Privacy, Anonimato e Data Breach nel sistema sanitario

Salvatore Colomba – *ESTAR*

SPID ed Identità Digitale: Mercificazione, risorsa, opportunità o pericolo?

Privacy, Anonimato e Data Breach nel sistema sanitario

- Raccolta dati
- Anonimizzazione
- Tecniche usate (Randomizzazione e Generalizzazione)
- Analisi delle caratteristiche e dei punti di debolezza
- Anonimizzazione utile in caso di Data Breach

La raccolta di dati personali e sensibili in generale e in particola in ambito sanitario non può che avvenire rispettando le norme in materia di privacy italiane in primis ed europee.

Dati raccolti con il consenso del soggetto interessato e trattati dal titolare del trattamento. Dati acquisiti in chiaro e conservati nelle base dati aziendali.

Come gestire questi dati? In quale forma? Come renderli fruibili a terzi?

Una soluzione utile a questi quesiti può arrivare dalle tecniche di **anonimizzazione**.

Efficacia e i limiti delle tecniche di anonimizzazione

Esamineremo l'efficacia e i limiti delle tecniche di anonimizzazione esistenti rispetto al quadro giuridico dell'UE in materia di protezione dei dati fornendo anche raccomandazioni per l'impiego di tali tecniche, tenendo conto del rischio residuo di identificazione insito in ciascuna di esse.

L'**anonimizzazione** ha un rilevante valore potenziale, in particolare come strategia per consentire alle persone e alle aziende in senso lato di fruire dei vantaggi dei "**dati aperti**", attenuando al contempo i rischi per le persone interessate, restando consapevoli comunque di quanto sia difficile creare insiemi di dati effettivamente anonimi mantenendo nello stesso tempo tutte le informazioni sottostanti necessarie per le attività di analisi.

Anonimizzazione?

La direttiva 95/46/CE e gli altri strumenti giuridici rilevanti dell'UE, parlano di **anonimizzazione** come risultato del trattamento di dati personali volto a impedire irreversibilmente l'identificazione del soggetto trattato.

L'anonimizzazione costituisce quindi **un trattamento successivo dei dati** personali; in quanto tale, deve soddisfare il requisito di compatibilità tenendo conto delle motivazioni giuridiche e delle circostanze del trattamento successivo.

I dati resi anonimi non rientrano più nell'ambito di applicazione della legislazione in materia di protezione dei dati. Tuttavia il soggetto interessato potrebbe comunque avere diritto a tale tutela in base ad altre disposizioni (ad esempio, quelle che proteggono la riservatezza delle comunicazioni).

Nel procedere in tal senso, i responsabili del trattamento devono tener conto di svariati elementi e prendere in considerazione tutti i mezzi che "**possono ragionevolmente**" essere utilizzati per l'identificazione.

Anonimizzazione?

Esaminando alcuni studi sui processi di anonimizzazione e analizzando alcune pubblicazioni di ricerca emerge chiaramente che non è semplice ottenere un insieme di dati **effettivamente anonimo** a partire da un ampio insieme di dati personali, mantenendo al contempo tutte le informazioni sottostanti necessarie per espletare l'attività di analisi richiesta.

Ad esempio, un insieme di dati considerato anonimo potrebbe essere combinato con un altro insieme di dati in maniera tale da consentire l'identificazione di uno o più soggetti.

Tecniche di anonimizzazione

Prenderemo in esame principalmente due tecniche di anonimizzazione: la **randomizzazione** e la **generalizzazione**.

In particolare, nell'ambito delle due classi esamineremo l'aggiunta del rumore statistico, le permutazioni, la privacy differenziale, l'aggregazione, il *k*-anonimato, la *l*-diversità e la *t*-vicinanza, illustrandone principi, punti di forza e di debolezza.

Conoscere i principali punti di forza e di debolezza di ciascuna tecnica è utile per decidere come progettare un processo di anonimizzazione adeguato in un determinato contesto.

