RIVISTA ITALIANA DI INFORMATICA E DIRITTO
PERIODICO INTERNAZIONALE DEL CNR-IGSG

ISSN 2704-7318
Fascicolo 1-2022
DOI: 10.32091/RIID0071
Articolo sottoposto a peer review

Una governance dei dati genetici per lo sviluppo della ricerca scientifica

Dimitri Martignago

Per la prima volta nel panorama normativo europeo, il GDPR ha introdotto il concetto di "dati genetici" che ha fra i principali campi di impiego quello della ricerca scientifica. Con l'obiettivo di contribuire ad un più ampio dibattito in materia, l'articolo intende affrontare la disciplina prevista dal dettato normativo per il trattamento dei dati genetici a fini di ricerca scientifica. Nello specifico, lo scopo dell'articolo è quello di esaminare ed approfondire l'insieme delle misure tecniche ed organizzative individuate dal legislatore europeo al fine di conciliare il rispetto dei diritti e delle libertà fondamentali degli individui con le necessità di sviluppo della ricerca scientifica e dunque le possibili deroghe ai diritti individuali. In particolare, partendo dalla disciplina introdotta dall'articolo 89 del Regolamento verranno fra l'altro presentate le tecniche di anonimizzazione e quelle di pseudonimizzazione per il trattamento dei dati genetici; il possibile spazio applicativo delle *distributed ledger technologies* e della *blockchain*, nonché quella parte di misure tecniche ed organizzative indicate dalle varie linee guida delle autorità nazionali per la protezione dei dati personali, oltre ai principi della *privacy by design* e *by default*. Un ultimo approfondimento sarà infine rivolto alle certificazioni e alla necessità di un impiego su scala internazionale di codici di condotta settoriali condivisi tra più Stati per superare le barriere nazionali che ad oggi rappresentano un freno alla ricerca genetica.

GDPR – Governance dei dati – Dati genetici – Ricerca scientifica

Sommario: *1. Introduzione – 2. L'anonimizzazione – 3. Misure tecniche e organizzative – 3.1. La pseudonimizzazione – 3.2. La protezione dei dati* by design *e* by default *– 3.3. Le misure di sicurezza e la valutazione del rischio – 3.4. La co-regolamentazione, tra codici di condotta e certificazioni – 4. Conclusione*

1. Introduzione

Gli sviluppi tecnologi e scientifici dell'ultimo decennio hanno consentito un'accelerazione improvvisa del progresso della ricerca genetica. La quantità di informazioni potenzialmente estraibili dai dati genetici li rende di estrema rilevanza tanto per ragioni sanitarie quanto per i risvolti nel contesto della ricerca scientifica. Per queste ragioni la tutela normativa dei dati genetici ha recentemente raccolto l'interesse di legislatori ed opinione pubblica con particolare riguardo alle questioni di carattere etico e pratico da risolvere, soprattutto a fronte del vertiginoso aumento dei test genetici diretti al consumatore (DTC) nonché dei potenziali rischi legati allo sviluppo della medicina personalizzata. La complessità e la sensibilità delle informazioni genetiche comportano un elevato rischio di abuso da parte di chi dispone dei dati e dei soggetti terzi in grado di identificare in maniera univoca i singoli interessati[1]. La quantità di informazioni potenzialmente estraibili da essi li rende di

D. Martignago si occupa di sviluppo dell'Innovazione nell'ambito della compliance e della responsabilità d'impresa in ReD OPEN S.r.l., spin-off partecipato dell'Università degli Studi Milano-Bicocca.

Questo contributo fa parte del numero speciale "La Internet governance e le sfide della trasformazione digitale" curato da Laura Abba, Adriana Lazzaroni e Marina Pietrangelo.

estrema rilevanza, soprattutto in virtù del fatto che buona parte del genoma di un individuo è condivisa anche dai suoi parenti. Questi rischi riguardano tanto le informazioni genetiche già estratte, quanto quelle potenzialmente estraibili da ogni genere di campione biologico. Ciononostante, il COVID-19 ed il rischio di nuove pandemie su scala globale ha evidenziato l'imperatività di una maggiore condivisione dei dati non solo per il progresso della ricerca scientifica, ma anche per favorire la proposta di nuovi trial clinici[2].

Nonostante il DNA di per sé non sia altro che un insieme di nucleotidi incatenati tra loro, le interpretazioni che vengono adottate e le informazioni da esso estratte possono determinare l'insorgere di discriminazioni o di violazioni di molteplice natura.

In considerazione della sostanziale riduzione dei costi di sequenziamento avvenuta negli ultimi anni, secondo recenti stime il mercato dei test genomici potrebbe subire una crescita repentina nei prossimi quattro o cinque anni, arrivando a toccare la soglia di 35,7 miliardi di euro. La conseguenza diretta di queste stime è che con il passare degli anni saranno sempre di più i soggetti coinvolti in maniera diretta o indiretta dal trattamento dei dati genetici, anche alla luce del fatto che i dati genetici riguardano tanto il soggetto da cui provengono quanto la sua parentela.

Dal 2016, con il GDPR, i dati genetici sono stati inclusi tra quelle categorie di dati particolari che, in virtù dell'articolo 9, non possono essere trattate a meno del ricorrere di una delle ipotesi previste al secondo paragrafo dello stesso. Ad esempio, laddove il trattamento sia condotto per finalità di ricerca scientifica questo dovrà avvenire, oltre che nel rispetto del principio di proporzionalità e dell'essenza del diritto alla protezione dei dati personali, adottando «delle misure adeguate e specifiche per tutelare i diritti fondamentali e gli interessi dell'interessato»[3] e in conformità con l'articolo 89, paragrafo 1 del GDPR. Quest'ultimo stabilisce dei generici standard di riferimento il cui rispetto è necessario non solo laddove ci si desideri avvalere della possibilità *ex* articolo 89(2) del GDPR di derogare ai diritti individuali per finalità di ricerca scientifica, ma, ancor prima, nei casi in cui per la stessa finalità, si vogliano trattare dati genetici in deroga al divieto imposto all'articolo 9 del GDPR.

Riconoscendo il trattamento per finalità di ricerca scientifica come una possibile eccezione al divieto di trattare i dati particolari, il Regolamento demanda ai singoli Stati membri, alle organizzazioni, alle associazioni e ai titolari del trattamento l'individuazione di una disciplina più specifica per quel che concerne sia le deroghe ai diritti individuali *ex* articolo 89(2), sia le misure con cui garantire il rispetto del Regolamento e la tutela dei diritti individuali. Per poter portare a termine un trattamento a fini di ricerca scientifica è dunque necessario tener presente che questo deve essere «soggetto a garanzie adeguate per i diritti e le libertà dell'interessato»[4], in conformità con il GDPR. Pur non precisando in maniera esaustiva in che cosa consistano dette garanzie[5], l'articolo 89(1) continua affermando che queste «assicurano che siano state predisposte misure tecniche e organizzative, in particolare al fine di garantire il rispetto del principio della minimizzazione dei dati»[6]. Se però la finalità di ricerca scientifica può essere conseguita «attraverso il trattamento ulteriore che non consenta o non consenta più di identificare l'interessato»[7], ossia con un dato anonimo o anonimizzato, la norma impone che venga conseguita in tal modo. In virtù di ciò, per poter trattare i dati a fini di ricerca scientifica la prima opzione da prendere in considerazione è l'anonimizzazione o comunque l'uso di dati anonimi. Solo una volta scartata questa ipotesi si potrà quindi valutare l'adozione delle altre misure tecniche e organizzative disponibili.

