Liguria Digitale

DE IDENTIFICAZIONE

Linux Day 23/10/2021

LIGURIA DIGITALE

Liguria Digitale, società ICT **in house** della Regione Liguria, lavora anche per il mercato, proponendo **soluzioni** e **infrastrutture tecnologiche** e **servizi digitali rivolti** a **imprese ed enti.**

GDPR COMPETENCE CENTER

Il **GDPR** Competence Center di Liguria Digitale è un **team di esperti** che supporta **aziende ed enti**, dalla fase di **progettazione** by design & by default alla **gestione** dei **dati**.

Chi sono e cosa faccio

- Ingegnere, 60 anni, amante della natura
- Lavoro su Unix, Linux dal 1984
- Lavoro sempre in ambito ICT
- Ultimamente (2012) mi occupo di:
  - Sicurezza Informativa
  - Privacy: sono RPD di numerosi Enti (AOU, ASL, Comuni) e sanità convenzionata
  - Cloud
- Per contattarmi m.pastore(at)liguriadigitale.it
- Servizi GDPR Liguria Digitale

Indice

Riferimenti normativi

- **General Data Protection Regulation n.679/2016**
  Regolamento Generale sulla protezione dei dati personali
- **D.Lgs 196/03 (Codice Privacy) novellato dal D.Lgs 101/2018**
- **Norme secondarie del Garante Privacy e del Comitato Europeo per la Protezione dei Dati**
- **Standard e Guidelines** (es. ENISA, ISO, etc.)

Obiettivi

- Il Regolamento UE protegge i diritti fondamentali delle persone fisiche, in particolare il **diritto alla protezione dei dati personali** e disciplina la loro **libera circolazione** nell'Unione Europea e fuori dall'UE.

- Protegge
  - ❖ Diritti
  - ❖ Libertà fondamentali

} **Delle persone FISICHE**

Dato personale (Art. 4)

*«Qualsiasi informazione riguardante una **persona fisica** identificata o **identificabile** («interessato»); […] direttamente o indirettamente»*

- Nome
- Numero identificazione (CF/matricola)
- Dati relativi all'ubicazione (GPS)
- Identificativo online (login, indirizzo IP)
- Elementi caratteristici dell'identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale (impronta, iride, comportamento, etnia, status sociale ed economico)

Definizione

Art. 4 GDPR

«pseudonimizzazione»: trattamento dei dati personali effettuato in modo tale che questi non possano più essere attribuiti a un interessato specifico, a meno che non si utilizzino **informazioni aggiuntive, conservate separatamente e soggette a misure tecniche e organizzative** volte a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile

Esempio

Venditori on line non vogliono permettere il colloquio diretto tra Fornitore e Compratore:

- L’indirizzo di posta di mittente e destinatario sono mappati su caselle del Venditore
- L’indirizzo dell’appartamento è mascherato e viene indicata solo un’area

Attenzione

- Senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative
- Non basta inserire un codice tra i due insiemi di dati in assenza di misure organizzative:
  - Dati direttamente identificabili
  - Dati personali (anche particolari)

Esempio navigazione internet

- In azienda
  - Il sito visitato vede l'indirizzo IP pubblico del Proxy
  - Il proxy vede l'indirizzo IP privato del PC
- Per il Titolare del sito il dato è pseudonimizzato
- Per l'Azienda no: L'amministratore del Proxy è in grado di correlare i dati
- Da casa:
  - Il vostro ISP conosce il vostro IP
  - Il sito visitato l'IP pubblico dell'ISP del PAT

Definizioni: dato anonimo

GDPR C 26

- I principi di protezione dei dati non dovrebbero pertanto applicarsi a **informazioni anonime**, vale a dire **informazioni che non si riferiscono a una persona fisica identificata o identificabile o a dati personali resi sufficientemente anonimi da impedire o da non consentire più l'identificazione dell'interessato**. Il presente regolamento non si applica pertanto al trattamento di tali informazioni anonime, anche per finalità statistiche o di ricerca.

