La condizione di anonimato è un concetto statico o dinamico?

di Riccardo Abeti

Presidente della Commissione

"New Technology, Personal Data and Communication Law" dell'UAE e

founding partner di EXP legal

agenda

- quadro normativo
- definizioni
- funzione e tecniche dell'anonimizzazione
- metodi diversi un unico comun denominatore
- casi di studio
- conclusioni

quadro normativo

95/46/CE

considerando che ...

(26) considerando che i principi della tutela si devono applicare ad ogni informazione concernente una persona identificata o identificabile; che, per determinare se una persona è identificabile, è opportuno prendere in considerazione l'insieme dei mezzi che possono essere ragionevolmente utilizzati dal responsabile del trattamento o da altri per identificare detta persona; che i principi della tutela non si applicano a dati resi anonimi in modo tale che la persona interessata non è più identificabile; che i codici di condotta ai sensi dell'articolo 27 possono costituire uno strumento utile di orientamento sui mezzi grazie ai quali dati possano essere resi anonimi e registrati in modo da rendere impossibile l'identificazione della persona interessata;

quadro normativo

02/58/CE

considerando che ...

(26) I dati relativi al traffico utilizzati per la commercializzazione dei servizi di comunicazione o per la fornitura di servizi a valore aggiunto dovrebbero inoltre essere cancellati o resi anonimi dopo che il servizio è stato fornito;

quadro normativo

Le direttive europee non specificano il come

ma il cosa si deve ottenere …

l'anonimizzazione deve essere permanente,

come la cancellazione

quadro normativo

d.lgs. 30 giugno 2003, n. 196

Articolo 4, comma 1

b) "dato personale", qualunque informazione relativa a persona fisica, identificata o identificabile, anche indirettamente, mediante riferimento a qualsiasi altra informazione, ivi compreso un numero di identificazione personale;

[...]

n) "dato anonimo", il dato che in origine, o a seguito di trattamento, non può essere associato ad un interessato identificato o identificabile;

quadro normativo

WP216

GRUPPO DI LAVORO ARTICOLO 29 PER LA PROTEZIONE DEI DATI

0829/14/IT
WP216

Parere 05/2014 sulle tecniche di anonimizzazione

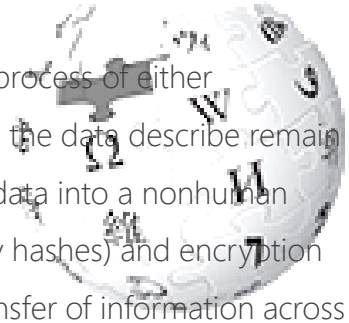
adottato il 10 aprile 2014

qualche definizione

Data anonymization

WIKIPEDIA
The Free Encyclopedia

Not to be confused with Data cleansing.



Data anonymization is a type of information sanitization whose intent is privacy protection. It is the process of either encrypting or removing personally identifiable information from data sets, so that the people whom the data describe remain anonymous. The Privacy Technology Focus Group defines it as "technology that converts clear text data into a nonhuman readable and irreversible form, including but not limited to preimage resistant hashes (e.g., one-way hashes) and encryption techniques in which the decryption key has been discarded."[1] Data anonymization enables the transfer of information across a boundary, such as between two departments within an agency or between two agencies, while reducing the risk of unintended disclosure, and in certain environments in a manner that enables evaluation and analytics post-anonymization. In the context of medical data, anonymized data refers to data from which the patient cannot be identified by the recipient of the information. The name, address, and full post code must be removed together with any other information which, in conjunction with other data held by or disclosed to the recipient, could identify the patient.[2] De-anonymization is the reverse process in which anonymous data is cross-referenced with other data sources to re-identify the anonymous data source.[3] Generalization and perturbation are the two popular anonymization approaches for relational data.[4]

qualche definizione

Data cleansing

WIKIPEDIA
The Free Encyclopedia

Data cleansing, data cleaning or data scrubbing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database. Used mainly in databases, the term refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. parts of the data and then replacing, modifying, or deleting this dirty data or coarse data.[1]

After cleansing, a data set will be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different stores.