2. L'anonimizzazione

Non esiste un'interpretazione dei termini *anonimia*, *anonimizzazione* e *identificazione* che sia condivisa dalla comunità internazionale. Nella stessa Unione europea il concetto di "dato anonimo" può essere inteso in maniera differente[8].

Nel GDPR viene innanzitutto stabilito che il «regolamento non si applica [...] al trattamento di [...] informazioni anonime, anche per finalità statistiche o di ricerca»[9]. Per poter anonimizzare dei dati personali però il sistema deve anzitutto essere alimentato con dei dati personali che potranno esser considerati anonimi solo a seguito dell'applicazione delle tecniche di anonimizzazione più adeguate[10]. Per questa ragione, il trattamento con cui si procede a privare i dati personali di tutti gli elementi identificativi, ossia l'anonimizzazione, è da considerarsi senza alcun dubbio un trattamento ulteriore dei dati personali e in quanto tale sottostà alle regole imposte dal Regolamento e all'obbligo del previo consenso, salve le eventuali eccezioni.

Esistono due generi di tecniche di anonimizzazione: la randomizzazione, che consiste nell'attenuare il legame tra la persona e i dati che lo riguardano così che, noti i dati randomizzati, non si possa risalire all'individuo[11], e la generalizzazione «che consiste nel rendere meno dettagliati o diluire gli attributi delle persone interessate presenti in una tabella modificando la rispettiva scala o ordine di grandezza in modo che più righe di quella tabella presentino la stessa combinazione di attributi generalizzati»[12].

In ambito genetico le tecniche che hanno generato maggior interesse da parte della comunità scientifica sono ad ora la *differential privacy*, tra i metodi di randomizzazione, la *k-anonimity* e la *l-diversity*, tra quelli di generalizzazione[13]. In ogni caso le rilevanti problematiche connesse all'anonimizzazione, laddove applicata al contesto dei dati genetici, sembrerebbero non renderla mai preferibile da un punto di vista scientifico[14]. Dal momento che i dati genetici sono la rappresentazione biologica di un singolo individuo, anche laddove il loro collegamento con l'identità civile dovesse essere cancellato, rimarrebbero estremamente personali e questo renderebbe difficile, se non impossibile, pensare che possano essere anonimizzati come avviene con altri dati[15].

Le conseguenze riguardano tanto i ricercatori quanto gli interessati. Nell'ambito delle ricerche svolte dalle biobanche, ad esempio, l'efficacia degli studi sui tessuti dipende dalle cosiddette informazioni di follow-up e quindi dalla storia clinica aggiornata dei soggetti da cui sono stati prelevati[16]. In questo settore un dato anonimizzato rischierebbe quindi di vanificare gli sforzi fatti, non permettendo di risalire al soggetto originario.

Sulla base dell'opinione rilasciata dal WP29, sono stati delineati tre fattori da tenere in considerazione laddove si desideri fare affidamento sulle tecniche di anonimizzazione dei dati[17].

In primo luogo, il titolare deve tenere conto dei mezzi necessari affinché un dato possa essere reso anonimo, valutando costantemente le tecnologie disponibili in termini di capacità di calcolo, nonché gli algoritmi sviluppati in grado di invertire il processo e quindi rendere identificabile la persona, de-anonimizzando i dati.

Bisogna poi sempre considerare il fatto che molti dataset disponibili al pubblico, nonostante vengano definiti anonimi, in realtà non rispondono agli standard di sicurezza richiesti. La portata di questo requisito è di particolare importanza, soprattutto considerato che impone in capo al titolare l'obbligo di considerare non solo la propria capacità nel de-anonimizzare i dati, ma anche quella di altri soggetti, conosciuti o meno, che, considerato lo sviluppo tecnologico e le altre fonti di dati pubblicamente disponibili, potrebbero risalire al soggetto interessato.

Già nel 2013 Yaniv Erlich dimostrò che, utilizzando informazioni pubbliche, era possibile risalire ad alcuni partecipanti che avevano contribuito con il proprio DNA ad un progetto di ricerca scientifica[18]. Lo stesso Erlich, in uno studio condotto nel 2018[19], arrivò a dimostrare che, da un dataset di 1,28 milioni di campioni prelevati ad individui anonimi, era possibile risalire al 60% della popolazione statunitense con origini europee. Lo studio documentò inoltre come, laddove un database genetico arrivi a contenere i dati del 2% degli adulti appartenenti ad una determinata etnia, nonostante l'anonimizzazione, sarebbe quasi certa la possibilità di individuare una corrispondenza tra ognuno di questi individui e un cugino di terzo grado o più prossimo. In aggiunta, ulteriori ricerche condotte nella comunità scientifica hanno dimostrato come, anche in mezzo ai dati di molti individui, sia possibile recuperare i dati personali di un singolo soggetto risalendo ad essi da aggregati di polimorfismi a singolo nucleotide, anche laddove si faccia uso di soli 25 loci selezionati in maniera casuale[20].

Tornando ai fattori da considerare nel caso in cui si desideri anonimizzare un dato, il terzo, ed essenziale, elemento aggiuntivo individuato corrisponde all'obbligo per il titolare del trattamento di non fare affidamento sulle proprie intenzioni, dovendo piuttosto valutare la propria disponibilità di altri dati il cui riferimento incrociato con il dataset potenzialmente anonimizzato potrebbe identificare i soggetti a cui i dati fanno riferimento. Molti progetti di ricerca, infatti, non trattano solo dati genetici, ma ne fanno un uso combinato ad altri dati riguardanti fattori non genetici come le variabili legate allo stile di vita o dati socioeconomici[21]. Lo stesso concetto è infatti ribadito al considerando 26 del GDPR in base al quale, se le informazioni anonime sono da considerarsi come quelle che non si riferiscono ad una persona fisica identificata o identificabile, «[p]er stabilire l'identificabilità di una persona è opportuno considerare tutti i mezzi [...] di cui il titolare del trattamento o un terzo può ragionevolmente avvalersi per identificare detta persona fisica direttamente o indirettamente»[22].

Pur non potendo affermare che i dati genetici non possano essere anonimizzati in generale[23], è comunque importante sottolineare che, in questo settore, nessuno scienziato responsabile potrebbe garantire un rispetto assoluto della privacy, ed infatti le linee guida internazionali per la ricerca stanno prendendo le distanze dall'anonimizzazione per ragioni legate alla qualità dei dati, al ritiro dei partecipanti alle ricerche e per la necessità di divulgare le scoperte oltre che di collegare in maniera continuativa questi dati con quelli clinici o altri ancora[24].