Applicabilità del GDPR al dato pseudonimizzato

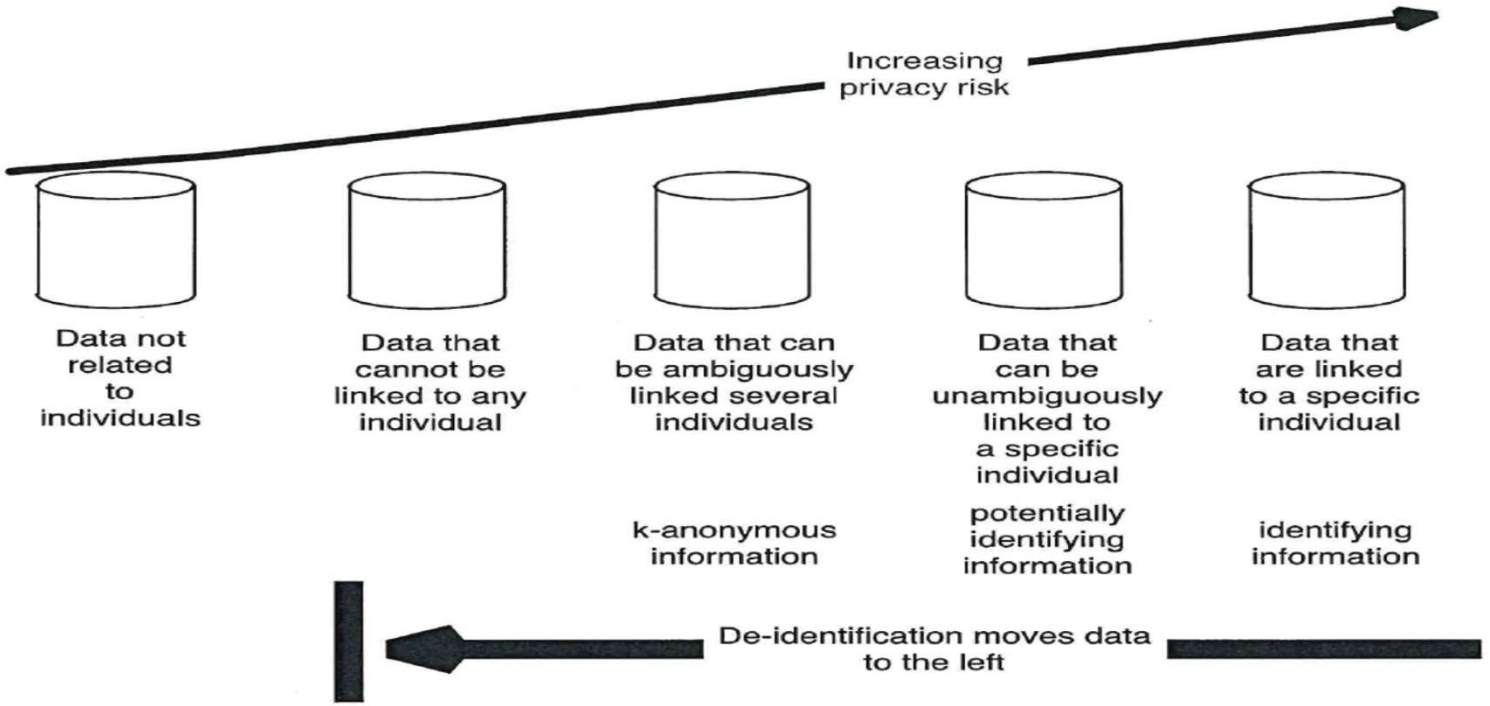
Considerando 26 GDPR

È auspicabile applicare i principi di protezione dei dati a tutte le informazioni relative a una persona fisica identificata o identificabile. **I dati personali sottoposti a pseudonimizzazione, i quali potrebbero essere attribuiti a una persona fisica mediante l'utilizzo di ulteriori informazioni, dovrebbero essere considerati informazioni su una persona fisica identificabile**. *Per stabilire l'identificabilità di una persona è opportuno considerare tutti i mezzi, come l'individuazione, di cui il titolare del trattamento o un terzo può ragionevolmente avvalersi per identificare detta persona fisica direttamente o indirettamente*. Per accertare la ragionevole probabilità di utilizzo dei mezzi per identificare la persona fisica, si dovrebbe prendere in considerazione l'insieme dei fattori obiettivi, tra cui i costi e il tempo necessario per l'identificazione, tenendo conto sia delle tecnologie disponibili al momento del trattamento, sia degli sviluppi tecnologici

Bestialità

- Manutentore di apparecchiature medicali contenenti dati sanitari pseudonimizzati (ID paziente è un codice) sostiene di non trattare dati personali
- L'Amministratore di un proxy senza autenticazione sostiene di non trattare dati personali e quindi non si pone il problema del corretto bilanciamento tra minimizzazione del trattamento (art. 5 GDPR) e della sicurezza (art. 32 GDPR)

Una scala «graduale»



Obiettivi

Soddisfacimento dei requisiti di:

1. Minimizzazione del trattamento (art. 5 GDPR)
2. Privacy per progettazione (art. 25 GDPR)
3. Sicurezza del trattamento (art. 32 GDPR)
4. Valutazione di impatto Privacy (DPIA - art. 35 GDPR)
5. Garanzie e deroghe relative al trattamento a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica a fini statistici (art. 89 GDPR)

Una vasta classe di problemi

- Occorre dare accesso ai dati:
  - Per altre finalità diverse da quelli che sono stati raccolti (esempio dati sanitari raccolti per finalità di cura utilizzati per ricerca scientifica)
  - A soggetti diversi: Esempio Ricercatori, case farmaceutiche
  - A un pubblico vasto: Esempio Pubblicazione scientifica
- Diminuire i rischi per l'interessato:
  - Identificazione (danno reputazionale, economico)
  - Integrità del dato (esempio scambio dati tra due soggetti con conseguenze anche fatali, esempio scambio sacche di sangue)
- Sostituire/integrare altre misure di protezione dei dati:
  - Autenticazione ed autorizzazione
- Fornire servizi data driven senza consegnare i dati: Esempi Facebook, Google

Anonimizzazione: i fraintendimenti più comuni EDPS – AEPD 2021

1) ~~Pseudonimizzazione = anonimizzazione~~

I dati pseudonimizzati possono essere attribuiti alla persona cui si riferiscono solo con l'utilizzo di informazioni aggiuntive. I dati anonimizzati non possono più essere associati a specifici individui.

2) ~~Anonimizzazione = crittografia~~

La crittografia può rappresentare uno strumento di pseudonimizzazione nella misura in cui rende i dati accessibili solo con l'uso della chiave per la decrittografia: la conoscenza di tale chiave rappresenta l'informazione aggiuntiva per la pseudonimizzazione. La cancellazione della chiave non rende i dati crittografati anonimi: la confidenzialità dei dati crittografati è influenzata da molti fattori.

3) ~~È sempre possibile anonimizzare i dati~~

Il contesto e la natura dei dati incidono sul rischio di reidentificazione che, per esempio ed in via generale, tanto più è elevato quanto è più basso il numero degli interessati i cui dati si intende anonimizzare.

Anonimizzazione: i fraintendimenti più comuni EDPS – AEPD 2021

4) ~~L'anonimizzazione è per sempre~~

Una volta anonimizzati, il rischio di reidentificazione muta nel tempo per essenzialmente due ragioni:

- Avanzamento tecnologico e nuovi strumenti che consentono la reidentificazione
- Rilevazione e conoscenza nel tempo di informazioni ulteriori tali da consentire la reidentificazione

5) ~~L'anonimizzazione elimina la probabilità di reidentificare le persone cui i dati si riferiscono~~

In base alle circostanze concrete, il rischio di identificazione può, talvolta, non essere azzerato e dunque deve essere preso in considerazione unitamente alle conseguenze che la reidentificazione causerebbe in capo agli interessati.

6) ~~L'anonimizzazione è un concetto binario che non può essere misurato~~

Salvo ipotesi specifiche in cui i dati sono altamente generalizzati, il rischio di reidentificazione non è mai zero; dunque è importante effettuare una valutazione e monitoraggio di tale rischio nel tempo.

Anonimizzazione i fraintendimenti più comuni EDPS – AEPD 2021

7) ~~L'anonimizzazione può essere integralmente automatizzata~~

Gli strumenti automatizzati sono utili nel processo di anonimizzazione ma l'intervento umano è necessario: solamente un'accurata analisi del contesto consente di tutelare adeguatamente i diritti delle persone i cui dati si vuole anonimizzare.

8) ~~L'anonimizzazione rende i dati inutili~~

Non è così. L'utilità dei dati anonimizzati si può valutare in relazione alla finalità che si intende perseguire con l'uso di tali dati. Inoltre, in alcuni casi, per non incorrere in violazione delle norme di legge, l'anonimizzazione deve essere necessariamente effettuata qualora si intenda continuare il trattamento.

Anonimizzazione: i fraintendimenti più comuni EDPS – AEPD 2021

9) ~~Seguire processi di anonimizzazione usati da altri conduce agli stessi risultati~~

Nel processo di anonimizzazione devono considerarsi: natura, portata, contesto e finalità del trattamento. Devono, inoltre, essere valutati i rischi diversi per probabilità e gravità rispetto ai diritti e le libertà delle persone fisiche coinvolte.

10) ~~Scoprire a chi si riferiscono i dati non è di interesse e non comporta rischi~~

La reidentificazione può avere un impatto più o meno grave sugli interessati in relazione alle differenti caratteristiche dei dati (es. natura). Non può prescindersi da una corretta valutazione del rischio di reidentificazione.

Tipi ti attacco

Quali sono le tipologie di attacco?

- Accusatore;
- Giornalista;
- Marketing;
- Identificabilità differenziale;
- Appartenenza;
- Inferenza.

Accusatore

Il datore di lavoro, che vuole accertarsi dello stato di salute di un candidato, consulta i dati di uno studio clinico sapendo che il candidato ne è parte non sapendo se è nel gruppo di controllo oppure se è affetto dalla patologia dello studio.

Giornalista

Un giornalista vuole discreditare un'organizzazione dimostrando che i dati che ha pubblicato possono portare all'identificazione di un interessato (non importa quale).

Marketing

Un’azienda vuole valutare il valore di database di marketing offerto da uno o più potenziali fornitori.