Data cleansing differs from data validation in that validation almost invariably means data is rejected from the system at entry and is performed at entry time, rather than on batches of data.

The actual process of data cleansing may involve removing typographical errors or validating and correcting values against a known list of entities. The validation may be strict (such as rejecting any address that does not have a valid postal code) or fuzzy (such as correcting records that partially match existing, known records).

Some data cleansing solutions will clean data by cross checking with a validated data set. Also data enhancement, where data is made more complete by adding related information, is a common data cleansing practice. For example, appending addresses with phone numbers related to that address.

qualche definizione

Data cleansing

WIKIPEDIA
The Free Encyclopedia

Data cleansing, data cleaning or data scrubbing is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database. Used mainly in databases, the term refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. parts of the data and then replacing, modifying, or deleting this dirty data or coarse data.[1]

After cleansing, a data set will be consistent with other similar data sets in the system. The inconsistencies detected or removed may have been originally caused by user entry errors, by corruption in transmission or storage, or by different data dictionary definitions of similar entities in different stores.

Data cleansing differs from data validation in that validation almost invariably means data is rejected from the system at entry and is performed at entry time, rather than on batches of data.

The actual process of data cleansing may involve removing typographical errors or validating and correcting values against a known list of entities. The validation may be strict (such as rejecting any address that does not have a valid postal code) or fuzzy (such as correcting records that partially match existing, known records).

Some data cleansing solutions will clean data by cross checking with a validated data set. Also data enhancement, where data is made more complete by adding related information, is a common data cleansing practice. For example, appending addresses with phone numbers related to that address.

qualche definizione

Pseudonymization

Con questo termine si fa riferimento alla riduzione della correlabiltià di un insieme di dati all'identità originaria

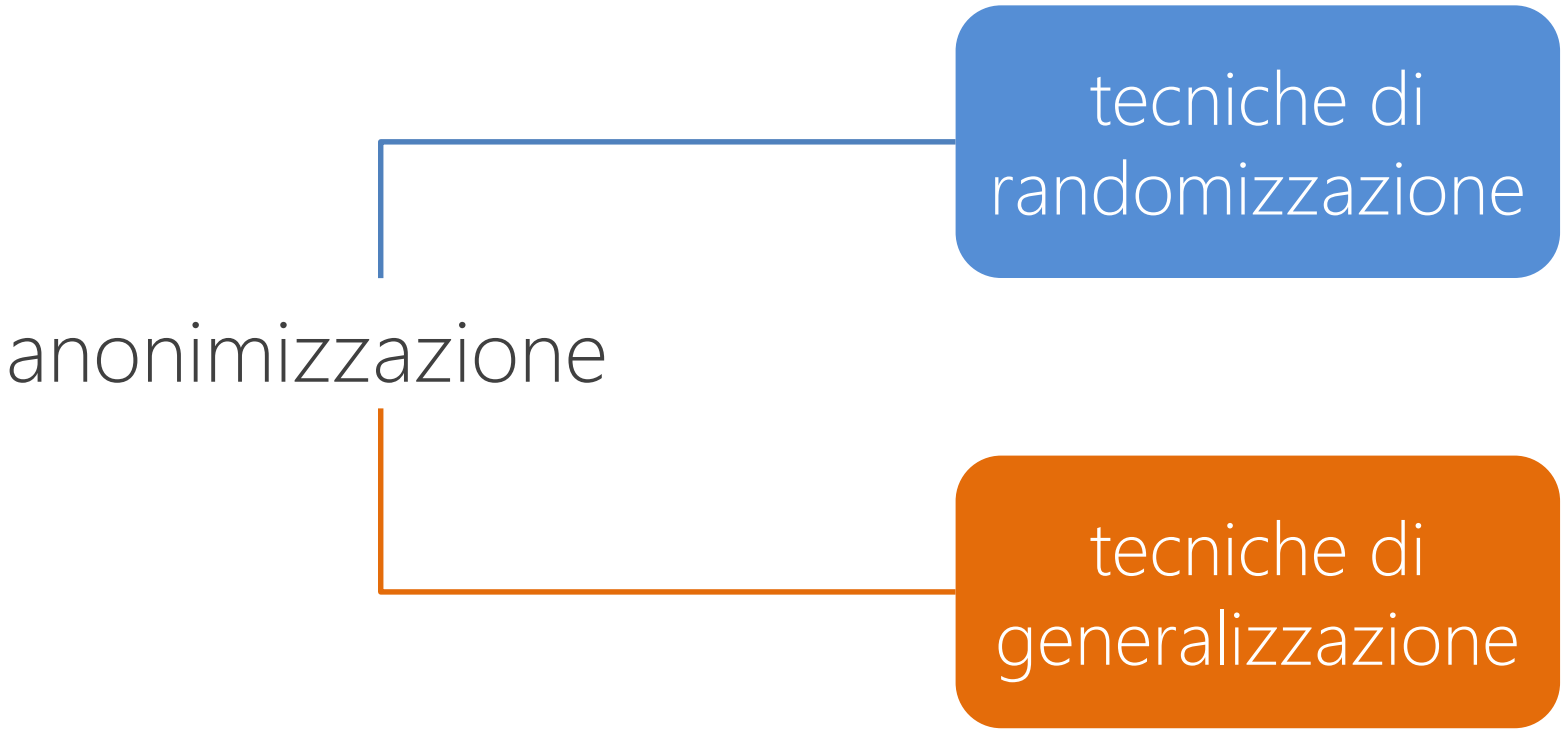
funzione e tecniche dell'anonimizzazione