Pseudonimizzazione, dove collocarla?

Considereremo anche la **pseudonimizzazione** al fine di chiarire alcune insidie e convinzioni erronee:

la pseudonimizzazione non è un metodo di anonimizzazione.

Si limita a ridurre la correlabilità di un insieme di dati all'identità originaria di un soggetto interessato, rappresentando pertanto una misura di sicurezza utile.

Anonimizzazione

Per rendere anonimi determinati dati, gli stessi devono essere privati di elementi sufficienti per impedire l'identificazione del soggetto interessato.

I dati devono essere trattati in maniera tale da non poter più essere utilizzati per identificare una persona fisica utilizzando "l'insieme dei mezzi che possono essere ragionevolmente utilizzati" dal responsabile del trattamento o da altri.

Caratteristica importante è che il **trattamento deve essere irreversibile**.

La normativa non specifica come si debba o si possa effettuare il processo di anonimizzazione.

Importa il risultato: i dati devono essere tali da **non consentire l'identificazione** della persona interessata mediante "l'insieme" dei mezzi che "possono" essere "ragionevolmente" utilizzati.

L'anonimizzazione viene definita anche in norme internazionali quali **ISO 29100** come processo nel quale le informazioni personali identificabili (IPI) sono modificate irreversibilmente in modo tale che un titolare di IPI non possa più essere identificato direttamente o indirettamente, né dal singolo responsabile del trattamento di IPI né dallo stesso in collaborazione con altri (ISO 29100:2011).

Anonimizzazione

Un'**efficace soluzione** di anonimizzazione impedisce a tutte le parti di identificare un soggetto in un insieme di dati, di collegare due dati all'interno di un insieme di dati (o tra due insiemi distinti di dati) e di dedurre informazioni da tale insieme di dati.

Eliminare elementi direttamente identificanti non è sufficiente a garantire che l'identificazione del soggetto interessato non sia più possibile.

In generale è necessario adottare **ulteriori misure** per prevenire l'identificazione, a seconda del contesto e degli scopi del trattamento cui sono destinati i dati resi anonimi.

Anonimizzazione

Le due classi di tecniche di anonimizzazione, randomizzazione e generalizzazione dei dati, presentano dei punti deboli; tuttavia, ognuna di esse può rivelarsi adeguata, in circostanze e contesti specifici, per conseguire lo scopo desiderato senza compromettere la sfera privata delle persone interessate.

Per "**identificazione**" non si intende solo la possibilità di recuperare il l'indirizzo di una persona e/o il suo nome, ma anche la potenziale identificabilità mediante **individuazione, correlabilità e deduzione**.

Inoltre, l'applicabilità della normativa in materia di protezione dei dati non dipende dalle intenzioni del responsabile del trattamento o del destinatario. **Nella misura in cui i dati sono identificabili, si applicano le norme in materia di protezione dei dati**.

Anonimizzazione

L'anonimizzazione è una tecnica che si applica ai dati personali al fine di ottenere una **deidentificazione irreversibile**.

L'assunto di partenza è che i dati personali devono essere stati raccolti e trattati in conformità alla legislazione applicabile in materia di conservazione dei dati in un formato identificabile.

In tale contesto, il processo di anonimizzazione, inteso come trattamento di dati personali per ottenerne l'anonimizzazione, rappresenta un "**trattamento successivo**".

La base giuridica per l'anonimizzazione può quindi essere individuata in ciascuna delle motivazioni espresse nelle direttive e nelle leggi in materia di trattamento dei dati personali.

Anonimizzazione

Nel momento in cui si decide di ricorrere alle tecniche di anonimizzazione, i responsabili del trattamento devono tener conto di alcuni fattori-rischio:

- considerare i dati pseudonimizzati equivalenti ai dati resi anonimi;
- ritenere che dati adeguatamente anonimizzati privino i soggetti di qualsivoglia salvaguardia, anzitutto per il motivo che all'utilizzo di tali dati potrebbero essere applicabili altri atti legislativi;
- non considerare l'impatto sulle persone, in determinate circostanze, di dati adeguatamente anonimizzati, soprattutto nel caso della definizione di profili.