Identificabilità differenziale

Un osservatore può capire se un soggetto ha una data patologia leggendo i risultati di un trail clinico prima su un campione nel quale egli non è incluso e, in seguito, sullo stesso campione aggiungendo i dati del medesimo soggetto. Se, per esempio, nel secondo test ci fosse un test positivo in più rispetto al precedente, l'osservatore potrebbe dedurre che il soggetto è affetto dalla patologia.

Appartenenza

Un’azienda che produce prodotti biologici può inviare delle offerte pubblicitarie a dei bersagli precisi sapendo che essi sono tra i firmatari di una petizione online contro gli agenti inquinanti usati sui prodotti alimentari.

Inferenza

Un osservatore può scoprire che un paziente è affetto da una determinata malattia avendo a disposizione le informazioni relative alla sua residenza e sapendo che tutti gli abitanti di una certa località sono affetti da una data malattia.

Variabili

Variabili che influenzano il processo:

1. Capacità dell'attaccante in termini di conoscenze tecniche e di risorse;
2. Facilità dell'accesso a fonti informative diverse;
3. Scorrere del tempo (i progressi tecnologici possono mettere a disposizione nuovi strumenti, nuove informazioni pubblicate).

Glossario

Termine	Definizione
Attributo chiave	Sinonimo di Identificativo indiretto
Dati aggregati	Dati che rappresentano un gruppo di interessati, quali un insieme di funzioni statistiche calcolati sul gruppo
Identificativo	Un insieme di attributi in un insieme di dati che consente l'identificazione univoca di un interessato all'interno di un contesto operativo
Identificativo diretto	Un attributo in un insieme di dati che da solo consente l'identificazione univoca di un interessato in un dato contesto
Identificativo indiretto	Un attributo in un insieme di dati che, quando considerato congiuntamente con altri attributi che può essere o meno all'interno di un insieme di dati, consente identificazione univoca di un interessato all'interno di un contesto operativo
Identificativo locale	Un insieme di attributi in un insieme di dati che insieme distinguono un interessato nell'insieme di dati
Identificatore unico	Un attributo in un insieme di dati che da solo distingue un interessato all'interno di un contesto, per esempio uno pseudonimo
Macrodati	Insieme di dati aggregati
Microdata	Insieme di dati riguardanti singoli interessati
Quasi identificativo	Un attributo in un insieme di dati che, considerato congiuntamente con altri attributi nell'insieme di dati, distingue un interessato
Record longitudinale di salute personale	Sequenza ordinata in ordine cronologico di informazioni significative su un paziente tenute per un lungo tempo
Studio osservazionale	Studio clinico che non prevede di intervenire sui pazienti interessati se non tramite quantità lievemente maggiori di prelievi comunque previsti dalla normale prassi clinica
Studio osservazione retrospettivo	Studio osservazionale che si basa su dati clinici acquisiti nel passato
Studio osservazionale prospettico	Studio osservazionale che si basa anche su dati clinici acquisibili nel futuro

Tassonomia del rischio

I destinatari dei dati de identificati possono essere logicamente suddivisi in due categorie:

- Fidati: soggetti con cui si ha o si costruisce una relazione di fiducia basata su un accordo.
- Non Fidati: soggetti con cui non si ha alcuna relazione di fiducia (pertanto in questi casi il rischio deve essere basso).

Rischio

Il rischio può essere caratterizzato in base a:

- Motivazioni dell'attaccante;
- Abilità dell'attaccante;
- Mezzi a disposizione dell'attaccante;
- Disponibilità di altri dati;
- Insieme di dati originari;
- Tecniche utilizzate nella de identificazione;
- Risultato a cui l'attaccante arriva.

Tecniche di de identificazione

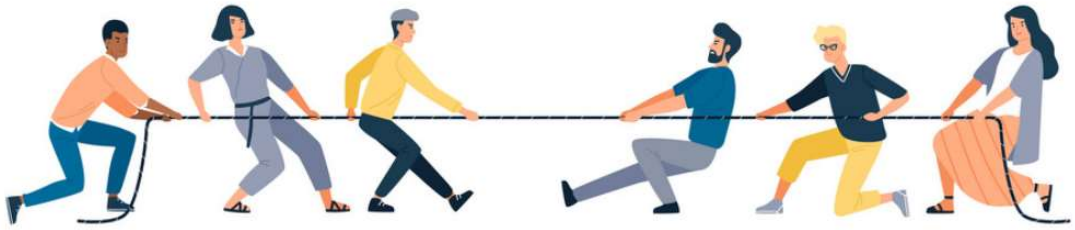
Le tecniche di de identificazione variano a seconda del:

- Tipo di utilizzo;
- Contesto;
- Numero di interessati;
- Tipologia di dato.

Qualifica del personale

È importante distinguere due gruppi di persone:

- Uno che determini le modalità di de identificazione;
- Uno che individui i rischi di identificazione più plausibili e la validità dei dati ai fini dello scopo di utilizzo.



Tecniche di de identificazione

Quali sono le tecniche di de identificazione?

1) Tecniche statistiche;
2) Tecniche crittografiche;
3) Tecniche di soppressione;
4) Tecniche di pseudonimizzazione;
5) Tecniche di dissezione;
6) Tecniche di generalizzazione;
7) Tecniche di randomizzazione;
8) Dati sintetici.

Tecniche di de identificazione

1) **Tecniche statistiche**:

Campionamento: tecnica che consiste nell'individuazione di un sottoinsieme di dati che si ritiene significativo per l'uso previsto, eliminando i dati di alcuni interessati.

Caratteristiche
- produce microdati;
- protegge da attacchi di tipo Appartenenza e Interferenza;
- tecnica adatta agli studi osservazionali retrospettivi.

https://en.wikipedia.org/wiki/Statistics

Tecniche di de identificazione

Dati presenti nelle cartelle cliniche:

Sdo	Codice fiscale	Patologia	Misura
27	PSTMRZ60H26E290S	Ipertensione	150
28	XXXXXXXXXXXXXXXX	Cardiopatia	170
29	YYYYYYYYYYYYYYYY	Aritmia	180
30	ZZZZZZZZZZZZZZZZZZ	Ipertensione	200
-…			
3000	KKKKKKKKKKKKKKKK	Ipertensione	190

Tramite la funzione excel CASUALE.TRA(27;3000) si ha un valore casuale tra il minimo e il massimo del numero di sdo. Ipotizzando che per la significatività statistica dello studio occorrano 150 record, ricalcolando un numero di volte superiore a 150 (potrebbero esserci valori ritornati dalla funzione duplicati) la funzione CAUSALE.TRA si ottengono i numeri di sdo da inserire nello studio. Quindi il dataset dopo il campionamento casuale sarà:

Sdo	Codice fiscale	Patologia	Misura
27O	LLLLLLLLLLLLLLLLLL	Ipertensione	150
283	JJJJJJJJJJJJJJJJJJJJ	Cardiopatia	170
29	YYYYYYYYYYYYYYYY	Aritmia	180
303	MMMMMMMMMM	Ipertensione	200
-…			

Tecniche di de identificazione

Aggregazione: tecnica che consiste nel produrre dei record aggregati che rappresentano i dati originari.