Alla luce di ciò, della scarsa utilità derivabile dall'anonimizzazione genetica e dei costi che simili tecniche potrebbero comportare, sembrerebbe necessario andare oltre, interrogandosi in merito a quelle misure tecniche e organizzative che il GDPR impone di predisporre in generale al fine di «garantire il rispetto delle disposizioni del [...] regolamento»[25], dei diritti e delle libertà degli individui e, più nello specifico, così da «attuare in modo efficace i prin-

cipi di protezione dei dati, quali la minimizzazione»[26], e soddisfare «i principi della protezione dei dati fin dalla progettazione e della protezione dei dati di default»[27].

3. Misure tecniche e organizzative

L'obiettivo di garantire il rispetto delle disposizioni del GDPR e di assicurare le adeguate garanzie ai diritti e alle libertà individuali tramite l'adozione di misure tecniche e organizzative appropriate non trae origine dall'articolo 89 del Regolamento, ma costituisce piuttosto un tassello essenziale di quel concetto espresso tramite il principio di accountability[28], su cui si sviluppa buona parte del testo legislativo.

Se infatti la precedente normativa era «focalizzata sui diritti dell'interessato»[29] il testo del GDPR «si sviluppa essenzialmente su processi, attività, misure tecniche ed organizzative, sanzioni e obblighi rivolti a[l] titolare e [al] responsabile del trattamento»[30]. E sulla base della formulazione adottata dal legislatore europeo, si può poi notare come, lì dove la norma tratta con maggior riguardo il principio di minimizzazione dei dati rispetto agli altri principi della protezione dei dati, nel caso delle singole misure da implementare, è la pseudonimizzazione ad esser stata posta come punto di riferimento in svariati articoli[31], nonostante le molte questioni aperte in merito al suo utilizzo.

3.1. La pseudonimizzazione

Tra i metodi di de-identificazione, il termine pseudonimizzazione normalmente raggruppa quelle tecniche volte a sostituire con uno pseudonimo gli identificatori diretti, quali ad esempio il nome, il cognome e il numero di telefono, conservando però quelli che vengono definiti come quasi-identificatori o identificatori indiretti[32]. La possibilità che un identificatore sia in grado di condurre all'identificazione di un soggetto dipende sempre dal contesto. In certi casi, infatti, un nome e un cognome potrebbero rivelarsi poco utili a fronte dell'assenza di altri riferimenti in grado di fornire ulteriori informazioni. In ogni caso, se è vero che un collegamento tra gli identificatori diretti e gli pseudonimi deve sempre essere mantenuto per consentire una re-identificazione; al fine di poter considerare la pseudonimizzazione una scelta utile a garantire un alto livello di sicurezza dei dati, l'associazione tra identificatore e pseudonimo non deve essere tanto palese da consentire a chiunque abbia accesso ai soli dati de-identificati di risalire al soggetto.

Seppur vero che la pseudonimizzazione, come definita dal GDPR[33], può fornire il giusto supporto alla tutela dei dati personali, è d'altronde necessario tenere in considerazione che non tutte le tecniche utilizzabili sono ugualmente efficaci e valide, potendo alcune comportare un rischio perfino maggiore, in virtù della falsa sensazione di sicurezza che si verrebbe a generare negli individui.

In molti casi lo pseudonimo è generato da un terzo fidato spesso autorizzato a re-identificare lo pseudonimo per ricollegarlo al soggetto originario[34]. Per queste ragioni l'Agenzia europea per la sicurezza delle reti e dell'informazione (ENISA) ha contribuito a definire degli obiettivi generali di progettazione che ogni titolare del trattamento farebbe meglio a tenere in considerazione nella valutazione delle migliori tecniche di pseudonimizzazione[35]. Il primo consiste nel far sì che gli pseudonimi non consentano una facile re-identificazione da parte di terzi in uno specifico contesto di trattamento. Una volta sinceratosi di ciò, il titolare deve evitare di rendere troppo semplice per un terzo la riproduzione degli stessi pseudonimi[36].

Il WP29 ha individuato due categorie di risultati che possono derivare dalla pseudonimizzazione: gli pseudonimi che possono essere derivati dai valori originali di un attributo o di una serie di attributi[37] e quelli indipendenti dal valore iniziale, come un numero generato casualmente o un identificativo scelto dall'interessato. Più nello specifico ENISA ha poi contribuito a circoscrivere le 5 tecniche attualmente più utilizzate per pseudonimizzare un singolo identificatore[38].

Nella tecnica del *Counter* l'identificatore viene sostituito con dei numeri individuati da un contatore monotono. Il problema di questo metodo è però rappresentato dalle difficoltà nell'implementazione e nella scalabilità laddove applicato a dataset avanzati e di grandi dimensioni. Allo stesso modo la tecnica del *Random Number Generator* (RNG) e quella della *Cryptographic Hash Function* hanno un'utilità limitata dalle circostanze in cui vengono adottate. Nel primo, infatti, sono frequenti i casi in cui si verificano delle *collisioni*, ossia situazioni in cui ad un unico pseudonimo vengono associati due identificatori. Nel caso della funzione di hash invece, nonostante questa sia in grado di assicurare un alto livello di integrità dei dati, allo stesso tempo è spesso soggetta ad attacchi di forza bruta quali la *dictionary search* che possono metterne in mostra le vulnerabilità. La quarta categoria è rappresentata dalla tecnica del *Message Authentication Code* (MAC), che utilizza sempre una funzione hash, ma che differisce da quella precedentemente vista per l'impiego di una chiave segreta nella generazione dello pseudonimo, senza la quale non è possibile mappare gli identificativi e gli pseudonimi. Il MAC è tuttora considerato una valida

tecnica di pseudonimizzazione nel settore della tutela dei dati personali e presenta diverse varianti che possono essere adottate in base ai requisiti di scalabilità e funzionalità richiesti. Ultima delle cinque è infine la crittografia che può essere utilizzata come tecnica di pseudonimizzazione[39] e si distingue in simmetrica e asimmetrica. Un esempio di cifratura simmetrica applicata nel settore della genetica è quello che prevede l'uso dell'algoritmo di cifratura a blocchi, *TwoFish*[40]. per convertire il numero di previdenza sociale (SSN) in un"'alphabet-derived string". La cifratura asimmetrica è stata spesso impiegata per processi di pseudonimizzazione, soprattutto nel settore della salute. In un caso ad esempio, partendo da un'analisi di «previously published privacy concepts regarding a streamlined de-pseudonymization process and a patient-based pseudonym as applicable to research with genomic data [...] approaches»[41], si è riusciti a ricostruire una tecnica di pseudonimizzazione apposita per progetti di ricerca transnazionali, in grado di generare pseudonimi univoci, così da evitare collisioni, e facendo in modo che l'addetto alla de-pseudonimizzazione non corrispondesse al soggetto incaricato della pseudonimizzazione[42]. In un altro contesto invece, la crittografia asimmetrica è stata applicata per sviluppare un nuovo approccio per la gestione dei dati particolari, denominato *Polymorphic Encryption and Pseudonymisation*. Pensato appositamente per risolvere problematiche recenti che sorgono dal sempre più diffuso utilizzo dei big data tramite una struttura di pseudonimizzazione polimorfica, garantisce che ogni individuo abbia in automatico «different pseudonyms at different parties»[43] e che questi possano essere de-pseudonimizzati solo da partecipanti, che conoscono l'identità originale.