Anonimizzazione: Tecniche

Esistono diverse pratiche e tecniche di anonimizzazione che presentano gradi variabili di affidabilità. I responsabili del trattamento devono prenderle in considerazione nell'applicarle e in particolare deve essere posta parecchia attenzione sul livello di garanzia che una data tecnica consente di raggiungere tenendo conto dello stato attuale della tecnologia e dei tre rischi essenziali per l'anonimizzazione:

- ***individuazione***, la possibilità di isolare alcuni o tutti i dati che identificano un soggetto all'interno dell'insieme di dati;
- ***correlabilità***, *ossia* la possibilità di correlare almeno due dati concernenti la medesima persona interessata o un gruppo di persone interessate (nella medesima banca dati o in due diverse banche dati);
- ***deduzione***, vale a dire la possibilità di desumere, con un alto grado di probabilità, il valore di un attributo dai valori di un insieme di altri attributi.

Randomizzazione e Generalizzazione

Due diversi approcci all'anonimizzazione: il primo si basa sulla ***randomizzazione***, il secondo si basa sulla ***generalizzazione.***

Randomizzazione

La randomizzazione è una famiglia di tecniche che modifica la veridicità dei dati al fine di eliminare la forte correlazione che esiste tra i dati e la persona. Se i dati sono sufficientemente incerti non possono più essere riferiti a una persona specifica. La randomizzazione non riduce l'unicità di ogni dato, in quanto ciascun dato può comunque essere ancora estrapolato da un'unica persona interessata, ma può rappresentare una tutela dagli attacchi/rischi di deduzione e può essere affiancata da tecniche di generalizzazione per fornire maggiori garanzie di tutela della sfera privata.

La tecnica dell'aggiunta del **rumore statistico** si rileva utile in particolare nel caso in cui gli attributi possano avere un effetto avverso importante sulle persone e consiste nel modificare gli attributi contenuti nell'insieme di dati in modo tale da renderli meno accurati, mantenendo nel contempo la distribuzione generale.

Randomizzazione: Rumore statistico

L'aggiunta del rumore statistico deve essere spesso affiancata da altre tecniche di anonimizzazione, quali l'eliminazione degli attributi ovvi e dei quasi-identificatori (combinazioni di attributi relativi a una persona interessata o a un gruppo di persone interessate).

Caratteristiche:

- *Individuazione:* è ancora possibile individuare i dati riferiti a una persona (magari in modo non identificabile) anche se i dati sono meno affidabili.
- *Correlabilità:* è ancora possibile correlare i dati della stessa persona, ma i dati sono meno affidabili e pertanto un dato reale può essere correlato a un altro che è stato aggiunto artificialmente (ad esempio, per creare rumore statistico). In alcuni casi, un'attribuzione errata potrebbe esporre una persona interessata a un livello di rischio significativo e persino maggiore di una corretta.
- *Deduzione:* gli attacchi tramite deduzione sono possibili, ma la probabilità di successo è minore e potrebbero comparire alcuni falsi positivi (e falsi negativi).

Randomizzazione: Rumore statistico

Punti deboli:

- Evitare di aggiungere rumore statistico incoerente. Se il rumore statistico non è semanticamente plausibile (ossia “fuori scala” e non rispetta la logica tra gli attributi in un dato insieme), un attaccante che acceda alla banca dati potrebbe filtrare il rumore statistico e, in alcuni casi, rigenerare le voci mancanti.

- Presumere che l’aggiunta di rumore statistico rappresenti una soluzione a sé stante per l’anonimizzazione. L’aggiunta di rumore statistico è una misura complementare che ostacola il recupero dei dati personali da parte di un eventuale intruso.