Caratteristiche

- produce macrodata;
- protegge dall'individuazione di un singolo record

Tecniche di de identificazione

I dati visti in precedenza possono essere modificati sostituendo la misura con la relativa media sotto riportata. I valori per gli ipertesi sono stati sostituiti dalla rispettiva media. Gli altri (cardiopatici, aritmia) rimangono invariati nell'ipotesi che non vi siano altri record con la stessa patologia.

Sdo	Codice fiscale	Patologia	Misura
27	PSTMRZ60H26E290S	Ipertensione	180
28	XXXXXXXXXXXXXXXX	Cardiopatia	170
29	YYYYYYYYYYYYYYYY	Aritmia	180
30	ZZZZZZZZZZZZZZZZZZ	Ipertensione	180
-…			
3000	KKKKKKKKKKKKKKKK	Ipertensione	180

Tecniche di de identificazione

2) Tecniche crittografiche:

Caratteristiche

- impongono di proteggere le chiavi (se usate);
- tipicamente ha un onere computazionale maggiore;
- normalmente viene utilizzata quando si vuole rendere possibile una re identificazione dei dati in un ambiente autorizzato/controllato.

https://en.wikipedia.org/wiki/Cryptography

Tecniche di de identificazione

Cifratura deterministica: tecnica che consiste nell'utilizzo di qualunque tecnica di cifratura (asimmetrica, a blocchi, a flussi) per trasformare gli attributi identificativi o i valori sensibili.

Caratteristiche

- consente di non distorcere i dati;
- non mantiene l'ordinamento (consente solo test di uguaglianza);
- chi possiede la chiave può risalire a tutti i dati originari.

Tecniche di de identificazione

I dati di cui sopra possono essere trattati sostituendo i valori dei codici fiscali e del numero di sdo con i risultati dell'applicazione dell'algoritmo AES128 ai rispettivi valori originari:

Sdo	Codice fiscale	Patologia	Misura
H123!njkldfgug#RTYJgfcv67	Jhbxqw189912ynxoixq!?1t	Ipertensione	150
Cwuincnwe2m3iojpy7892h	Cwuincnwe2m3iojpy7892h	Cardiopatia	170
jkahbxWCW23hl4bnHAO6Y	……	Aritmia	180
….	……..	Ipertensione	200

Si noti che:

1. I campi cifrati hanno cambiato lunghezza: l'uso di un algoritmo a blocchi (AES128) implica che ciascuna rappresentazione cifrata sia di lunghezza di multipli interi della lunghezza della chiave (in questo caso 128bit/8bit = 16 byte. Il risultato è un valore binario che viene normalmente rappresentato in base64 per essere stampabile. Questo determina una maggior lunghezza del 33%, nel caso specifico 22 caratteri. L'uso di chiavi maggiori, come nel caso della crittografia asimmetrica (1024 o 2048) accentua il fenomeno;
2. I campi cifrati non hanno più il formato originario;
3. L'ordinamento dei dati cifrati non corrisponde a quello dei dati originari.

Tecniche di de identificazione

Cifratura che preserva l'ordinamento: tecnica variante della cifratura deterministica con il vantaggio che i confronti sui dati de identificativi equivalgono ai confronti su dati originari.

Caratteristiche

- è tipicamente applicata a valori numerici;
- svantaggio dell'onerosità computazionale dell'algoritmo, della relativa debolezza e dell'assenza di standardizzazione.

Tecniche di de identificazione

Cifratura che conserva il formato: algoritmi che preservano il formato (lunghezza della stringa e spazio dei caratteri).

Caratteristiche
- è adatto al test;
- ci sono algoritmi standard e sicuri.

https://en.wikipedia.org/wiki/Format-preserving_encryption

Tecniche di de identificazione

Crittografia omomorfa: forma di crittografia random.

Caratteristiche

- consente una o più operazioni sui dati cifrati;
- non distorce i dati;
- produce microdata;
- è possibile risalire ai dati originari solo conoscendo la chiave;
- è adatto al data mining se le operazioni previste non distorcono i dati (tranne quelli soppressi).
- Onere computazionale molto alto e dipendente dalla precisione dei dati numerici

Tecniche di de identificazione

3) Tecniche di soppressione:

Si tratta di tecniche che comportano la perdita di informazione tramite la soppressione di un attributo (colonna) per tutti i record/ per alcuni record/ per alcuni campi di alcuni record.

Caratteristiche

- produce microdata;
- i valori aggregati su dati de identificati potrebbero essere distorti se calcolati su dati de identificati.

Tecniche di de identificazione

Mascheratura: tecnica che consiste nell'eliminazione dei campi direttamente identificativi di (alcuni o tutti) i campi identificativi.

Caratteristiche

- è la più utilizzata;
- fornisce i risultati peggiori in termini di sicurezza: per tale ragione i dati de identificati con questo algoritmo devono essere protetti con misure poco più rilassate di quelle dei dati originari.
- Alti rischi di inferenza

Tecniche di de identificazione

I dati visti in precedenza vengono mascherati sopprimendo la chiave univoca Codice Fiscale e un attributo indirettamente identificativo (SDO).

Patologia	Misura
Ipertensione	150
Cardiopatia	170
Aritmia	180
Ipertensione	200
Ipertensione	190

Si noti che:

- Non è possibile correlare i dati con altri insieme di dati con le certezze necessarie per condurre uno studio prospettico, non essendoci più i dati identificativi. Quindi eventuali controlli di qualità sui dati devono essere compiuti prima della mascheratura;
- La correlazione tra Patologia e Misura può essere fatta tranquillamente;
- Gli attacchi di tipo inferenziale possono avere ancora luogo: sapendo che l'interessato è nel dataset e che la sua misura è 180 posso dedurre che un Aritmia. Oppure sapendo che è cardiopatico che ha 170 come misura. Questo nell'ipotesi che non ci siano altri record con questi valori.

Tecniche di de identificazione

Soppressione locale: tecnica che consiste nella soppressione di valori di singoli record per specifici attributi.

Caratteristiche
- è tipicamente utilizzato per eliminare valori rari;
- deve essere eseguito su un numero consistente di record;
- è adoperato ad attributi di categorizzazione.

Tecniche di de identificazione

Soppressione di record: tecnica che consiste nell'eliminazione di record con combinazioni rare di valori.

Tecniche di de identificazione

Si tolgono i record contenenti i valori rari Aritmia e Cardiopatia.

Patologia	Misura
Ipertensione	150
Ipertensione	200
…	…
Ipertensione	190

Si noti che:

- La media della misura potrebbe essere diversa da quella effettuata sui dati originari;
- Lo studio non può dare indicazioni sulle patologie eliminate (non avrebbe comunque potuto darli per il numero limitato di valori).

Tecniche di de identificazione

4) Tecniche di pseudonimizzazione:

Si tratta di tecniche consistenti nella sostituzione degli identificativi diretti con identificativi indiretti appositamente creati.

Caratteristiche

- consentono il collegamento tra diverse fonti senza rilevare direttamente l'identità degli interessati;
- producono microdata;
- producono tavola di assegnazione degli pseudonimi, chiavi crittografiche

tali informazioni possono essere utilizzate per re identificare gli interessati (le informazioni devono essere adeguatamente protette).

Tecniche di de identificazione

Come avviene la pseudonimizzazione?

