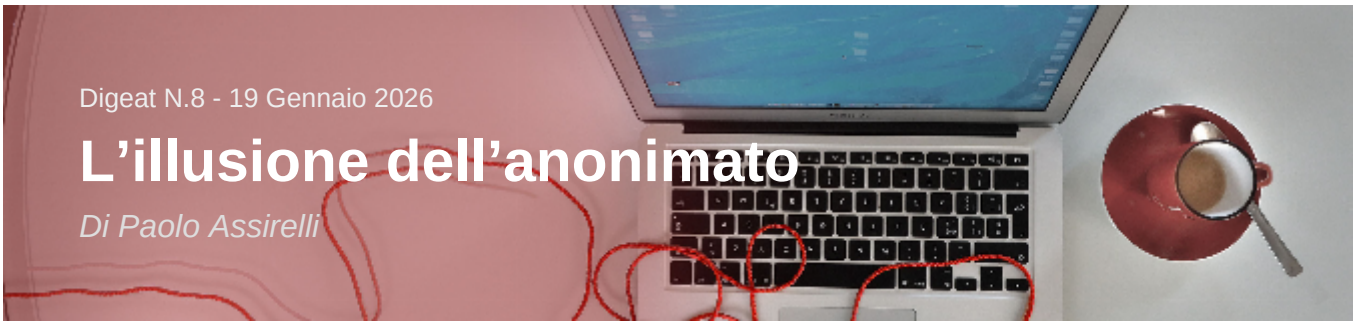

Digeat N.8 - 19 Gennaio 2026

# L’illusione dell’anonimato

*Di Paolo Assirelli*

## Abstract

La digitalizzazione sanitaria ha reso critico l’equilibrio tra innovazione e tutela dei diritti fondamentali. Nonostante i progressi nelle tecniche di anonimizzazione, la re-identificazione dei dati sanitari rimane concretamente possibile per l’aumento della granularità, interoperabilità dei sistemi e potenza dei big data analytics. Le recenti evoluzioni normative – dal GDPR al progetto europeo EHDS – evidenziano i limiti delle tradizionali misure tecniche e spingono verso un approccio risk-based, dove pseudonimizzazione robusta, DPIA specifiche e governance multilivello assumono un ruolo centrale. L’articolo propone un’analisi multidisciplinare della dicotomia tra anonimato e identificabilità in ambito sanitario.

## Indice



## Introduzione

**La rivoluzione digitale nel settore sanitario ha trasformato radicalmente l’ecosistema della gestione delle informazioni cliniche**. L’adozione degli strumenti digitali[1] ha portato a una crescita esponenziale dei dati generati, raccolti ed elaborati[2], con una dinamica superiore alla media di qualsiasi altro comparto, cui l’evoluzione dell’IA ha dato ulteriore spinta.

Questo dataset costituisce una risorsa straordinaria per ricerca scientifica, sanità pubblica, epidemiologia e personalizzazione delle cure, **ma la natura intrinsecamente sensibile dei dati[3] impone cautele particolari per la vulnerabilità** degli interessati **e il potenziale lesivo** derivante da incidenti per data breach o usi illeciti.[4]

È in questa **dicotomia** che la distinzione tra **anonimizzazione** e **deidentificazione** spiega la sua importanza: **la prima prevede trasformazione irreversibile dei dati** escludendo qualsiasi ricollegamento all’interessato; **la seconda lascia margini a possibili re-identificazioni**, specie in ambito sanitario caratterizzato da alta granularità e combinabilità delle informazioni.

> La praticabilità effettiva dell'anonimizzazione nel contesto sanitario è oggetto di profonda revisione critica, evidenziandone l'inadeguatezza strutturale di fronte a dataset longitudinali, omogenei o genomici**.**

**Le soluzioni di governance più efficaci si basano su combinazione multilivello di misure tecniche** (pseudonimizzazione, minimizzazione, encryption), **organizzative** (DPIA, data access committees, audit etici) **e giuridiche**.

## Il quadro normativo

Il GDPR rappresenta la pietra miliare nella protezione dei dati personali. L'art. 4 fornisce definizioni fondamentali; l'art. 5 stabilisce principi di liceità, trasparenza, minimizzazione, limitazione della conservazione.