La scelta rispetto a quale tecnica di pseudonimizzazione utilizzare dipende in ogni caso da diversi parametri, tra cui ovviamente il livello di protezione desiderato e il vantaggio che si vuole ottenere dal dataset. Per di più, come visto nel caso delle tecniche di anonimizzazione applicate ai dati genetici e come anche affermato dall'ENISA, è bene tener presente che ci sarà sempre il rischio di una re-identificazione degli individui[44]. Può infatti darsi che un individuo, inizialmente non identificabile, lo divenga successivamente per via dell'aggiunta di ulteriori informazioni, collegate allo pseudonimo. Proprio per questa ragione il titolare deve avere ben chiaro lo scopo della pseudonimizzazione e deve selezionare la tecnica appropriata ad esso, in quanto un livello di protezione non adeguato sarà certamente in contrasto con i requisiti *ex* articolo 5 del GDPR.

Lo status dei dati pseudonimizzati resta una delle incertezze che più impattano sull'uso dei dati genetici[45] e, se la pseudonimizzazione può rappresentare una misura utile a garantire il rispetto del principio di minimizzazione, bisogna però tener presente che i vantaggi che possono derivare dal suo impiego sono legati al singolo contesto e che dunque potrebbero esserci soluzioni più adeguate.

Nel GDPR la pseudonimizzazione viene menzionata oltre che nell'articolo 89, anche negli articoli 25 e 32. In entrambi viene inclusa all'interno di quel macro insieme di misure tecniche e organizzative che, a seconda della situazione, possono rivelarsi più o meno adatte a «soddisfare i requisiti del [...] regolamento e a tutelare i diritti degli interessati»[46] in generale, e nello specifico a garantire «un livello di sicurezza adeguato al rischio»[47]. La loro utilità è quindi legata al contesto, e per questo è sempre bene tener presente il quadro generale dei requisiti imposti dal GDPR piuttosto che concentrarsi su una sola delle possibili misure.

3.2. La protezione dei dati *by design* e *by default*

L'inserimento della protezione dei dati *by design* e *by default* tra le misure tecniche e organizzative del GDPR ha rappresentato un passo enorme che per poter esser compreso richiede innanzitutto di prendere le mosse dai tre paragrafi dell'articolo 25 in materia di protezione dei dati *by design* e *by default*. Quest'ultimo, infatti, parte dal presupposto per cui le tecnologie digitali e i sistemi possono essere progettati in maniera tale da garantire la privacy e la protezione dei dati degli individui. Con un incipit che ricalca in buona parte quello dell'articolo 24 sul principio di accountability, il primo paragrafo individua i parametri che il titolare del trattamento deve tenere in considerazione nel valutare l'adeguatezza delle misure da mettere in atto per «attuare in modo efficace i principi di protezione dei dati, quali la minimizzazione» e di «integrare nel trattamento le necessarie garanzie al fine di soddisfare i requisiti del [...] Regolamento e tutelare i diritti degli interessati»[48]. L'elenco dei parametri non è da considerarsi esaustivo, ma consente comunque di condurre un'analisi di sostenibilità, utile ad orientare il titolare verso scelte adeguate, per assicurare la conformità del trattamento alla legge[49] e non solo ai principi di minimizzazione dei dati o di sicurezza[50].

In virtù del principio di *data protection by design*, la valutazione sull'adeguatezza delle misure tecniche e organizzative deve avvenire già in fase di progettazione ovvero «sia al momento di determinare i mezzi del trattamento sia all'atto del trattamento stesso»[51], in maniera dinamica e costante nel tempo. Il titolare del trattamento deve quindi rimanere

aggiornato sui più recenti sviluppi tecnologici e su come questi possano comportare dei rischi per la protezione dei dati nelle operazioni di trattamento. Il tutto deve essere tenuto in considerazione sin dalle fasi di organizzazione del trattamento, pianificando e spendendo sulla base dei costi necessari a far fronte ai rischi per i diritti e le libertà degli interessati, implementando ogni principio del Regolamento nel trattamento.

Il secondo paragrafo dell'articolo 25 è dedicato al principio della protezione dei dati *by default*. Se nel caso della tutela dei dati *by design* le misure tecniche e organizzative adottate devono essere volte all'implementazione di tutti i principi *ex* articolo 5 del Regolamento, nel caso della protezione *by default* il legislatore si è concentrato nello specifico su quei principi maggiormente correlati al concetto di necessità, ovvero la minimizzazione, la limitazione delle finalità e della conservazione, con l'intento di limitare il trattamento solo a ciò che è strettamente necessario.

Con una formulazione che riprende il testo dell'articolo 5(1)(c), il principio in questione persegue l'obiettivo di tutelare gli interessati proteggendo direttamente i loro dati e il modo in cui questi vengono trattati[52]. Per questa ragione in capo al titolare del trattamento spetta quindi l'obbligo di mettere «in atto misure tecniche e organizzative adeguate per garantire che siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento»[53].

Se è vero che la struttura dei principi di protezione dei dati *by design* e *by default* potrebbe sembrare all'apparenza identica, in realtà così non è. Mentre il primo principio è più orientato ai processi, il secondo è invece più incentrato sui risultati[54]. L'obbligo sancito al secondo paragrafo dell'articolo 25 è formulato in termini assoluti: il trattamento deve avere ad oggetto, per impostazione predefinita, solo quei dati necessari alla specifica finalità perseguita. Per questo, in tutti i casi in cui venga fatto uso di impostazioni di default, queste devono rispettare i requisiti di necessità correlati allo scopo prefissato.

Nel caso della protezione dei dati *by design* invece, l'imperatività dell'obbligo sancito dal legislatore è affievolita dall'insieme dei fattori che il titolare deve tenere in considerazione e bilanciare durante il processo decisionale relativo alle misure tecniche e organizzative da adottare[55]. Il tono adottato in questo caso dal legislatore sembrerebbe conferire all'obbligo normativo un carattere meno assoluto rispetto a quello visto sopra. In effetti la maggiore attenzione generalmente dedicata nel GDPR al principio di accountability e alla gestione del rischio[56] trova qui un contemperamento nello stato dell'arte e nei costi di implementazione, da valutare nella scelta delle misure adeguate con cui attuare in modo efficace i principi di protezione dei dati e integrare le garanzie necessarie per la tutela degli interessati e il rispetto del Regolamento.

Un approccio simile lo si ritrova poi nell'articolo 32 del GDPR, specificamente dedicato alla sicurezza del trattamento. Nonostante la formulazione, più congrua con il principio della protezione dei dati *by design*, è però bene precisare fin da ora che la norma in questione è collegata tanto alla protezione dei dati *by design* quanto a quella *by default*[57].

3.3. Le misure di sicurezza e la valutazione del rischio

Sulla base dell'articolo 32 del GDPR, il titolare del trattamento è oggi tenuto a mettere in atto «misure tecniche e organizzative adeguate per garantire un livello di sicurezza adeguato al rischio»[58] tenendo in considerazione lo stato dell'arte, i costi di attuazione e tutti i restanti fattori poc'anzi analizzati in merito alla protezione dei dati *by design*. L'adeguatezza del livello di sicurezza è rapportata ai potenziali rischi del trattamento e detti rischi sono stati suddivisi in tre diverse categorie, potendo essi comportare: la violazione della disponibilità, dell'integrità o della riservatezza dei dati[59].