Due fasi:

- Selezione degli attributi (tutti gli identificativi diretti e potenzialmente tutti o quasi gli identificativi indiretti)
- Creazione di pseudonimi

La creazione di pseudonimi è influenzata da:
- oneri computazionali;
- necessità della re identificazione;
- possibilità che due diverse chiavi abbiano lo stesso pseudonimo.

Tecniche di de identificazione

Nell’esempio precedente:

- Il Codice Fiscale è identificativo diretto;
- Il Codice Fiscale è un identificativo locale;
- Il Codice Fiscale è un identificatore unico;
- La sdo è un identificativo locale;
- La sdo è identificativo indiretto e quindi attributo chiave;
- La sdo è un identificatore unico.

Tecniche di de identificazione

Pseudonimi indipendenti dagli attributi identificativi: in tal caso occorre mantenere una tabella di corrispondenza tra attributi identificativi e lo pseudonimo.

Pseudonimi dipendenti dalle chiavi tramite crittografia: in questa ipotesi occorre individuare una funzione crittografica che in base all'attributo (o attributi) identificativo determini lo pseudonimo.

- cifratura;
- uso di funzioni di hash.

https://en.wikipedia.org/wiki/Hash

Tecniche di de identificazione

Applicazione di un numero progressivo.

La tabella di esempio ha già una colonna (sdo) che rappresenta il progressivo di ricovero (nell'anno). Quindi mantenere questa colonna, seppur tramite dati del contesto, consente l'identificazione dell'interessato. Infatti accedendo a una delle seguenti fonti di informazioni: cartelle cliniche (in formato digitale o cartaceo), flussi verso Regione e Ministero, sistema informativo ospedaliero modulo ADT, eventuali sistemi verticali (RIS, LIS) che per i pazienti interni registrino sia il Codice Fiscale che il numero di SDO consente l'identificazione dell'interessato. Inoltre il primo ricoverato dell'anno può essere più facilmente individuato: molti quotidiani pubblicano l'elenco dei primi nati dell'anno, almeno con il nome della madre. Quindi questa soluzione migliora leggermente la sicurezza del dato rispetto a mantenere il Codice Fiscale dell'interessato, ma non può essere considerata sicura: lo è quanto il più debole dei sistemi sopracitati presso la struttura ospedaliera o altri soggetti (Regione, Ministero).

Tecniche di de identificazione

Uso di un progressivo locale.

La mappatura tra i dati identificativi (SDO, Codice Fiscale) e il nuovo progressivo va mantenuta con adeguate misure di protezione, oppure la copia originale dei dati va mantenuta nello stesso ordine in cui sono stati attribuiti i progressivi. Questa seconda opzione espone all'attacco del giornalista e anche a quella di marketing: un attaccante potrebbe provare diversi ordinamenti (per SDO e quindi temporale, per Codice Fiscale, ect.) e conoscendo i dati di almeno un soggetto potrebbe identificare la totalità dei soggetti tramite l'accesso a dati di contesto (esempio elenco pazienti reclutati).

Tecniche di de identificazione

Hash SHA256 del Codice Fiscale (con mascheramento della colonna SDO).

Hash 256	Patologia	Misura
53dfa930b74cb3c6793f012fbf52059ff296b1d085f10cad70d475e045be7cf9	Ipertensione	150
72c84ba99d77ee766e9468a0de36433a44888e5dec4afb84f8019777800b7364	Cardiopatia	170
c023200b20ee92af51d10a4a47ef74d1f6bc76937b7d9080dde4123aa1e1e9c3	Aritmia	180
cfefb6d182428582bac03cb663bf65c69de4fb16b717168e1055ecc978d0c2af	Ipertensione	200
-…		
a1f7746b449edffa84e5950aebaeecb52d68d728d56e2fef7805afafdee2dec6	Ipertensione	190

Si noti che:

- La lunghezza degli hash è costante (dipende solo dall'algoritmo scelto, ma non dal valore in input alla funzione) e non preserva il formato originale;
- Non è necessario conservare una tabella di corrispondenza tra hash e valori degli identificativi originali in quanto la funzione di hash produce lo stesso risultato a parità di input;
- Chi disponesse dell'elenco dei Codici Fiscali dei pazienti inclusi nello studio potrebbe eseguire tutti gli attacchi con successo. Unica difficoltà è l'individuazione del campo sui cui si è costruito hash. L'algoritmo è desumibile dalla lunghezza (tranne alcuni casi in cui algoritmi diversi producono lunghezze identiche). Una potenziale difesa è concatenare il campo con una stringa nota solo ai ricercatori (per esempio !8P) e calcolare hash della stringa concatenata: esempio !8PPSTMRZ60H26E290S che fornisce il valore e20a5b4c9970a9cc1a313848684873aa04fe04870ab87609e93d9bbf02e8d4b1 diverso da quello calcolato con il solo PSTMRZ60H26E290S. La difesa è robusta quanto la segretezza del valore !8P. E' un esempio di hash con chiave (banale) ma vi sono algoritmi ben più robusti.

  https://en.wikipedia.org/wiki/Message_authentication_code

Tecniche di de identificazione

5) Tecniche di dissezione:

Si tratta di tecniche consistenti nel:

- raggruppare i record in base alle combinazioni dei dati quasi identificativi, aggiungendo qualche attributo non quasi identificativo;
- creare un identificativo per ciascuna classe di equivalenza;
- mantenere una tabella (ove si voglia rendere possibile la re identificazione) con la relazione tra le classi di equivalenza e i dati identificativi in modo protetto.

Caratteristiche
- non distorce i dati prodotti.

Tecniche di de identificazione

Esempio:

Sdo	Codice fiscale	Patologia	Misura	Città
27	PSTMRZ60H26E290S	Ipertensione	180	Roma
28	XXXXXXXXXXXXXXXX	Cardiopatia	170	Milano
29	YYYYYYYYYYYYYYYY	Aritmia	180	Milano
30	ZZZZZZZZZZZZZZZZZZ	Ipertensione	180	Milano
-…				
3000	KKKKKKKKKKKKKKK	Ipertensione	180	Roma

Viene prodotta una tabella con le classi di equivalenza (A) ed una con il resto dei dati, collegata alla prima dall'identificatio di classe (B): :

Tabella A

Id classe	Patologia	Misura	Città
1	Aritmia	180	Milano
2	Cardiopatia	170	Milano
3	Ipertensione	180	Milano
4	Ipertensione	180	Roma
4	Ipertensione	180	Roma

Tabella B

Sdo	Codice fiscale	Id Classe
27	PSTMRZ60H26E290S	4
28	XXXXXXXXXXXXXXXX	2
29	YYYYYYYYYYYYYYYY	3
30	ZZZZZZZZZZZZZZZZZZ	4
-…		
3000	KKKKKKKKKKKKKKK	4

Tecniche di de identificazione

6) Tecniche di generalizzazione:

Si tratta di tecniche che, allo scopo di aumentare il numero di record che condividono una determinata combinazione di valori delle colonne, comporta una rinuncia alla precisione (non necessaria) di alcune colonne di dati.

Algoritmi specifici:

1. Arrotondamento;
2. Valore minimo e massimo;
3. Combinazione di un insieme di attributi in un unico attributo;
4. Sostituzione di valori rari (generalizzazione locale).

Tecniche di de identificazione

7) Tecniche di randomizzazione:

Si tratta di tecniche che consistono nella modifica di uno o più attributi modificando i valori con un valore random.