> **Centrale è la nozione di "identificabilità": un dato è personale se permette l'identificazione diretta o indiretta, anche tramite combinazione con altre informazioni.**

L'art. 9 introduce il divieto di trattamento dei dati sulla salute, salve deroghe. Il Considerando 26 chiarisce che il GDPR non si applica ai dati anonimizzati, impossibili da ricondurre a individui "utilizzando mezzi ragionevolmente a disposizione". **La pseudonimizzazione è misura di sicurezza, non garanzia di uscita dal perimetro GDPR.** L'art. 89 disciplina finalità scientifiche e statistiche.

Il quadro normativo e interpretativo in materia di dati sanitari si è progressivamente arricchito e articolato su più livelli. A livello nazionale, il Codice in materia di Protezione dei Dati Personali (D.Lgs. 196/2003), adeguato al Regolamento (UE) 2016/679, contiene oggi disposizioni di dettaglio specificamente dedicate all'ambito sanitario, confermando la necessità di garanzie rafforzate per il trattamento di dati particolarmente sensibili.

In questo contesto, il Garante per la Protezione dei Dati Personali ha svolto un ruolo centrale, definendo criteri puntuali in materia di sicurezza, valutazione d'impatto e modelli di governance, con particolare riferimento al Fascicolo Sanitario Elettronico 2.0, quale infrastruttura strategica del sistema sanitario digitale.

Sul piano europeo, **la proposta di Regolamento sullo Spazio Europeo dei Dati Sanitari (European Health Data Space – EHDS) segna un ulteriore passo in avanti, promuovendo l'interoperabilità tra i sistemi nazionali e introducendo modelli di consenso più evoluti**, capaci di coniugare circolazione dei dati, tutela dei diritti e finalità di interesse pubblico. A ciò si affianca l'apporto degli standard tecnici internazionali, come la norma ISO/IEC 20889:2018, che fornisce metodologie e criteri per valutare il rischio di re-identificazione, chiarendo che l'anonimizzazione non può essere considerata un risultato meramente formale, ma richiede una valutazione concreta e dinamica del rischio.

La giurisprudenza della Corte di Giustizia dell'Unione europea ha ulteriormente contribuito a definire i confini applicativi della disciplina, precisando la distinzione tra dato anonimo e dato pseudonimizzato e ribadendo che quest'ultimo rientra, a pieno titolo, nell'ambito di applicazione del GDPR. In modo coerente, le Opinion del Gruppo Articolo 29 e le successive Linee guida dell'European Data Protection Board (EDPB) hanno consolidato tali principi interpretativi, offrendo indicazioni operative essenziali per autorità, operatori e titolari del trattamento.

**Questo complesso di fonti non resta confinato a un piano teorico, ma trova concreta applicazione in strumenti e contesti operativi già attivi, quali il Fascicolo Sanitario Elettronico,**

**la disciplina delle biobanche e le policy regionali in materia di dati sanitari**, che rappresentano oggi veri e propri laboratori di sperimentazione giuridica, tecnologica e organizzativa della sanità digitale.

## Fondamenti tecnici e metodologie

**Nel trattamento dei dati sanitari avanzati non esistono soluzioni di anonimizzazione perfette: esistono solo tecniche più o meno efficaci**, ciascuna con limiti e rischi, che devono essere valutati, combinati e governati in modo consapevole, soprattutto in contesti ad alta intensità informativa come biobanche, studi clinici e sanità digitale. Le tecniche tradizionali comprendono generalizzazione, soppressione e aggregazione statistica, spesso insufficienti per i dati clinici moderni. Anche modelli teoricamente più avanzati quali la k-anonymity, la l-diversity e la t-closeness mostrano limiti significativi nella pratica applicativa, risultando vulnerabili a diverse forme di attacco, tra cui i linking attacks, i differencing attacks e gli attacchi basati sulla conoscenza pregressa del contesto (background knowledge attacks).[5] Tali criticità mettono in evidenza come la protezione dei dati non possa essere affidata esclusivamente a soluzioni formali o statiche.