Con riferimento alle modalità di valutazione del rischio[60] il Regolamento ha preferito affidarsi ad un disegno basato su tre moduli scalari piuttosto che su un unico procedimento valutativo di carattere generale. Il primo è rappresentato dalla generica analisi, svolta dal titolare o dai suoi responsabili sull'entità e la probabilità del rischio. Laddove questa non dovesse rivelarsi sufficiente, considerato l'elevato grado di rischio potenziale per i diritti e le libertà dell'interessato, l'articolo 35 prevede che il titolare debba portare a termine una valutazione d'impatto sulla protezione dei dati (DPIA). In ultima istanza, l'articolo 36 prevede l'obbligo di consultazione preventiva dell'autorità di controllo qualora dalla DPIA condotta dal titolare risulti un grado di rischio elevato in assenza di misure adottate per attenuarlo. Più alto è il rischio per il diritto dell'interessato, più rigorose devono essere le misure adottate dal titolare o dal responsabile del trattamento.

Anche l'articolo 32 distingue le misure di sicurezza adeguate al rischio tra tecniche e organizzative. Le prime consistono in quelle soluzioni demandate al lavoro di una macchina, un elaboratore, mentre le seconde sono affidate alle persone, che devono adottare comportamenti conformi a standard operativi. È poi lo stesso articolo a fornire un elenco esemplificativo

delle altre possibili misure di sicurezza da adottare[61]. Detta elencazione necessita però di essere sdoppiata per ciascun punto in: una regola tecnica e una regola organizzativa[62]. Quindi, nel caso della pseudonimizzazione o della cifratura, la loro implementazione dovrà avvenire sia con tecniche apposite, sia tramite «policies che impongono a funzioni aziendali di ricorrere a tali strumenti e di monitorare il ricorso a tali strumenti»[63]. Proprio per questa ragione, al fine di mettere in atto un sistema di sicurezza adeguato, è sempre essenziale procedere preventivamente a formare gli addetti al trattamento dei dati.

Ad oggi sono molte le autorità nazionali intente a fornire il proprio contributo nel tracciare una serie di linee guida generali per consentire alle organizzazioni e alle aziende di individuare il mezzo più adatto al proprio contesto, opponendosi all'illusione di una soluzione *one size fits all*.

Nel campo della ricerca genetica, ad esempio, è innanzitutto essenziale limitare l'accesso ai dati alle sole persone autorizzate, fino a quando questo si riveli necessario ai fini della ricerca stessa. È quindi essenziale prevedere dapprima una procedura di autorizzazione all'uso dei dati, definendo i differenti profili di autorizzazione ai sistemi e separandoli sulla base delle attività e delle responsabilità. Scaduto il contratto o conclusosi il periodo prestabilito, l'autorizzazione dovrà essere prontamente rimossa, così da garantire la riservatezza in maniera continuativa. Per questa ragione è bene effettuare una revisione annuale delle autorizzazioni, riallineando i diritti dell'utente alle sue funzioni. Una garanzia ulteriore sarebbe inoltre offerta, secondo il CNIL, da una politica di controllo degli accessi che sia specifica per i trattamenti messi in atto dall'organizzazione. Registrando parte delle azioni avvenute sui sistemi informatici si potrebbero infatti gestire gli incidenti o gli abusi, verificando l'origine degli accessi illeciti con *file* di registro o *log*, tramite il tracciamento.

Considerati i rischi di uso improprio dei dati è poi necessario integrare le normali misure per la deidentificazione e per l'acquisizione del consenso con ulteriori misure per una governance efficace, che solitamente si concretizzano in sistemi di controllo ulteriori per limitare l'accesso ai dati ai soli utenti autorizzati[64] e che coinvolgono tanto gli aspetti tecnici, quanto quelli manageriali, o organizzativi. Proprio in virtù di ciò, oggi per buona parte dei repository di dati genetici è previsto uno step preliminare, in base al quale un comitato apposito può acconsentire o meno a fornire l'accesso ai dati, ma solo dopo aver determinato se il progetto di ricerca presentato dal ricercatore rientri nell'ambito del consenso informato fornito dagli interessati. Purtroppo, un simile processo richiede spesso molti mesi per essere portato a termine e, pur garantendo agli interessati e alle biobanche di esercitare un controllo più preciso sui dati, ne limita la condivisione, non contribuendo a risolvere le preoccupazioni legate ad eventuali violazioni da parte dei ricercatori stessi[65].

Per far fronte a simili problematiche un insieme di nuove tecnologie tra cui la *blockchain*, dette *Distributed Ledger Technologies* (DLT), ha recentemente raccolto l'interesse di parte della comunità scientifica. Se infatti i metodi finora utilizzati sono stati contraddistinti da una particolare lentezza e in certi casi da uno scarso controllo effettivo sui dati stessi, la tecnologia sviluppata nel 2008 dal soggetto conosciuto con lo pseudonimo di Satoshi Nakamoto presenta delle caratteristiche che, nonostante i limiti ancora irrisolti, la rendono una valida soluzione per contribuire al progresso scientifico.

Le evidenti potenzialità alla base della *blockchain*, tra cui la decentralizzazione ed il controllo degli utenti sui propri dati[66], vanno però ponderate con alcune problematiche. Innanzitutto, il relativo panorama normativo è contrassegnato da una forte incertezza e frammentazione e mettere dei dati su una *blockchain* sarebbe una decisione difficile da revocare, che potrebbe richiedere sforzi particolarmente intensi, andando quindi in contrasto con il diritto all'oblio[67]. Per di più le DLT non sono state progettate per avere a che fare con «omics-sized data sets»[68] o con i dati derivanti dal sequenziamento di un intero genoma che possono arrivare ad occupare 0,5 terabyte per genoma, escludendo le immagini. Non per ultimo, nonostante aziende tra cui *Nebula Genomics* e *Encrypgen* facciano già uso della *blockchain* per acquisire e trasferire dati genetici, le strategie di monetizzazione alla base di simili modelli di business fanno sorgere nuove questioni etiche tuttora inesplorate.

Date le attuali incognite una certa cautela è quindi d'obbligo, senza però dimenticare che l'obiettivo da perseguire nel contesto europeo dovrebbe risiedere non nella creazione dell'ennesimo mercato europeo di test diretti al consumatore, ma piuttosto nell'applicazione delle DLT come uno strumento per i ricercatori che permetta agli individui di contribuire alla ricerca genomica[69], tenendo presente che i ricercatori europei si distinguono da sempre per gli sforzi compiuti nella sensibilizzazione rispetto ai pericoli e alle implicazioni etiche dell'uso dei dati genetici e del loro accesso su larga scala[70].

In conclusione, è poi da tenere a mente che sia l'articolo 32 che l'articolo 24, in tema di responsabilità del titolare del trattamento, affermano che, per dimostrare il rispetto degli obblighi sanciti in capo al titolare e la conformità ai requisiti *ex* articolo

32(1) poc'anzi analizzati, è possibile fare affidamento all'«adesione a un codice di condotta approvato di cui all'articolo 40 o a un meccanismo di certificazione approvato di cui all'articolo 42».