Randomizzazione: Permutazione

La **permutazione** consiste nel mescolare i valori degli attributi all'interno di una tabella in modo tale che alcuni di essi risultino artificialmente collegati a diversi soggetti. Risulta utile quando è importante mantenere l'esatta distribuzione di ciascun attributo all'interno dell'insieme di dati.

La permutazione è considerata una forma speciale di aggiunta di rumore statistico. Nella tecnica classica di aggiunta del rumore, gli attributi vengono modificati mediante valori randomizzati.

La generazione di rumore statistico coerente può rappresentare un'operazione difficile da effettuare, mentre modificare solo marginalmente i valori degli attributi potrebbe non tutelare adeguatamente la sfera privata.

Le tecniche di permutazione modificano invece i valori contenuti nell'insieme di dati semplicemente permutandoli da un dato all'altro. Tali scambi garantiscono che gamma e distribuzione dei valori rimangano invariate, a differenza delle correlazioni tra valori e persone.

Randomizzazione: Permutazione

Se tra due o più attributi sussiste un legame logico o una correlazione statistica e gli stessi vengono permutati in maniera indipendente, tale legame verrà meno. Può pertanto essere importante permutare un insieme di attributi correlati in modo da non spezzare il legame logico, altrimenti un attaccante potrebbe individuare gli attributi permutati e invertirne la permutazione.

Ad esempio, se consideriamo un sottogruppo di attributi in un insieme di dati sanitari come "motivi del ricovero/sintomi/reparto/prestazioni", nella maggior parte dei casi tra i valori sussisteranno legami logici forti e la permutazione di un unico valore verrebbe pertanto individuata e potrebbe persino essere invertita.

Randomizzazione: Permutazione

Anche per la permutazione valgono le considerazioni fatte per il rumore statistico.

Caratteristiche:

- *Individuazione:* permane la possibilità di individuare i dati di una persona, ma gli stessi sono meno affidabili.
- *Correlabilità:* se la permutazione riguarda attributi e quasi-identificatori, potrebbe impedire una correlazione “corretta” di attributi a un insieme di dati sia internamente sia esternamente, ma consentirebbe comunque una correlabilità “non corretta”, in quanto un’immissione reale potrebbe essere associata a una persona interessata diversa.
- *Deduzione:* permane la possibilità di trarre deduzioni dall’insieme di dati, specialmente se gli attributi sono correlati o uniti da un legame causale logico forte. Non sapendo quali attributi sono stati permutati, un intruso deve tener conto del fatto che la propria deduzione si basa su un’ipotesi errata e quindi può ricorrere soltanto alla deduzione probabilistica.

Randomizzazione: Permutazione

Punti deboli:

- Evitare di selezionare l'attributo sbagliato: la permutazione di attributi non sensibili o non rischiosi non offre grandi vantaggi in termini di protezione dei dati personali. Se gli attributi sensibili/rischiosi fossero ancora associati all'attributo originario, un intruso potrebbe ancora estrapolare informazioni sensibili sulle persone.
- Permutazione casuale degli attributi: se tra due attributi sussiste una forte correlazione, la loro permutazione casuale non fornisce garanzie degne di nota.
- Evitare che la permutazione sia sufficiente: di per sé non garantisce l'anonimato e dovrebbe essere affiancata da altre tecniche, come l'eliminazione degli attributi ovvi.

Randomizzazione: Privacy differenziale

La privacy differenziale adotta un approccio diverso rispetto alle altre tecniche: mentre l'inserimento del rumore statistico interviene prima, al momento dell'eventuale pubblicazione dell'insieme di dati, la privacy differenziale può essere utilizzata quando il responsabile del trattamento genera opinioni anonimizzate di un insieme di dati e conserva al contempo una copia dei dati originali.