Il valore potenziale dell'anonimizzazione è quello di consentire i molteplici utilizzi delle informazioni attenuando i rischi, in termini di protezione dei dati, per le persone interessate

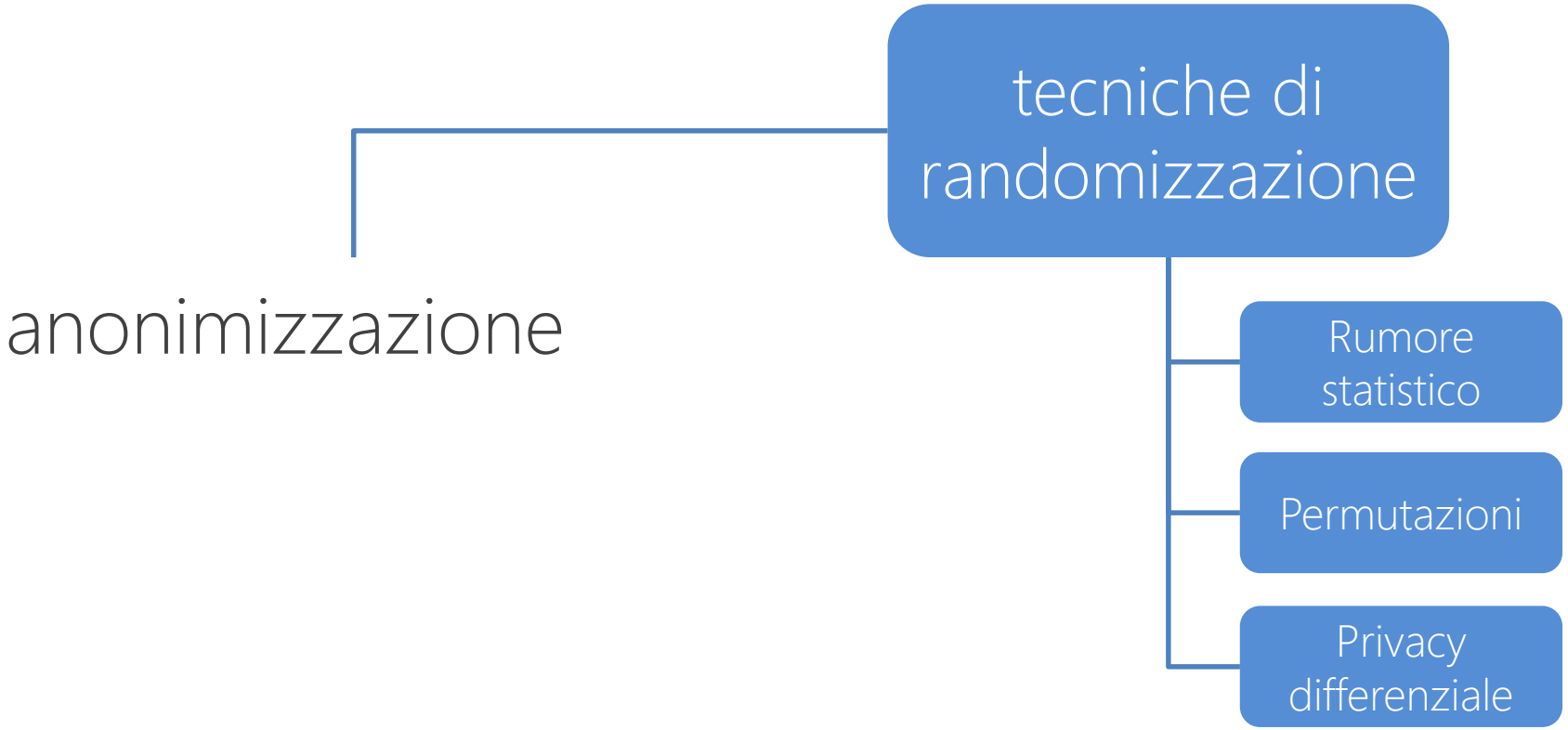
funzione e tecniche dell'anonimizzazione