3.4. La co-regolamentazione, tra codici di condotta e certificazioni

I codici di condotta nel settore della tutela dei dati personali non rappresentano una novità. Ancor prima del GDPR, la Direttiva 95/46/CE, all'articolo 27, prevedeva una formulazione secondo cui era compito degli Stati membri e della Commissione incoraggiare «l'elaborazione di codici di condotta destinati a contribuire, in funzione delle specificità settoriali, alla corretta applicazione delle disposizioni nazionali di attuazione della presente direttiva, adottate dagli Stati membri»[71].

Se nel caso della Direttiva i codici avevano una funzione limitata, dovendo semplicemente facilitare l'applicazione della normativa e non integrarla o sostituirla, nel GDPR si assiste ad un importante cambio di direzione, proprio in virtù del fatto che tali codici sono oggi considerati degli strumenti di co-regolazione[72]. Stando alla definizione offerta dal Comitato europeo per la protezione dei dati personali (EDPB), i codici di condotta servono a dimostrare il rispetto del principio di accountability da parte di alcune categorie di titolari e di responsabili del trattamento, tramite la messa in pratica delle regole qui previste per la tutela dei dati personali[73]. Il contenuto è quindi costituito da norme che descrivono in maniera dettagliata i comportamenti più appropriati ed etici in un determinato settore[74].

Ormai da tempo si è avuto modo di evidenziare come la legislazione nazionale si sia rivelata inidonea a fronteggiare le sfide poste dal progresso tecnologico e dalla tutela dei dati personali[75]. Per queste ragioni è stata da più parti avanzata la possibilità di adottare dei codici di condotta settoriali approvati anche a livello internazionale. Una simile opportunità avrebbe il potenziale vantaggio di ridurre non solo l'incertezza legislativa ma anche i problemi connessi allo scambio di dati tra Stati, con particolare riguardo a temi quali il consenso informato, la giustificazione del trattamento dei dati senza il consenso dell'interessato o del trattamento di informazioni genetiche[76]. Ad oggi però i tentativi messi in atto per adottare un codice di condotta internazionale si sono rivelati inconcludenti, essendo stati tutt'al più pubblicati sotto forma di dichiarazioni non vincolanti[77]. Ciononostante, soprattutto in ambito genetico, questi sono visti come lo strumento in grado di apportare benefici concreti tanto nell'immediato quanto nel lungo periodo, potendo essere utilizzati per promuovere un vero ecosistema europeo per la ricerca biomedica congiunta e una più facile condivisione dei dati a livello internazionale[78].

Gli Stati membri, le autorità di controllo, il Comitato europeo per la protezione dei dati personali (EDPB) e la Commissione sono oggi esortate a incoraggiare le associazioni e gli altri organismi rappresentanti determinate categorie di titolari del trattamento o di responsabili, a elaborare codici di condotta, modificandoli e, nel caso, prorogandoli, per precisare l'applicazione del GDPR in funzione delle specificità del settore. Questo è il caso del Codice di condotta polacco approvato recentemente dalla *General Assembly of Biobanking and BioMolecular Resources Research Infrastructure* in Polonia[79].

Tra gli elementi che il Regolamento enumera a titolo esemplificativo come meritevoli di una particolare disciplina all'interno dei codici di condotta, meritano particolare attenzione «d) la pseudonimizzazione dei dati personali; e) l'informazione fornita al pubblico e agli interessati; f) l'esercizio dei diritti degli interessati» nonché «h) le misure e le procedure di cui agli articoli 24 e 25 e le misure volte a garantire la sicurezza del trattamento di cui all'articolo 32»[80]. Una disciplina apposita per ognuno di questi elementi all'interno di un codice di condotta adottato su larga scala, pur non costituendo una panacea di fronte ad eventuali conflitti tra normative nazionali, sarebbe comunque in grado di contribuire a tracciare delle valide linee guida vincolanti per l'armonizzazione della disciplina del trattamento dei dati genetici per finalità di ricerca scientifica. L'adozione di ulteriori codici deontologici promossi dalle autorità di controllo sarebbe inoltre un'alternativa valida tanto nelle circostanze «in cui il GDPR riserva agli Stati membri la libertà di mantenere o introdurre disposizioni più specifiche»[81], quanto nel caso dell'articolo 9, dove il testo del Regolamento rinvia «ai legislatori nazionali l'individuazione di garanzie apposite»[82].

Le disposizioni relative ai codici di condotta «si correlano a quelle sulla certificazione e sugli organismi privati a ciò abilitati»[83]. Anche le certificazioni, infatti, servono a dimostrare l'ottemperanza volontaria da parte del titolare e del responsabile alle norme e ai principi del regolamento. Il GDPR prevede che gli Stati membri, le autorità di controllo, il comitato e la Commissione sono tenuti a incoraggiare l'istituzione di simili meccanismi, oltre che di sigilli e marchi, relativamente alla protezione dei dati personali per dimostrare la conformità al Regolamento[84]. In questo caso il focus normativo è incentrato sulla dimostrazione della conformità piuttosto che sulla conformità in sé[85], ragion per cui l'articolo 42 e le

certificazioni sono da ritenersi una diretta conseguenza del principio di accountability[86], servendo a «dimostrare che un titolare del trattamento ha rispettato la disposizione e che quindi ha definito e attuato misure appropriate che sono state periodicamente sottoposte a revisione»[87].

Entrambe le normative, sia quella sui codici di condotta, sia quella relativa alle certificazioni, sono da ritenersi di estrema importanza soprattutto a fronte delle affermazioni della *Global Alliance for Genomics and Health* (*GA4GH*), secondo cui, già nel 2017, si riteneva plausibile che entro il 2025 sarebbe stato sequenziato il genoma di almeno 60 milioni di persone[88]. Recenti stime affermano infatti che ad oggi esistono 800 terabyte di dati genomici nel mondo[89]. Con l'obiettivo di massimizzarne l'utilità le istituzioni hanno però il bisogno di scambiare i propri dati al di fuori dei confini nazionali, una prassi rivelatasi evolutiva tanto per la ricerca contro le malattie rare quanto contro il cancro[90]. Lo scopo ultimo dei codici di condotta è di rispondere a delle problematiche in maniera accettabile e comprensibile tanto per il ricercatore quanto per il soggetto interessato dal trattamento. Benché la BBMRI-ERIC (Infrastruttura di Ricerca Europea delle Biobanche e delle Risorse BioMolecolari), ad esempio, nel 2017 abbia annunciato che avrebbe sviluppato un codice di condotta per l'intera Unione europea al fine di facilitare le dinamiche per il trattamento di dati legati alla salute[91] ad oggi questo non è ancora stato pubblicato.

Una circostanza in cui l'adozione di un codice di condotta o una certificazione possono rivelarsi utili è data dal caso in cui a un'istituzione che aggreghi i dati personali ricevuti da altre istituzioni venga richiesto di dimostrare l'esistenza e la validità del consenso al trattamento raccolto da un'altra istituzione. In questo caso appunto l'adozione di un codice di condotta sarebbe sufficiente a dimostrare che «the providing data controller follows the approved code of conduct and is certified under Article 42 of the GDPR»[92]. La mancanza di un organismo o di un'associazione che rappresenti davvero tutte le istituzioni di ricerca, la frammentazione e il contrasto tra gli interessi delle singole branche scientifiche e, infine, la sovrapposizione tra la ricerca scientifica e altri settori che possono richiedere il loro codice di condotta personale rappresentano però ad oggi un grosso freno all'adozione di un codice di condotta condiviso[93].