Caratteristiche

- producono microdata perturbati;
- a determinate condizioni le funzioni statistiche per la produzione di dati aggregati danno risultati coerenti con quelli operati su dati originali.

Aggiunta di rumore: consiste nel sommare o moltiplicare per una variabile random i valori di alcune colonne, tipicamente numeriche. Le distribuzioni del rumore devono essere scelte con accuratezza per non inficiare i risultati statistici ottenuti sui dati perturbati.

Tecniche di de identificazione

Permutazione: tecnica che consiste nello scambiare i dati tra un record e l'altro, scelti con una funzione random.

Caratteristiche
- mantenimento funzioni aggregate sui valori scambiati;
- vengono inficiate le eventuali correlazioni con altre variabili non permutate o permutate con un algoritmo diverso

Microaggregazione:
- Individuazione classi di equivalenza sufficientemente ampie per rendere la procedura robusta agli attacchi;
- Calcolo della media dei valori in ciascuno dei gruppi individuati;
- Sostituzione dei valori dei rispettivi record con la media.

Tecniche di de identificazione

Permutazione

Un questionario senza dati direttamente identificativo di pazienti è somministrato tramite la compilazione di diverse pagine. Se si ha cura di porre sulla medesima pagina le variabili da correlare e di mettere eventuali attributi quasi identificativi su pagine diverse, si possono mescolare le pagine come si farebbe per un mazzo di carte.

In questo modo si diminuisce la probabilità di re identificazione e nel contempo si riesce ad effettuare studi sul collegamento di fenomeni tra di loro.

Tecniche di de identificazione

Micro aggregazione:

Premesso che l'esempio non è sufficientemente ampio si procede a fare la media tra i valori di Misura nell'ambito della stessa patologia e a sostituire i valori con la media stessa. Ovviamente questa operazione non consente di effettuare correlazioni tra altre variabili e la Misura. Si noti che qua la media viene effettuata su sottoinsiemi di dati. Su esempi più numerosi si potrebbe fare media in k intervalli di Misura (tenendo conto o meno della Patologia): per esempio a tuti i valori di Misura compresi tra 100 e 120 viene sostituita la relativa media, quelli tra 120 e 140 viene sostituita la relativa media e si fa la media tra le Misure > 140 e si sostituisce nel record il relativo valore.

Patologia	Misura
Aritmia	180
Cardiopatia	170
Ipertensione	150
Ipertensione	200
Ipertensione	190

Patologia	Misura
Ipertensione	180
Cardiopatia	170
Aritmia	180
Ipertensione	180
Ipertensione	180

Tecniche di de identificazione

8) Dati sintetici:

Si tratta di una tecnica consistente nel generare un nuovo insieme di dati che abbia le stesse proprietà statistiche, comprese le correlazioni tra le diverse variabili.

Caratteristiche

- per produrre i dati (che per definizione non sono attribuibili ai singoli interessati) si ricorre a randomizzazione e campionamento;
- a determinate condizioni le funzioni statistiche per la produzione di dati aggregati danno risultati coerenti con quelli operati su dati originali;
- è indicato per produrre dati a scopi didattici/test.

Sintesi delle tecniche

Tecnica	Produce	Perturbati	Formato	Ordinamento	Previene	Onere computazionale	Reversibile	Non adatto
Campionamento	Microdata	No	Mantenuto	Mantenuto	Inferenza Appartenenza	Medio basso	No	Numero record appena sufficiente
Aggregazione	Macrodata	Si	Mantenuto	Mantenuto solo sui dati aggregati	Accusatore Giornalista Marketing	Medio	No	Presenza valori rari non aggregati
Cifratura simmetrica	Microdata	No	Non mantenuto	Non mantenuto	Accusatore Giornalista Marketing	Medio Alto	Si con chiave	Pochi valori diversi
Cifratura asimmetrica	Microdata	No	Non mantenuto	Non mantenuto	Accusatore Giornalista Marketing	Alta	Si con chiave	Pochi valori diversi
Hash	Microdata	No	Non mantenuto	Non mantenuto	Accusatore Giornalista Marketing	Medio bassa	Si	Valori predeterminabili
Cifratura che preserva ordinamento	Microdata	Si	Mantenuto	Mantenuto	Accusatore Giornalista Marketing	Medio alta	Si con chiave	Solo valori numerici
Cifratura che conserva il formato	Microdata	Si	Mantenuto	Non mantenuto	Accusatore Giornalista Marketing	Medio alta	Si con chiave	Adatto per test o per uso applicazioni non modificabili
Cifratura omomorfa	Microdata	No	Mantenuto	Mantenuto	Accusatore Giornalista Marketing	Alta	Si con chiave	Solo valori numerici

Sintesi delle tecniche

Tecnica	Produce	Perturbati	Formato	Ordinamento	Previene	Onere computazionale	Reversibile	Non adatto
Mascheratura	Microdata	No, tranne informazioni soppresse	Mantenuto	Mantenuto, tranne informazioni soppresse		Bassa	No. Possibile solo mantenendo altrove	
Soppressione locale	Microdata	Solo per informazioni soppresse e su aggregazioni dei rispettivi attributi	Mantenuto per altri attributi e per record non modificati	Non mantenuto per i valori soppressi	Inferenza Giornalista	Basso	No	
Soppressione di record	Microdata	Solo per informazioni soppresse e su aggregazioni dei rispettivi attributi	Mantenuto	Non mantenuto per i valori soppressi	Inferenza Giornalista	Basso	No	
Pseudonimi indipendenti dai dagli attributi identificativi	Microdata	No	Mantenuto	Per i campi chiave dipende dall'algoritmo scelto. Preferibile che non sia mantenuto	Accusatore Giornalista Marketing	Medio basso	Mantenendo in modo sicuro tabella di corrispondenza	
Pseudonimi tramite hash	Microdata	No	Non mantenuto per attributi identificativi	Non mantenuto per attributi identificativi	Accusatore Giornalista Marketing	Medio basso	Si	Valori predeterminabili

Sintesi delle tecniche

Tecnica	Produce	Perturbati	Formato	Ordinamento	Previene	Onere computazionale	Reversibile	Non adatto
Pseudonimi tramite cifratura deterministica	Microdata	No	Non mantenuto per attributi identificativi	Non mantenuto per attributi identificativi	Accusatore Giornalista Marketing	Medio per crittografia simmetrica Alto per crittografia asimmetrica	Si con chiave	Pochi valori diversi Valori predeterminabili
Hash con salt	Microdata	No	Non mantenuto per attributi identificativi	Non mantenuto per attributi identificativi	Accusatore Giornalista Marketing	Medio	Si conservando salt	Pochi valori diversi Valori predeterminabili
Hash con salt e chiave	Microdata	No	Non mantenuto per attributi identificativi	Non mantenuto per attributi identificativi	Accusatore Giornalista Marketing	Medio	Si conservando salt e chiave	
Dissezione	Microdata	No	Mantenuto	Mantenuto	Inferenza	Medio basso	Si con tabella	
Arrotondamento	Microdata	Si, contenuta	Mantenuto	Si, tranne all'interno dell'arrotondamento	Inferenza Giornalista	Basso	No	Valori non numerici
Valore minimo e massimo	Microdata	Si	Mantenuto	Si tranne per i valori minori del minimo o maggiori del massimo	Inferenza Giornalista	Basso	No	Valori non numerici