Le opinioni anonimizzate sono solitamente generate attraverso un sottogruppo di query per terzi specifici. Il sottogruppo presenta una certa dose di rumore statistico casuale aggiunto appositamente a posteriori.

La privacy differenziale suggerisce al responsabile del trattamento la quantità e la forma di rumore statistico che va aggiunto per ottenere le garanzie di tutela della sfera privata.

Randomizzazione: Privacy differenziale

Per contenere gli attacchi tramite deduzione e correlabilità occorre tenere traccia delle interrogazioni formulate da un soggetto e osservare le informazioni acquisite sulle persone interessate.

Per questi motivi le banche dati di “privacy differenziale” non dovrebbero essere utilizzate su motori di ricerca aperti che non offrono alcuna rintracciabilità dei soggetti che formulano le interrogazioni.

Caratteristiche:

- *Individuazione:* se vengono prodotte solo statistiche e i criteri applicati all’insieme dei dati sono scelte in maniera oculata, non dovrebbe essere possibile utilizzare le informazioni finali per individuare una persona.
- *Correlabilità:* utilizzando richieste multiple potrebbe essere possibile correlare le informazioni relative a una persona specifica tra due risposte.
- *Deduzione:* è possibile dedurre informazioni su persone o gruppi ricorrendo a richieste multiple.

Randomizzazione: Privacy differenziale

Punti deboli:

- Non aggiungere una quantità sufficiente di rumore statistico per impedire un collegamento con le conoscenze di base. La sfida consiste nell'aggiungere una quantità adeguata di rumore alle risposte vere in modo da tutelare la sfera privata delle persone e mantenere contemporaneamente l'utilità delle risposte pubblicate.

Generalizzazione

La **generalizzazione** rappresenta la seconda famiglia di tecniche di anonimizzazione e consiste nel generalizzare, o diluire, gli attributi delle persone interessate modificando la rispettiva scala o ordine di grandezza (ad esempio una regione anziché una città, un mese anziché una settimana).

Sebbene possa essere efficace per impedire l'individuazione, la generalizzazione non consente un'anonimizzazione che risulti efficace in tutti i casi; in particolare, presuppone approcci quantitativi specifici e sofisticati per impedire la correlabilità e la deduzione.

Generalizzazione: Aggregazione e k-anonimato

Aggregazione e *k*-anonimato

Le tecniche di aggregazione e *k*-anonimato sono volte a impedire l'individuazione di persone interessate mediante il loro raggruppamento con almeno k altre persone. A tale scopo, i valori degli attributi sono sottoposti a una generalizzazione tale da attribuire a ciascuna persona il medesimo valore.

Ad esempio, riducendo il grado di dettaglio di una località da città a Stato, si include un numero più elevato di persone interessate.

Questi metodi possono essere utilizzati nei casi in cui la correlazione di valori puntuali di attributi possa creare quasi-identificatori.

Generalizzazione: Aggregazione e k-anonimato

Caratteristiche:

- *Individuazione:* dato che i medesimi attributi sono condivisi da k utenti, non dovrebbe essere più possibile individuare una persona all'interno di un gruppo di k utenti.
- *Correlabilità:* benché la correlabilità sia limitata, permane la possibilità di collegare i dati per gruppi di k utenti. All'interno di tale gruppo, la probabilità che due dati corrispondano agli stessi pseudoidentificatori è pari a 1/k (che potrebbe essere significativamente più elevata della probabilità che tali informazioni siano non correlabili).
- *Deduzione:* il difetto principale del modello di *k*-anonimato consiste nel fatto che non protegge da alcun tipo di attacco tramite deduzione. In effetti, se tutte le k persone rientrano in uno stesso gruppo e se è noto a quale gruppo appartiene una persona, è piuttosto semplice recuperare il valore di tale proprietà.