In questo scenario, **la differential privacy offre maggiore protezione in contesti epidemiologici, in quanto rafforza la tutela dell'individuo anche in presenza di analisi su larga scala**, ma implica trade-off tra utilità scientifica e protezione poiché comporta inevitabili compromessi tra il livello di protezione dei dati personali e la qualità informativa e scientifica dei risultati ottenuti.  Numerosi casi di studio hanno dimostrato come anche dataset resi pubblici e qualificati come anonimi possano consentire la re-identificazione degli interessati, usando tecniche di data triangulation – ossia l'incrocio di più fonti informative e l'impiego di algoritmi predittivi sempre più sofisticati, che permettono di ricostruire profili individuali partendo da informazioni apparentemente innocue.

Con l'obiettivo di di ridurre il rischio di identificazione diretta si collocano le pratiche di deidentificazione e pseudonimizzazione che utilizzano architetture crittografiche, sistemi di tokenizzazione e tecniche di data masking.[6]

Il rischio di re-identificazione risulta ulteriormente amplificato laddove appaia una elevata  granularità come accade in presenza dei dati genomici, longitudinali e temporali. In questo contesto, biobanche, trial clinici e registri epidemiologici rappresentano spazi nei quali innovazione scientifica e rischio privacy convivono rappresentando dei veri e propri laboratori dove innovazione e rischio sono inscindibili, dove l'impossibilità di anonimizzazione perfetta è riconosciuta. Proprio a partire da questa consapevolezza ci stiamo aprendo ad approcci tecnologici emergenti innovativi che includono il privacy-preserving record linkage, il federated learning e il synthetic data generation, sempre sotto controllo di risk assessment. **Ciò significa che tali soluzioni non eliminano il rischio ma lo ridimensionano e lo rendono gestibile attraverso un costante presidio con valutazioni strutturate e dinamiche del rischio basandosi su un modello di responsabilità, governance e controllo continuo.**

## Analisi critica della dicotomia

La "perfezione" dell'anonimizzazione è da considerarsi, ad oggi, una mitologia teorica.

> **Le evidenze giuridiche e tecniche, consolidate dalla letteratura scientifica dimostrano come il rischio di re-identificazione permanga anche in dataset apparentemente bonificati**, specie quando il background informativo o la struttura dei dati consente linkage imprevisti.

Il teorema **“no free lunch”**[7] della privacy **implica che ogni incremento di protezione riduce l’utilità del dato**. **Il paradosso utilità-privacy si esprime in un continuum, dove le gradazioni di rischio evolvono come funzione temporale e contestuale, con variabilità tecnologica non sempre governabile.**

La standardizzazione del concetto di “mezzi ragionevolmente utilizzabili” si scontra con fenomeni come la crescita delle risorse computazionali (quantum computing), l’amplificazione delle fonti open data e la frequente presenza di attributi “quasi-identificativi” (es. dati genomici), che rendono la vera anonimizzazione in sanità virtualmente impossibile.

I dati genomici, paradigmatici per unicità e inferenze su terzi, evidenziano l’insostenibilità della dissociazione totale.

> **La vera tutela consiste nell’affinamento multidisciplinare di minimizzazione, pseudonimizzazione robusta, data access committees e audit etici permanenti**.

## Applicazioni e casi d’uso nel settore sanitario

**L’impiego delle tecnologie di anonimizzazione e pseudonimizzazione trova applicazione diretta nelle principali infrastrutture digitali sanitarie italiane ed europee.** Il FSE 2.0 rappresenta laboratorio di governance avanzata, con separazione tra database identificativi e clinici, chiave crittografica protetta e controllo multilivello degli accessi.[8]

Nel FSE 2.0 la pseudonimizzazione consente l’uso secondario dei dati per finalità di governo clinico, statistica, controllo e ricerca, minimizzando il rischio di re-identificazione tramite architetture di sistema dedicate e valutazioni d’impatto sulla Protezione dei Dati (DPIA) specifiche. I modelli di consenso, sempre più granulari e modulabili, integrano informativa differenziata e scelte personali orientate a tutela, con la supervisione del Garante Privacy.

**La ricerca biomedica, i trial clinici e gli studi osservazionali adottano protocolli di pseudonimizzazione differenziata**: i biocampioni e i dati associati sono separati, mentre i ricercatori hanno accessi differenziati secondo necessità scientifica e misure di sicurezza. Le biobanche di patologia, genomiche e popolazionali utilizzano comitati di data access e audit etici, con revisione costante delle policy di minimizzazione e validazione scientifica dei dataset.