Sintesi delle tecniche

Tecnica	Produce	Perturbati	Formato	Ordinamento	Previene	Onere computazionale	Reversibile	Non adatto
Combinazione di attributi	Microdata	Si	Non mantenuto per gli attributi combinati	Dipende dalla funzione usata, sarebbe meglio di no	Accusatore Giornalista Marketing	Basso	Dipende dalla funzione usata	
Generalizzazione locale	Microdata	Si, i record modificati	Mantenuto per altri attributi e per record non modificati	Non mantenuto per i valori modificati	Inferenza Giornalista	Basso	No	
Aggiunta di rumore	Microdata	No	SI	No	Inferenza Giornalista	Medio basso	No	
Permutazione	Microdata	Si (in termini di consistenza dei diversi record)	Si	Si (tranne che per i valori permutati)	Inferenza Giornalista	Basso	No	
Microaggregazione	Microdata	Si	Mantenuto	NO (almeno per i valori aggregati)	Inferenza	Medio bassa	No	
Dati sintetici	Microdata	Si	Mantenuto	No	Tutti	Alto	No	Adatto solo per didattica o test

Ambiti di utilizzo

Le operazioni effettuate sull'insieme di dati ne possono compromettere l'utilità per diversi scopi.

Inoltre le misure idonee ad ottenere il risultato possono inficiare alcune proprietà del dato:

1. **Disponibilità** → l'uso di algoritmi crittografici può portare alla perdita delle informazioni crittografiche in caso di indisponibilità della chiave;
2. **Integrità** → le diverse tecniche utilizzate implicano la modifica dei dati.

Ambiti di utilizzo

Per tali ragioni le tecniche di de identificazione sono adoperate nei cd. “usi secondari”:

- Test del sw o di sistema;
- Statistiche;
- Pubblicazione dei dati a scopi di ricerca (PPDP);
- Analisi dati per conto terzi (PPDM);
- Accesso ed elaborazione dei dati particolari “veri” (ossia non cifrati) da personale autorizzato;
- Collegamento di più fonti dati per lo stesso interessato da parte del personale autorizzato.

Ordine attività

L'applicabilità dei principi può essere più semplice se svolto nel seguente ordine (anche se spesso si incorre in re iterazione della procedura):

1. Individuazione obblighi legali;
2. Individuazione obiettivi di business e liceità degli stessi per finalità/usi primari;
3. Individuazione obiettivi di business ed eventuali obblighi legali specifici per gli utilizzi secondari;
4. De identificazione per usi primari;
5. Definizione modalità, supporti e organizzazione usi primari;
6. De identificazione per usi primari (reiterazione se rischio elevato);
7. De identificazione per uno o più usi secondari;
8. Definizione modalità, supporti e organizzazione per uno o più usi secondari;
9. De identificazione per usi secondari;
10. Revisione ad opera del black team;
11. Rivalutazione complessiva dei rischi residui.

Metodi suggeriti

El Eman e Malin hanno suggerito un processo suddiviso in undici passi per procedere alla de identificazione:

1. Individuazione identificativi diretti;
2. Soppressione e pseudonimizzazione identificativi diretti;
3. Valutazione minacce e informazioni a disposizione;
4. Determinazione utilità minima dei dati;
5. Identificazione livello di rischio accettabile;
6. Estrazione parziale dei dati;
7. Valutazione rischio di reidentificazione;
8. Comparazione rischio valutato con la soglia;
9. Revisione tecniche usate se il rischio è troppo elevato e procedure con eventuali trasformazioni;
10. Diagnosi sui dati trasformati e valutazione utilità e il potenziale di re identificazione;
11. Rilascio dati per uso secondario.

Ordine logico di utilizzo delle tecniche

La scelta delle tecniche da utilizzare nei singoli casi non è affatto banale in quanto occorre diminuire il rischio di re identificazione con mantenimento di altre proprietà dei dati che danno valore al loro utilizzo.

Ordine logico suggerito

1. Soppressione (mascheratura);
2. Campionamento;
3. Aggregazione;
4. Classificazione attributi (gli identificativi diretti vanno rimossi/pseudonimizzati/crittografati);
5. Gestione altri attributi (identificativi indiretti e sensibili).

Valutazione rischio residuo

Due approcci

Formale: tra le soluzioni formali si hanno modelli “matematici-logici” complessi ed astratti.

1. K anonimization: un insieme di dati è K anonimizzato se per ogni combinazione possibile di valori degli attributi ci sono almeno k-1 interessati (più è alto il k maggiore è il potere identificativo). E’ un metodo efficiente per prevenire gli attacchi dell’accusatore (non è però adatto ad altri tipi di attacco).
2. L diversity: questo metodo aggiunge il requisito che gli attributi sensibili per ciascuna classe di equivalenza abbiano almeno i diversi valori (ciò aggiunge robustezza agli attacchi inferenziali).

Valutazione rischio residuo

Pratico

Tra le soluzioni pratiche quella più significativa è la seguente:

Metodo HIPAA Safe Harbour Method: consiste nella rimozione dalle informazioni mediche diciotto tipologie di dati.

Caratteristiche

- metodo basato unicamente sulla mascheratura e sulla generalizzazione;
- non fornisce sempre garanzie;
- è applicabile essenzialmente in fase di pubblicazione di dati aggregati.

HIPAA Safe Harbour Method

1. Nomi;
2. Tutte le suddivisioni geografiche più piccole di uno Stato (Americano), eccetto le prime tre cifre del codice di avviamento postale se sussistono determinate condizioni;
3. Tutti i riferimenti temporali collegati ai pazienti sono generalizzati all'anno;
4. Numeri di telefono;
5. Numeri di fax;
6. Indirizzi mail;
7. Social Security Number;
8. Numero di registrazione medica;
9. Health Plan Beneficiary Number;
10. Numero di conto;
11. Numeri di certificati, patente di guida;
12. Targa di autoveicolo e numero di telaio;
13. Numeri identificativi di apparati;
14. URL;
15. Indirizzi IP;
16. Identificativi biometrici;
17. Immagini della faccia;
18. Ogni altro codice identificativo o caratteristica.

Approfondimenti

- Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymization Tecniques, April 2014
- EDPS-AEPD-10 MISUNDERSTANDINGS RELATED TO ANONYMISATION
- ICO Anonymisation: managing data protection risk code of practice
- ICO Introduction to anonymisation Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance
- ICO Draft Chapter 2: How do we ensure anonymisation is effective?
- AEPD- K anonymity
- ISO/IEC 20889 Privacy enhancing data de identification terminology and classification of tecnicques
- ISO 25237 Health Informatics – Pseudonymization
- NIST, NISTIR 8053 De identification of Personal Information, October 2015
- DICOM PS3:15 2016 Annex E

Strumenti sw

- https://arx.deidentifier.org/overview/related-software/

Esempi di utilizzo

Quali sono le tecniche da utilizzare?

1. Campionamento;
2. Tecniche di pseudonimizzazione;
3. Mascheratura;
4. Tecniche di generalizzazione;
5. Soppressione locale;
6. Soppressione di record.