L'effetto atteso dall'applicazione di un procedimento di anonimizzazione è quello di impedire irreversibilmente l'identificazione dei soggetti rappresentati ...

funzione e tecniche dell'anonimizzazione



funzione e tecniche dell'anonimizzazione



funzione e tecniche dell'anonimizzazione

anonimizzazione

- tecniche di generalizzazione
  - Aggregazione
  - K-anonimato
  - L-diversità
  - T-vicinanza

funzione e tecniche dell'anonimizzazione

In questo quadro che ruolo hanno i quasi-indicatori?

funzione e tecniche dell'anonimizzazione

La soluzione ottimale dovrebbe essere decisa caso per caso ...

funzione e tecniche dell'anonimizzazione

… se possibile utilizzando una combinazione di tecniche diverse, tenendo sempre presente:

- ❑ i fattori contestuali e
- ❑ i fattori circostanziali;

inoltre …

funzione e tecniche dell'anonimizzazione

… inoltre l'anonimizzazione presuppone un riesame periodico dei rischi di re-identificazione, allo stato dell'evoluzione delle tecniche utilizzate e della disponibilità di nuove fonti informative …

funzione e tecniche dell'anonimizzazione

In sostanza l'obiettivo da raggiungere è configurabile nella "ragionevole impossibilità" di reidentificare gli interessati e nella considerazione che in questa materia:

le ricerche + gli strumenti + la potenza di calcolo

sono in continua evoluzione!!!

funzione e tecniche dell'anonimizzazione

dato il rischio intrinseco nel processo di anonimizzazione del dato esso deve essere applicato ad un trattamento legittimo … per questo il processo di anonimizzazione viene definite un "trattamento successivo" …

funzione e tecniche dell'anonimizzazione

Per esempio, nel caso dei titolari del trattamento cui si applichi la direttiva relative alla vita private e alle comunicazioni elettroniche, se un trattamento non è ammesso ex articolo 6 (direttiva 58 del 2002) non può sussistere interesse legittimo ex articolo 7, comma 1, let. f) (direttiva 46 del 1995)