La Real World Evidence – basata sull’analisi di dati osservativi provenienti e raccolti dalla pratica clinica ordinaria e quotidiana – richiede modelli di gestione dei dati costruiti secondo i principi di privacy by design. In tale prospettiva, assumono un ruolo centrale approcci come il federated learning e forme strutturate di data sharing internazionale. La validazione degli algoritmi di IA in medicina, sia per set di training che di test, è sottoposta a processi di sintetizzazione dati e decentramento, finalizzati a conservare validità scientifica e protezione individuale delle persone.[9]

La pandemia COVID-19 ha offerto un caso paradigmatico: sistemi di sorveglianza epidemiologica e contact tracing hanno imposto bilanciamenti emergenza-privacy, con ricorso temporaneo a deroghe normative (art. 9 GDPR, DL n. 28/2020) e implementazione di tecniche di minimizzazione, pseudonimizzazione, tracciamento decentralizzato e informativa pubblica.[10]

**Medicina personalizzata e farmacogenomica richiedono consenso informato specifico e audit etici per la non dissociabilità tra informazione genetica e identità.**[11]

## Profili di responsabilità e governance

**La responsabilità nel trattamento dei dati sanitari coinvolge molteplici attori**: titolare, responsabile, sub-responsabili, DPO sanitario. Il contratto tra le parti deve definire ruoli, compiti, ambiti di accesso e modalità di esercizio della supervisione. Fondamentali sono la tenuta e l'aggiornamento del registro dei trattamenti, la pianificazione obbligatoria delle DPIA, la formazione continua e le misure tecniche e organizzative proporzionate ai rischi.[12]

**La compliance si realizza tramite audit periodici, certificazioni ISO, procedure di notifica e gestione del data breach (artt. 33-34 GDPR), segnalazioni e reportistica al Garante Privacy**, con particolare attenzione alle sanzioni amministrative e pecuniarie sempre più stringenti nei casi di violazione nel settore sanitario.[13]

> I modelli di privacy by design e by default sono adottati in architetture FSE2.0, biobanche, trial clinici, dove la DPIA non è solo obbligatoria, ma si trasforma in strumento di indirizzo tecnico-scientifico e di governance. I comitati etici e i data access committees svolgono funzioni cruciali di supervisione, allocazione rischi e revisione dei processi: il coinvolgimento multidisciplinare (clinici, giuristi, informatici, bioeticisti) è elemento determinante per l'efficacia.[14] La formazione del personale sanitario, la deontologia professionale, la consapevolezza informatica e giuridica sono oggi parte integrante dei percorsi di aggiornamento e accreditamento, orientando la cultura della responsabilità in modo trasversale e costante.

## Dimensioni etiche e deontologiche

La gestione dei dati sanitari è permeata dai princìpi bioetici di autonomia informativa, beneficialità, non maleficenza, giustizia distributiva, solidarietà e bene comune. Il rispetto delle scelte e delle aspettative del paziente è principio-guida, come pure la trasparenza comunicativa e la fiducia nel sistema sanitario.

> **La percezione della privacy nella società contemporanea incide direttamente sulla fiducia verso le istituzioni: la trasparenza dei processi, la chiarezza dell'informativa e la rendicontazione sull'uso secondario dei dati sono condizioni necessarie per il mantenimento di un rapporto fiduciario.**

La deontologia impone obblighi specifici al personale sanitario: segreto professionale, responsabilità verso i pazienti, formazione etica continua. Ugualmente i data scientist, i responsabili informatici e i giuristi sono tenuti a rispettare principi di prudenza, proporzionalità e competenza interdisciplinare. Il ruolo del legale nella data governance risulta centrale nel bilanciamento tra interessi contrapposti, nel disegno delle policy aziendali e nel presidio della compliance normativa e etica.

## Conclusioni

**L'anonimizzazione perfetta dei dati sanitari è sostanzialmente impraticabile.** Il rischio di re-identificazione persiste in dataset apparentemente bonificati. Il dato pseudonimizzato rimane dato personale finché può essere ricondotto a un individuo tramite informazioni aggiuntive.