Alla luce di simili difficoltà, è quindi legittimo domandarsi se anche ora, nonostante i recenti traguardi conseguiti grazie al GDPR, la richiesta dei ricercatori di adottare delle linee guida uniformi per il trattamento di dati genetici continuerà a rimanere una questione irrisolta.

4. Conclusione

Per la prosecuzione e lo sviluppo della ricerca scientifica risulta oggi sempre più essenziale l'implementazione da parte dei titolari del trattamento delle misure tecniche e organizzative necessarie per attuare i principi di protezione dei dati e garantire il rispetto dei diritti e delle libertà degli individui. Questo risulta ancor più rilevante laddove si consideri che, a fronte dell'attuale e futuro progresso tecnologico, l'anonimizzazione dei dati genetici, oltre a rivelarsi poco utile per i ricercatori, sarà a dir poco irraggiungibile.

In attesa dell'adozione di strumenti di *soft law* flessibili e specifici per il settore della ricerca scientifica in grado di integrare le norme europee e nazionali, i titolari dovranno dunque tenere a mente che le tecniche efficaci oggi, un domani potrebbero non esserlo più, richiedendo ad essi un'opera di costante aggiornamento e un dispiegamento di forze non indifferente nel nome della privacy e della sicurezza.

Note

[1] C. Ziegenhain, R. Sandberg, *BAMboozle removes genetic variation from human sequence data for open data sharing*, in "Nature Communications", 2021, n. 12.

[2] J. Scheibner, J.L. Raisaro, J.R. Troncoso-Pastoriza et al., *Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis*, in "Journal of Medical Internet Research", vol. 23, 2021, n. 2.

[3] Art. 9, co. 2, lett. j), Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati).

[4] *Ivi*, art. 89, co. 1.

[5] C. Staunton, S. Slokenberga, D. Mascalzoni, *The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks*, in "European Journal of Human Genetics", vol. 27, 2019, p. 1159-1167.

[6] Art. 89, co. 1, Regolamento (UE) 2016/679.

[7] *Ibidem*.

[8] T. Finnegan, A. Hall, *Identification and genomic data*, 2017, p. 20.

[9] Considerando 26, co. 1, Regolamento (UE) 2016/679.

[10] K. El Emam, C. Álvarez, *A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques*, in "International Data Privacy Law", vol. 5, 2015, n. 1, p. 79.

[11] G. D'Acquisto, M. Naldi, *Big Data e Privacy by Design: Anonimizzazione Pseudonimizzazione Sicurezza*, Giappichelli, 2017, p. 67.

[12] *Ivi*, p. 81.

[13] S.S. Samani, *Assessing Disclosure Risks with Genomic Data*, A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities, 2018, p. 53-54; per ulteriori studi: J. Scheibner, J.L. Raisaro, J.R. Troncoso-Pastoriza et al., *op. cit.*; D. Froelicher, J.R. Troncoso-Pastoriza, J.L. Raisaro

et al., *Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption*, in "Nature Communications", 2021, n. 12.

[14] R. Cippitani, *Il trattamento di dati genetici a fini di ricerca scientifica*, in "Diritto e Processo", 2018, pp. 95-134, a p. 118.

[15] D. Hallinan, M. Friedewald, P. De Hert, *Genetic Data and the Data Protection Regulation: Anonymity, multiple subjects, sensitivity and a prohibitionary logic regarding genetic data?*, in "Computer Law & Security Review", vol. 29, 2013, n. 4, p. 317-329, a p. 322.

[16] M. Macilotti, U. Izzo, G. Pascuzzi, M. Barbareschi, *La disciplina giuridica delle biobanche*, in "Pathologica", 2008, n. 100, pp. 86-108, a p. 87.

[17] P. Quinn, L. Quinn, *Big genetic data and its big data protection challenges*, in "Computer Law & Security Review", vol. 34, 2018, n. 5, pp. 1000-1018, a p. 1003.

[18] E.C. Hayden, *The genome hacker*, in "Nature", vol. 497, 2013, p. 172-174.

[19] Y. Erlich, T. Shor, I. Pe'er, S. Carmi, *Identity inference of genomic data using long-range familial searches*, in "Science", vol. 362, 2018, n. 6415, pp. 690-694.

[20] M. Philips, *Can Genomic Data Be Anonymised?*, Global Alliance for Genomic Health, 10 October 2018.

[21] P. Quinn, L. Quinn, *op. cit.*, p. 1003.

[22] Considerando 26, Regolamento (UE) 2016/679.

[23] P. Quinn, L. Quinn, *op. cit.*, p. 1003.

[24] M. Phillips, B. Knoppers, *The discombobulation of de-identification*, in "Nature Biotechnology", vol. 34, 2016, n. 11, p. 1102-1103, a p. 1102

[25] Considerando 78, Regolamento (UE) 2016/679.

[26] *Ivi*, art. 25, co. 1.

[27] *Ivi*, considerando 78.

[28] *Ivi*, art. 5, co. 2 e art. 24.

[29] M. Siano, *Art. 24*, in G.M. Riccio, G. Scorza, E. Belisario (a cura di), "GDPR e Normativa Privacy Commentario", Wolters Kluwer, 2018, p. 237.

[30] *Ibidem.*

[31] V. artt. 25, 32 e 89 Regolamento (UE) 2016/679.

[32] M. Hintze, K. El Emam, *Comparing the Benefits of Pseudonymization and Anonymization Under the GDPR*, Privacy Analytics, 17 August 2017, p. 3.

[33] Art. 4, co. 1, n. 5, Regolamento (UE) 2016/679.

[34] A. Gabel, I. Schiering, *Privacy Patterns for Pseudonymity*, in E. Kosta, J. Pierson, D. Slamanig (eds.), "Privacy and Identity Management. Fairness, Accountability, and Transparency in the Age of Big Data", Springer, 2018, pp. 155-172, a p. 159.

[35] ENISA, *Recommendations on shaping technology according to GDPR provisions. An overview on data pseudonymisation*, November 2018.

[36] *Ivi*, p. 19.

[37] Article 29 Data Protection Working Party, *Opinion 5/2014 on Anonymisation Techniques*, 10 April 2014, p. 20.

[38] ENISA, *op. cit.*, p. 21-23.

[39] *Ibidem.*

[40] F.K. Dankar, A. Ptitsyn, S.K. Dankar, *The development of large-scale de-identified biomedical databases in the age of genomics – principles and challenges*, in "Human Genomics", 2018, n. 12, p. 16.

[41] H. Aamot, C.D. Kohl, D. Richter, P. Knaup-Gregori, *Pseudonymization of patient identifiers for translational research*, in "BMC Medical Informatics and Decision Making", vol. 13, 2013, n. 75, p. 1.

[42] *Ivi*, p. 10.

[43] E. Verheul, B. Jacobs, C. Meijer et al., *Polymorphic Encryption and Pseudonymisation for Personalised Healthcare. A Whitepaper*, Institute for Computing and Information Sciences Radboud University Nijmegen, 2016, p. 1.