Generalizzazione: Aggregazione e k-anonimato

Punti deboli:

- Con valore di k troppo basso, il peso di una persona all'interno di un raggruppamento è troppo significativo e gli attacchi tramite deduzione possono avere maggior successo;
- Trascurare alcuni quasi-identificatori: un parametro cruciale è la soglia di k. Più alto è il valore di k, maggiori sono le garanzie di tutela della sfera privata. La riduzione dei quasi-identificatori agevola la creazione di raggruppamenti di *k*-utenti;
- Non protegge bene dagli attacchi tramite deduzione;

Anno	Sesso	Codice P ostale	Diagnosi
1957	M	561*	Attacco cardiaco
1957	M	561 *	Colesterolo
1957	M	561 *	Colesterolo
1964	M	561 *	Attacco cardiaco
1964	M	561 *	Attacco cardiaco

se si conosce l'anno di nascita ad esempio 1964 si può dedurre il tipo di diagnosi e magari anche da dove proviene

Generalizzazione: L-diversità e T-vicinanza

L-L-diversità

La *l*-*l*-diversità amplia il *k*-anonimato per impedire gli attacchi tramite deduzione deterministica facendo sì che in ciascuna classe di equivalenza ogni attributo abbia almeno *l* valori diversi.

Obiettivo fondamentale è limitare la presenza di classi di equivalenza con una scarsa variabilità degli attributi, consentendo quindi eventuali attacchi con un grado di incertezza molto significativo.

La *l*-*l*-diversità è utile per proteggere i dati dagli attacchi tramite deduzione quando i valori degli attributi sono ben distribuiti.

Se vengono a mancare questi requisiti la *l*-*l*-diversità è soggetta ad attacchi tramite deduzione probabilistica.

La ***t*-vicinanza** rappresenta un affinamento della *l*-*l*-diversità nel senso che mira a creare classi equivalenti che assomigliano alla distribuzione iniziale di attributi.

Generalizzazione: L-diversità e T-vicinanza

La t-vicinanza è utile quando è importante mantenere i dati quanto più possibile prossimi a quelli originali.

Alla classe di equivalenza viene imposto un ulteriore vincolo:

- non solo devono esistere almeno *l* valori diversi all'interno di ogni classe di equivalenza, ma anche che ogni valore è rappresentato tante volte quante sono necessarie per rispecchiare la distribuzione iniziale di ciascun attributo.

Caratteristiche:

- *Individuazione:* la *l*-*l*-diversità e la *t*-vicinanza garantiscono che i dati relativi a una persona non possano essere individuati all'interno della banca dati.
- *Correlabilità:* la *l*-diversità e la *t*-vicinanza non rappresentano un miglioramento rispetto al *k*-*k*-anonimato per quanto riguarda la non correlabilità. Il problema è analogo a quello di ogni raggruppamento: la probabilità che le stesse informazioni appartengano alla medesima persona interessata è più elevata di 1/N (dove N rappresenta il numero di persone interessate nella banca dati).
- *Deduzione:* il principale vantaggio offerto dalla *l*-diversità e dalla *t*-vicinanza rispetto al *k*-anonimato consiste nel fatto che viene eliminata la possibilità di attaccare tramite deduzione una banca dati "*l*-diversa" o "*t*-*t*-vicina" con una sicurezza del 100%.

Pseudonimizzazione

La pseudonimizzazione consiste nel sostituire un attributo (di solito un attributo univoco) di un dato con un altro. La persona fisica potrebbe pertanto essere ancora identificata in maniera indiretta.

La pseudonimizzazione, se utilizzata da sola, non consente di ottenere un insieme di dati anonimo.

La pseudonimizzazione riduce la correlabilità di un insieme di dati all'identità originale di un soggetto; è dunque una misura di sicurezza utile, ma non un metodo di anonimizzazione.

Il risultato della pseudonimizzazione può essere indipendente dal valore iniziale (ad esempio usando un numero casuale o nome scelto dalla persona interessata) o può essere prodotto dai valori originali di un attributo o insieme di attributi usando una funzione di hash o un sistema di crittografia.