funzione e tecniche dell'anonimizzazione

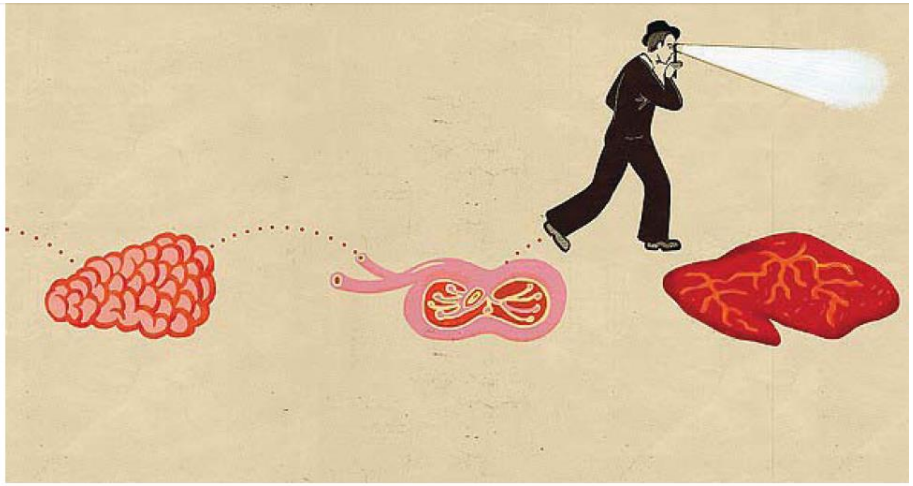
Attenzione poi alla perenne "oscillazione" tra l'interesse legittimo del "data controller" (ovvero il titolare del trattamento secondo la norma italiana) e il diritto dell'interessato ...

metodi diversi un unico comun denominatore

	Sussiste rischio individuazione?	Sussiste rischio correlabilità?	Sussiste rischio deduzione?
Pseudonimizzazione	Si	Si	Si
Rumore statistico	Si	Forse no	Forse no
Sostituzione	Si	Si	Forse no
K-anonimato	No	Si	Si
L-diversità	No	Si	Forse no
Privacy differenziale	Forse no	Forse no	Forse no
Hashing	si	Si	Forse no

cases

dalle schede di dimissione ospedaliere. Combinati con le liste elettorali, reperibili a poco prezzo, consentono di risalire con precisione quasi assoluta a nomi e cognomi dei ricoverati. Una pacchia per chi voglia conoscere la storia sanitaria di una persona prima di assumerla, erogarle un mutuo, venderle prodotti mirati



Aiuto, la nostra salute non è un segreto

Una falla del sistema permette di identificare i pazienti Così datori di lavoro e industrie potrebbero servirsene

di SERENA DANNA *e* SIMONA RAVIZZA

I dati personali sono l'oro del terzo millennio. Il nuovo petrolio, come li definisce il World Economic Forum report. Li vogliono e li cercano — attraverso l'uso delle nuove tecnologie — imprese, società di marketing, banche e assicurazioni desi-

di residenza. Sono tre elementi che, combinati con le banche dati dell'anagrafe, consentono di risalire al paziente. A ciascuno di noi.

I tre «quasi-identificatori» dovrebbero essere chiusi in cassaforte. Così non è. Ricercatori di importanti isti-

ILLUSTRAZIONE DI BEPPE GIACOBBE

disporre di questa variabile e assicura il trattamento corretto del dato secondo la normativa sulla protezione dei dati personali e sensibili, data di nascita, comune di residenza e sesso dei pazienti possono essere rilasciati insieme».

cases

conclusioni

Quali dati possiamo dunque ritenere, certamente, anonimi?

E quanti possiamo ritenerli "al sicuro" in quanto anonimi?

correlazioni

finalità

spionaggio

hacktivism

Grazie per l'attenzione!
@riccardoabeti

backup

@riccardoabeti

titolo

95/46/CE

Considerando che …

(26) Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible;

passive data

un nuovo livello di informazioni …