[44] ENISA, *op. cit.*, p. 11.

[45] C. Mitchell, J. Ordish, E. Johnson et al., *The GDPR and Genomic Data. The impact of the GDPR and DPA 2018 on genomic healthcare and research*, PHG Foundation, 2020, p. 50.

[46] Art. 25, co. 1, Regolamento (UE) 2016/679.

[47] *Ivi*, art. 32, co. 1.

[48] *Ivi*, art. 25, co. 1.

[49] G. D'Acquisto, *Art. 25*, in G.M. Riccio, G. Scorza, E. Belisario (a cura di), "op. cit.", p. 247.

[50] A. Ciccia Messina, N. Bernardi, *Privacy e Regolamento Europeo*, 2017, Wolters Kluwer, p. 112.

[51] Art. 25, co. 1, Regolamento (UE) 2016/679.

[52] European Data Protection Supervisor (EDPS), *Opinion 5/2018. Preliminary Opinion on privacy by design*, 31 May 2018, p. 7.

[53] Art. 25, co. 2, Regolamento (UE) 2016/679.

[54] L.A. Bygrave, *Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements*, in "Oslo Law Review", vol. 4, 2017, n. 2, p. 105-120, a p. 116.

[55] Art. 25, co. 1, Regolamento (UE) 2016/679.

[56] A. Mantelero, *Il nuovo approccio della valutazione del rischio nella sicurezza dei dati*, in G. Finocchiaro (a cura di), "Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali", Zanichelli, 2017, p. 306.

[57] ENISA, *op. cit.*, p. 7.

[58] Art. 32, co. 1, Regolamento (UE) 2016/679.

[59] Gruppo di lavoro articolo 29 per la protezione dei dati, *Parere 3/2014 sulla notifica delle violazioni dei dati personali*, 25 marzo 2014, p. 5.

[60] A. Mantelero, *op. cit.*, p. 307.

[61] Art. 32, co. 1, Regolamento (UE) 2016/679.

[62] A. Ciccia Messina, N. Bernardi, *op. cit.*, p. 113.

[63] *Ibidem.*

[64] Nuffield Council on Bioethics, *The collection, linking and use of data in biological research and health care: ethical issues*, 2015, p. 78.

[65] B. Berger, H. Cho, *Emerging technologies towards enhancing privacy in genomic data sharing*, in "Genome Biology", vol. 20, 2019, n. 128.

[66] M. Alghazwi, F. Turkmen, K.J. van der Velde, D. Karastoyanova, *Blockchain for Genomics: A Systematic Literature Review*, arXiv, 2021.

[67] S. Thiebes, M. Schlesner, B. Brors, A. Sunyaev, *Distributed Ledger Technology in genomics: A call for Europe*, in "European Journal of Human Genetics", vol. 28, 2020, p. 139.

[68] *Ibidem.*

[69] *Ivi*, p. 140.

[70] European Society of Human Genetics, *Statement of the ESHG on direct-to-consumer genetic testing for health-related purposes*, in "European Journal of Human Genetics", vol. 18, 2010, p. 1271-1273.

[71] Art. 27, co. 1, Direttiva 95/46/CE del Parlamento europeo e del Consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati.

[72] A.R. Popoli, *Codici di Condotta e Certificazioni*, in G. Finocchiaro (a cura di), "Il nuovo Regolamento europeo sulla privacy e sulla protezione dei dati personali", Zanichelli, 2017, cit., pp. 394-395

[73] European Data Protection Board (EDPB), *Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679*, 12 February 2019, p. 6.

[74] Art. 40, co. 1, Regolamento (UE) 2016/679.

[75] D.D. Hirsch, *In Search of the Holy Grail: Achieving Global Privacy Rules Through Sector-Based Codes of Conduct*, in "Ohio State Law Journal", vol. 74, 2013, n. 6, pp. 1030-1069.

[76] M. Koščík, M. Myška, *Data protection and codes of conduct in collaborative research*, in "International Review of Law, Computers & Technology", vol. 32, 2018, n. 1, p. 141-154, a p. 143.

[77] *Ibidem.*

[78] F. Molnár-Gábor, J.O. Korbel, *Genomic data sharing in Europe is stumbling. Could a code of conduct prevent its fall?*, in "EMBO Molecular Medicine", vol. 12, 2020, n. 3, p. 1-7, a p. 1.

[79] D. Krekora-Zając, B. Marciniak, J. Pawlikowski, *Recommendations for Creating Codes of Conduct for Processing Personal Data in Biobanking Based on the GDPR art. 40*, in "Frontiers in Genetics", vol. 12, 2021, n. 12, p. 2189.

[80] Art. 40, co. 2, Regolamento (UE) 2016/679.

[81] R. D'Orazio, *Art. 40*, in G.M. Riccio, G. Scorza, E. Belisario (a cura di), "op. cit.", pp. 357-358.

[82] *Ibidem.*

[83] *Ivi*, p. 359.

[84] Art. 42, co. 1, Regolamento (UE) 2016/679.

[85] I. Kamara, R. Leenes, E. Lachaud et al., *Study on Certification Mechanisms under Articles 42 and 43 of the General Data Protection Regulation (GDPR) (EU) 2016/679*, 2019, p. 19.

[86] Gruppo di lavoro articolo 29 per la protezione dei dati, *Parere 3/2010 sul principio di responsabilità*, 12 luglio 2010, p. 18.

[87] *Ibidem.*

[88] E. Birney, J. Vamathevan, P. Goodhand, *Genomics in healthcare: GA4GH looks to 2022*, in "bioRxiv", 2017, p. 1.

[89] M. Phillips, F. Molnár-Gábor, H.O. Korbel et al., *Genomics: data sharing needs an international code of conduct*, in "Nature", vol. 578, 2020, p. 31-33, a p. 31.

[90] F. Molnár-Gábor, J.O. Korbel, *op. cit.*, p. 1.

[91] BBMRI-ERIC, *Code of conduct for using personal data in health research*, 2017.

[92] M. Koščík, M. Myška, *op. cit.*, p. 149.

[93] *Ivi*, p. 151.

* * *

Genetic data governance for the development of the scientific research

Abstract: For the first time in the European regulatory landscape, the GDPR has introduced the concept of "genetic data". A concept that is especially relevant in the context of scientific research. With the aim of contributing to a broader debate on the subject, this article aims to address the legal framework for the processing of genetic data for scientific research purposes. In particular, the article aims to examine and deepen the set of technical and organisational measures identified by the European legislator to reconcile the respect for the fundamental rights and freedoms of individuals with the needs of the development of scientific research and therefore the possible derogations to individual rights. Starting from the discipline introduced by Article 89 of the Regulation, the anonymisation and pseudonymisation techniques for the processing of genetic data will be presented, as well as the possible application of distributed ledger technologies (DLT) and blockchain This will be accompanied by a presentation of the technical and organisational measures set out in the various guidelines of national data protection authorities, as well as the principles of privacy by design and by default. A final in-depth analysis will be devoted to certification and to the need for international use of sectoral codes of conduct to overcome the national barriers that currently hinder genetic research.

Keywords: GDPR – Data governance – Genetic data – Scientific research