Pseudonimizzazione: Tecniche

Tecniche di pseudonimizzazione:

- **crittografia con chiave segreta**: chi conosce la chiave può facilmente risalire all'identificazione di ogni persona interessata decrittando l'insieme di dati, poiché i dati personali sono ancora contenuti all'interno dell'insieme di dati, anche se in forma crittografata;
- **funzione di hash**: una funzione che, a partire da un insieme di dati di qualsiasi dimensione (l'insieme potrebbe essere costituita da un unico attributo o da un insieme di attributi), restituisce un valore di dimensione fissa; tale funzione non può essere invertita, cioè non esiste più il rischio di inversione associato alla crittografia.
  Tuttavia, se l'intervallo di valori di input relativi alla funzione di hash è noto, la funzione stessa consente di riprodurli al fine di desumere il valore corretto associato a un dato specifico. Ad esempio, se un insieme di dati è stato pseudonimizzato effettuando l'hashing del codice fiscale, lo stesso può essere estrapolato semplicemente effettuando l'hashing di tutti i possibili valori di immissione e raffrontando il risultato con i valori contenuti nell'insieme di dati;

Pseudonimizzazione: Tecniche

- **funzione di hash cifrato con chiave memorizzata**: una funzione di hash particolare che utilizza una chiave segreta quale input aggiuntivo;
- **crittografia deterministica o funzione di hash cifrato con cancellazione della chiave**: questa tecnica può essere equiparata alla selezione di un numero casuale quale pseudonimo di ciascun attributo contenuto nell'insieme di dati seguita dalla cancellazione della tabella delle corrispondenze.
- **tokenizzazione**: tecnica applicata solitamente nel settore finanziario per sostituire i numeri delle carte d'identità con valori che presentano un'utilità ridotta per un eventuale intruso.

Pseudonimizzazione: Tecniche

Caratteristiche:

- *Individuazione:* permane la possibilità di individuare i dati delle persone, in quanto queste ultime sono ancora identificate da un attributo unico che è il risultato della funzione di pseudonimizzazione (= l'attributo pseudonimizzato).
- *Correlabilità:* la correlabilità rimane un'operazione di semplice effettuazione tra dati che utilizzano lo stesso attributo pseudonimizzato per fare riferimento allo stesso soggetto, soprattutto se non è stato eliminato ogni legame tra l'attributo originario e quello pseudonimizzato.
- *Deduzione:* gli attacchi all'identità reale di un soggetto tramite deduzione sono possibili all'interno dell'insieme di dati o tra diversi insiemi di dati che utilizzano lo stesso attributo pseudonimizzato per una persona, oppure se gli pseudonimi sono molto evidenti e non mascherano adeguatamente l'identità originale della persona interessata.

Pseudonimizzazione: debolezze

Nome, Indirizzo e CF	Esenzione	Validità	Numero di riferimento della coorte di ricerca
	048	Illimitata	2B48HFG
	007	Illimitata	SD289K9
	E01	31/12/2016	QA5FRD4
	C02	Illimitata	RC3URPQ
	003	31/12/2019	5E1FL7Q
	048	31/12/2020	6F1EL8B

Un insieme di dati relativi ad esenzioni con dati identificativi oscurati. Il numero di riferimento della coorte di ricerca è stato generato dai dati cancellati mediante una funzione hash.

Anche se il nome, l'indirizzo e il CF sono stati cancellati, è possibile recuperare i dati di un soggetto se si conosce la funzione hash usata e i dati che hanno generato il numero di riferimento.

Anonimizzazione in aiuto in caso di Data Breach

Usare tecniche di anonimizzazione consente di rendere il dato e l'informazione fruibile a terzi salvaguardando il diritto alla riservatezza delle persone.

L'anonimizzazione costituisce anche un valido strumento di difesa in caso di data breach, ossia in caso di violazione dei dati da parte di attaccanti esterni.