La soluzione passa da pseudonimizzazione robusta costantemente aggiornata. Il teorema "no free lunch" implica che ogni incremento di protezione riduce l'utilità del dato. La prospettiva futura impone approcci risk-based multidisciplinari con valutazione caso per caso.

L’EHDS e la L. 132/2025 consolidano questa direzione. La vera tutela non risiede nell’illusione dell’anonimato, ma nell’affinamento multidisciplinare di minimizzazione, pseudonimizzazione robusta, Data Access Committees e audit etici permanenti.

La gestione di campioni biologici e dati genetici impone bilanciamento tra interesse pubblico alla ricerca e diritti all’identità, poiché il materiale biologico contiene il genoma, parte dell’identità della persona.

> **Solo una governance avanzata garantisce innovazione responsabile e sostenibilità del sistema sanitario digitale, assicurando sviluppo antropocentrico e trasparente delle tecnologie.**

---

**NOTE**

[1] FSE 2.0, registri di patologia, piattaforme di telemedicina.

[2] Volume globale dati sanitari digitali: crescita annua ~37% nel 2025.

[3] Informazioni su salute, dati genetici e biometrici (art. 9 GDPR).

[4] Ransomware ospedalieri, vulnerabilità portali FSE, re-identificazione dataset “anonimizzati”.

[5] R. Nobile, Differential Privacy, DPCE online 2/2024.

[6] A. Bozzo, Pseudonimizzazione, Diritto Rovescio 2.09.2024.

[7] Wolpert e Macready: “due algoritmi sono equivalenti mediando su tutti i possibili problemi”.

[8] Architettura EDS, ICT Security Magazine 8.03.2025; L. Baldacci, cit.

[9] Synthetic hEalthcare dAta goveRnanCe Hub, CE 13.09.2024; Biobanca UPO, D.R. 1031/7.07.2021.

[10] Contact tracing Covid-19, Diritto dell’Informatica 27.03.2020; Garante Privacy 2020-2022.

[11] R. Di Giammarco, Privacy Dati Genetici, DNA Express 25.07.2025.

[12] Registro trattamenti: garanteprivacy.it/registro-delle-attivita-di-trattamento.

[13] Accountability: garanteprivacy.it/regolamentoue/approccio-basato-sul-rischio.

[14] E. Ferioli, M. Picozzi, Biobanche, Università Cattolica 30.08.2011.

## Bibliografia


- GDPR, D.Lgs. 196/2003, EHDS proposal, Standard ISO/IEC 20889:2018
- Sentenze CGUE “Deloitte”, Provvedimenti Garante Privacy 2025
- WP29 Opinion n. 05/2014, EDPB Guidelines 01/2022 e 01/2025.
- Lucio Baldacci, *L’evoluzione dal Fascicolo Sanitario Elettronico 2.0 all’Ecosistema Dati Sanitari in Regione Umbria*.
- Giorgia Bianchini, *Trattamento dei dati sanitari: FSE 2.0 e l’intervento del Garante privacy*, in Nt+Diritto del 17 Luglio 2024.
- Alberto Bozzo, *Il Fascicolo Sanitario Elettronico 2.0*, in Diritto Rovescio del 4.08.2024.

- Alberto Bozzo , *La pseudonimizzazione nel trattamento dei dati personali,* in Diritto Rovescio, 2.09.2024.
- Roberto Di Giammarco, *Privacy dei Dati Genetici: La Protezione delle Informazioni più Intime dell'Essere Umano*, in DNA Express 25.07.2025.
- Elena Ferioli e Mario Picozzi, *La conservazione del materiale biologico finalizzato alla ricerca scientifica: questioni giuridiche e riflessioni etiche sulle biobanche*, Università Cattolica, 30.08.2011.
- Roberta Nobile, *Differential Privacy: la nuova frontiera per il miglioramento della privacy*, in DPCE online n. 2/2024.
- *Contact tracing e tutela della Privacy al tempo del Covid-19: è giusto controllare gli spostamenti dei cittadini attraverso la rete cellulare?*, in Diritto dell'Informatica, 27.03.2020.
- *Uso dei dati di localizzazione e degli strumenti per il tracciamento dei contatti nel contesto dell'emergenza legata al COVID-19*, Garante Privacy 2020-2022.