Esempi di utilizzo

Tutte le tecniche di de identificazione indicate potranno/dovranno essere affiancate a tecniche crittografiche almeno fino alla fine del controllo qualità dei dati al fine di fornire una protezione adeguata dei dati.

Conclusioni

- La protezione dei dati è necessaria;
- I potenziali tipi di attacco sono numerosi e diversificati;
- Le tecniche di de identificazione variano a seconda di:

1. Tipo di utilizzo;
2. Contesto;
3. Numero di interessati;
4. Tipologia di dato

- Le tecniche di de identificazione sono numerose, e ognuna presenta vantaggi e svantaggi;
- L'ordine logico di utilizzo delle tecniche di de identificazione è il seguente:

1. Soppressione (mascheratura);
2. Campionamento;
3. Aggregazione;
4. Classificazione attributi (gli identificativi diretti vanno rimossi/pseudonimizzati/crittografati);
5. Gestione altri attributi (identificativi indiretti e sensibili).

- Occorre sempre valutare il rischio residuo!

Proviamo insieme

Utilizzando le informazioni contenute nella seguente tabella:

SDO	Codice fiscale	Nome	Cognome	Patologia	Età	Misura
1	jsdhadjhasncas	Giulio	Bianchi	Ipertensione	38	111
2	dnadasmdams	Fausto	Rossi	Cardiopatia	76	176
3	sadasdagfgdsf	Lorena	Bianchi	Ipertensione	17	111
4	asdafsrgbbvsd	Salvo	Rossi	Cardiopatia	78	104
5	asdafrgsddvfa	Cesare	Bianchi	Aritmia	46	189
6	asagdrhdbdds	Arianna	Conte	Ipertensione	32	191
7	asfrgbsvewt5	Franca	Vecchi	Ipertensione	53	183
8	gadfg5y5tfds	Silvia	Vecchi	Cardiopatia	30	172
9	dasgetafa34	Ramona	Bianchi	Ipertensione	72	139
10	zfdfsgsrty65y	Anneo	Rossi	Ipertensione	49	114
11	dsgadfgaergav	Giulio	Conte	Cardiopatia	36	104
12	adfdsfgererrg	Marta	Bianchi	Aritmia	75	117

Proviamo insieme

1. Si individuino un **identificativo diretto** e un **identificativo indiretto**;
2. Si selezioni un gruppo di record utilizzando la **tecnica del campionamento**;
3. Si utilizzi la **tecnica dell'aggregazione**;
4. Si utilizzi la **tecnica della mascheratura** per anonimizzare il campione. In seguito si usi la **tecnica della generalizzazione** per categorizzare i record per fasce di età;
5. Si applichi la **tecnica della soppressione** di record per eliminare i valori rari.

Proviamo insieme

Soluzioni:

1. Il Codice Fiscale è identificativo diretto. La sdo è identificativo indiretto e quindi attributo chiave;
2. Un esempio di selezione potrebbe essere il seguente:

SDO	Codice fiscale	Nome	Cognome	Patologia	Età	Misura
2	dnadasmdams	Fausto	Rossi	Cardiopatia	76	176
5	asdafrgsddvfa	Cesare	Bianchi	Aritmia	46	189
7	asfrgbsvewt5	Franca	Vecchi	Ipertensione	53	183
8	gadfg5y5tfds	Silvia	Vecchi	Cardiopatia	30	172
10	zfdfsgsrty65y	Anneo	Rossi	Ipertensione	49	114
12	adfdsfgererrg	Marta	Bianchi	Aritmia	75	117

Proviamo insieme

3. Utilizzo la tecnica dell'aggregazione:

SDO	Codice fiscale	Nome	Cognome	Patologia	Età	Misura
1	jsdhadjhasncas	Giulio	Bianchi	Ipertensione	38	141.5
2	dnadasmdams	Fausto	Rossi	Cardiopatia	76	139
3	sadasdagfgdsf	Lorena	Bianchi	Ipertensione	17	141.5
4	asdafsrgbbvsd	Salvo	Rossi	Cardiopatia	78	139
5	asdafrgsddvfa	Cesare	Bianchi	Aritmia	46	153
6	asagdrhdbdds	Arianna	Conte	Ipertensione	32	141.5
7	asfrgbsvewt5	Franca	Vecchi	Ipertensione	53	141.5
8	gadfg5y5tfds	Silvia	Vecchi	Cardiopatia	30	139
9	dasgetafa34	Ramona	Bianchi	Ipertensione	72	141.5
10	zfdfsgsrty65y	Anneo	Rossi	Ipertensione	49	141.5
11	dsgadfgaergav	Giulio	Conte	Cardiopatia	36	139
12	adfdsfgererrg	Marta	Bianchi	Aritmia	75	153

Proviamo insieme

4. Utilizzo la tecnica della mascheratura:

SDO	Patologia	Età	Misura
1	Ipertensione	38	111
2	Cardiopatia	76	176
3	Ipertensione	17	111
4	Cardiopatia	78	104
5	Aritmia	46	189
6	Ipertensione	32	191
7	Ipertensione	53	183
8	Cardiopatia	30	172
9	Ipertensione	72	139
10	Ipertensione	49	114
11	Cardiopatia	36	104
12	Aritmia	75	117

Proviamo insieme

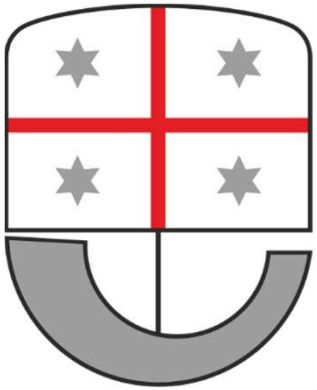
Utilizzo della tecnica della generalizzazione :

SDO	Patologia	Età	Misura
1	Ipertensione	30<età<40	111
2	Cardiopatia	70<età<80	176
3	Ipertensione	10<età<20	111
4	Cardiopatia	70<età<80	104
5	Aritmia	40<età<50	189
6	Ipertensione	30<età<40	191
7	Ipertensione	50<età<60	183
8	Cardiopatia	30<età<40	172
9	Ipertensione	70<età<80	139
10	Ipertensione	40<età<50	114
11	Cardiopatia	30<età<40	104
12	Aritmia	70<età<80	117

Proviamo insieme

5. Utilizzo la tecnica della soppressione di record:

SDO	Codice fiscale	Nome	Cognome	Patologia	Età	Misura
1	jsdhadjhasncas	Giulio	Bianchi	Ipertensione	38	111
2	dnadasmdams	Fausto	Rossi	Cardiopatia	76	176
3	sadasdagfgdsf	Lorena	Bianchi	Ipertensione	17	111
4	asdafsrgbbvsd	Salvo	Rossi	Cardiopatia	78	104
6	asagdrhdbdds	Arianna	Conte	Ipertensione	32	191
7	asfrgbsvewt5	Franca	Vecchi	Ipertensione	53	183
8	gadfg5y5tfds	Silvia	Vecchi	Cardiopatia	30	172
9	dasgetafa34	Ramona	Bianchi	Ipertensione	72	139
10	zfdfsgsrty65y	Anneo	Rossi	Ipertensione	49	114
11	dsgadfgaergav	Giulio	Conte	Cardiopatia	36	104



Liguria Digitale

Grazie dell'attenzione