Non è una questione di **SE** un'organizzazione potrà essere oggetto di violazione, ma piuttosto **QUANDO**.

Anonimizzazione in aiuto in caso di Data Breach

- Prevedere o fermare ogni forma di attacco è quasi impossibile. Oggi, gli utenti intesi come organizzazioni, aziende e singoli individui, hanno bisogno di accedere ad una miriade di sistemi critici, dati, applicazioni, informazioni per svolgere il loro lavoro.
- Queste attività non solo sono svolte dietro un classico firewall aziendale, un classico sistema di protezione perimetrale, ma la crescente esigenza di sistemi **SaaS** che sussistono fuori della rete aziendale e l'uso sempre più diversificato di strumenti per l'accesso ai dati con postazioni mobili rende molto complesso l'intero ambiente.

Anonimizzazione in aiuto in caso di Data Breach

- La tradizionale rete perimetrale sta rapidamente scomparendo, o almeno stiamo assistendo ad un affiancamento di un nuovo modello lavorativo. Basarsi solo su un valido e robusto sistema di protezione perimetrale intorno alla rete aziendale non è più una sufficiente forma di sicurezza.
- Si sta assistendo ad un incoraggiante segnale di cambiamento. Le aziende stanno mostrando maggiore attenzione sulle attività preventive e di monitoraggio, su chi ha accesso alle applicazioni ed ai dati su sistemi tradizionali o sul **CLOUD**, indipendentemente dal dispositivo usato.

Anonimizzazione in aiuto in caso di Data Breach

- Mettere al centro della propria strategia di sicurezza una gestione efficacie dell'identità consente alle organizzazioni di reagire velocemente e meglio ad una violazione; capire meglio chi è e che cosa è a rischio e potenzialmente arrestare un attacco al suo nascere.
- Mentre è necessario fare tutto il possibile per proteggersi contro una violazione, ci sono azioni e passaggi ben precisi che un'organizzazione può prendere per incrementare la propria **resilienza** e potenzialmente ridurre l'impatto negativo di una violazione quando accade.
- In extremis è la gravità della perdita dei dati e non semplicemente il fatto che essi siano stati violati che impatta sulla vita dell'azienda, sul suo modello di business, danneggiandola.
- L'anonimizzazione e l'uso delle varie tecniche presentate prima costituiscono valide armi per fronteggiare una violazione dei dati.

Anonimizzazione: Conclusioni

Le tecniche di anonimizzazione forniscono garanzie di protezione della sfera privata e possono essere utilizzate per creare efficaci procedure di anonimizzazione.

La loro applicazione deve essere progettata in maniera adeguata, ossia i requisiti preliminari (contesto) e l'obiettivo o gli obiettivi della procedura di anonimizzazione devono essere definiti in modo chiaro per poter ottenere l'anonimizzazione auspicata producendo nello stesso tempo dati e informazioni utili.

La soluzione ottimale dovrebbe essere decisa caso per caso, possibilmente utilizzando una **combinazione** di tecniche diverse.

Anonimizzazione: Conclusioni

I responsabili del trattamento dovranno essere consapevoli che un insieme di dati resi anonimi può comunque presentare rischi **residui** per i soggetti interessati.

Anonimizzazione e reidentificazione sono argomenti attivi di ricerca con pubblicazioni regolari su nuove scoperte in materia e, inoltre, persino i dati resi anonimi, ad esempio le statistiche, possono essere oggetti di arricchimento dei profili esistenti dei soggetti, determinando di fatto nuovi problemi di protezione dei dati.

L'anonimizzazione **non va pertanto considerata come un processo statico**, **una tantum**. Rischi connessi devono essere oggetto di un continuo riesame da parte dei responsabili del trattamento.

Privacy, Anonimato e Data Breach nel sistema sanitario

Grazie,

Salvatore Colomba

s.colomba@estar.toscana.it - ESTAR

APiHM