che le Certificazioni possono trovare adesione anche da parte di titolari del trattamento non soggetti al GDPR perché residenti in Paesi terzi o organizzazioni internazionali. Questa rappresenta una grande apertura per riconoscere le tutele e la legittimità dei trattamenti di dati da e verso Paesi terzi all'Unione Europea con grande vantaggio per le potenzialità di sistemi di AI che esprimono il massimo della loro potenzialità nel contesto della condivisione dei dati e nella collaborazione tra organizzazioni pubbliche e/o private anche di stampo internazionale.

Partnership pubblico-privata: il modello del payback del dato clinico

Il modello del Repository Sanitario Nazionale fonda le sue basi etiche sul principio della solidarietà mutualistica del dato sanitario tra i cittadini e sulla visione proprietaria del dato. Il cittadino, in qualità di proprietario dei propri dati sanitari, decide su base volontaria di cedere i propri dati, mantenendone pe-

rò il controllo e l'esercizio dei propri diritti regolamentati dal GDPR. La concessione del dato viene quindi a configurarsi come un contributo alla comunità e la stessa, compreso il cittadino stesso contribuente, riceve in cambio i benefici sanitari derivanti dallo sviluppo di tecnologie sanitarie basate su algoritmi creati dall'analisi dei dati sanitari presenti nel Repository Sanitario Nazionale. La visione proprietaria e contributiva del dato assume ulteriormente coerenza se si considera l'innegabile valore economico attribuibile al dato. A questo proposito è di particolare rilievo la sentenza della Corte di Cassazione in data 2 luglio 2018 (Cass. 17278/18) con la quale, in merito al consenso al trattamento dei dati personali prestato dall'interessato, si stabilisce che: *"nulla … impedisce al gestore del sito - … concernente un servizio né infungibile, né irrinunciabile-, di negare il servizio offerto a chi non si presti a ricevere messaggi promozionali… Insomma, l'ordinamento non vieta lo scambio di dati personali, ma esige tuttavia che tale scambio sia frutto di un consenso pieno ed in nessun modo coartato"*.

Il corollario più evidente derivante dalla decisione della Suprema Corte è il riconoscimento di un valore economico dei dati degli interessati che, prestando il consenso - purché libero ed informato - al trattamento dei dati personali in cambio di un determinato servizio, effettuano uno scambio. Il valore dei dati, in quanto tali e/o in forma aggregata (c.d. Big Data), è stato ribadito anche dal Garante della Privacy italiano, Antonello Soro, il quale già nel 2013 sottolineava che: "Non dovremmo permettere che i dati personali, che hanno assunto un valore enorme in chiave predittiva e strategica, diventino di proprietà di chi li raccoglie"[13].

La possibilità di cedere i propri dati sanitari, e con essi il valore economico associato, ad una organizzazione pubblica non evidenzia criticità etiche sin quando all'utilizzo dei suddetti dati consegua un prodotto sanitario di proprietà e utilizzo pubblico e quindi a beneficio della comunità di contribuenti che lo ha finanziato con i propri dati.

Tuttavia appare evidente come sia necessario trovare soluzioni etiche e giuridiche ad hoc nel caso in cui un'organizzazione privata abbia interesse ad attingere ai dati sanitari presenti nel Repository Sanitario Nazionale allo scopo di produrre soluzioni commerciali. D'altro canto limitare l'accesso dei dati sanitari alla sfera pubblica limiterebbe fortemente il potenziale del Repository Sanitario Nazionale escludendo la spinta degli investimenti privati.

In linea col principio della solidarietà mutualistica del dato sanitario e con la visione proprietaria dello stesso si ritiene che il dato sanitario contenuto nel Repository Sanitario Nazionale possa essere trasmesso - previa anonimizzazione ai sensi dell'art. 11 del GDPR - ad una organizzazione privata

per scopi di sviluppo industriale secondo il modello del payback del dato. Il dato reso successivamente anonimo fuoriesce dal campo di applicazione del GDPR poiché non comporta più la possibilità di identificare l'interessato.

La componente più innovativa nella presente prospettiva risiede non tanto nel rapporto tra interessato, titolare e responsabile, ottimamente normato nel GDPR a tutela dell'interessato, quanto nel concetto stesso di "payback". L'azienda interessata a raccogliere un certo tipo di dato per motivi di sviluppo industriale dovrebbe porre richiesta all'ente di governance del Repository Sanitario Nazionale che si pone come ente governativo a garanzia dei diritti del cittadino e dei doveri delle istituzioni stesse e della parte privata. Si pone quindi la questione della corrispondenza da parte privata a fronte dei dati ottenuti. Il cittadino quindi vedrebbe corrisposto il proprio contributo in termini di dato sanitario con un contributo diretto al cittadino stesso o in alternativa un contributo indiretto nei confronti della collettività intesa come Paese o sue Organizzazioni Sanitarie da parte dell'organizzazione privata che ne ha richiesto l'uso.

Nella considerazione secondo cui l'algoritmo può essere migliorato anche dopo la prima immissione in commercio proprio grazie alla capacità di apprendere da ulteriori dati, è necessario considerare che il dato sanitario mantiene il suo valore non solo al momento del prelievo dal Repository Sanitario Nazionale, ma anche qualora la macchina nel suo utilizzo diffuso raccolga dati da un paziente. In considerazione di questo aspetto, il dato raccolto dalle macchine intelligenti nel corso del proprio utilizzo dovrebbe necessariamente essere veicolato nel Repository Sanitario Nazionale e non all'azienda che lo ha immesso in commercio.

Un'ulteriore richiesta di trattamento da parte di un'organizzazione privata per aggiornare un algoritmo già in commercio dovrebbe quindi passare per una nuova richiesta e l'instaurazione di una nuova trattativa di payback del dato con i singoli interessati.

Appare chiara sin da subito la complessità necessaria per definire il valore del dato sanitario e la definizione di "contributo" il relazione al ritorno dell'investimento privato, ma allo stesso tempo si ritiene che lo sforzo necessario a definire queste norme ripagherebbe il sistema Paese con un formidabile strumento in grado di accelerare sviluppo e investimenti nel rispetto dei principi etici in un settore particolarmente delicato, ma contemporaneamente strategico, come quello della ricerca e sviluppo industriale in campo biomedico.