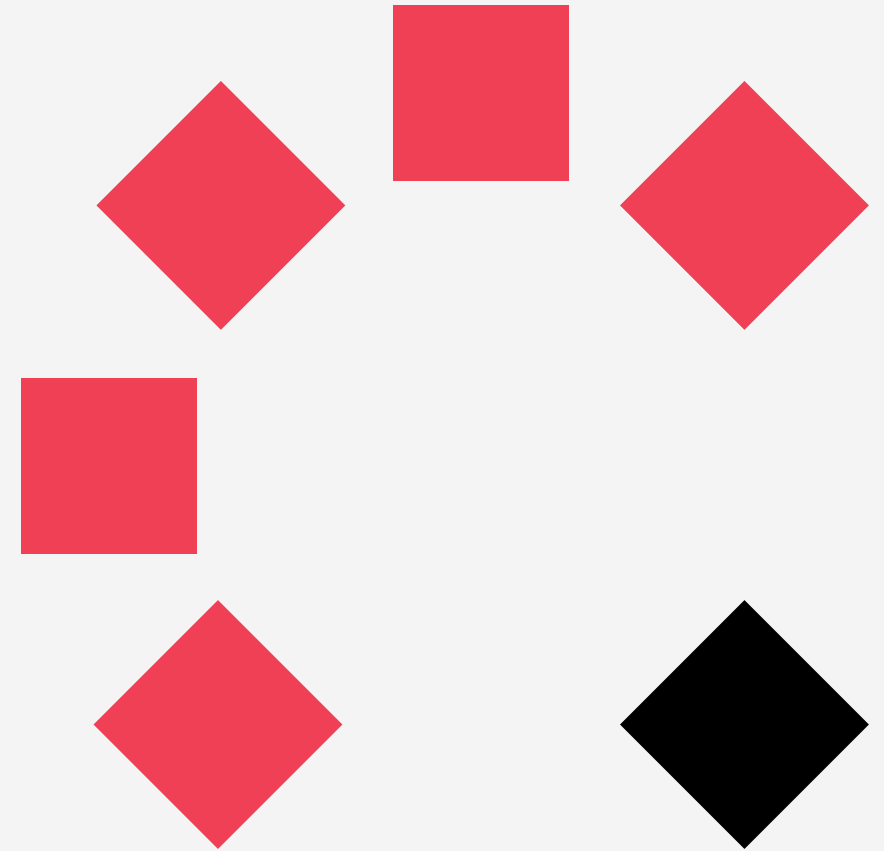




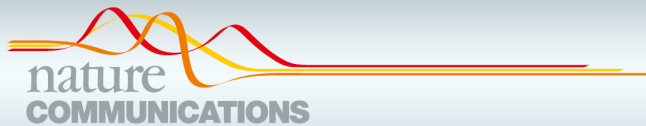
*Da ll'anonimizzazione a ll'enablement:
il dato clinico come motore
dell'innovazione sanitaria*

Key Note Speech – Aldo Cingolani

AI Legal – Seminario 1
10/02/2026



Perché l'anonimizzazione «classica» non basta



ARTICLE

<https://doi.org/10.1038/s41467-019-10933-3>

OPEN

Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher^{1,2,3}, Julien M. Hendrickx¹ & Yves-Alexandre de Montjoye^{2,3}

While rich medical, behavioral, and socio-demographic data are key to modern data-driven research, their collection and use raise legitimate privacy concerns. Anonymizing datasets through de-identification and sampling before sharing them has been the main tool used to address those concerns. We here propose a generative copula-based method that can accurately estimate the likelihood of a specific person to be correctly re-identified, even in a heavily incomplete dataset. On 210 populations, our method obtains AUC scores for predicting individual uniqueness ranging from 0.84 to 0.97, with low false-discovery rate. Using our model we find that 99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes. Our results suggest that even heavily sampled anonymized datasets are unlikely to satisfy the modern standards for anonymization set forth by GDPR and seriously challenge the technical and legal adequacy of the de-identification release-and-forget model.

Le lacune dell'anonimizzazione classica

- **Rischio di re-identificazione:**
il rischio cresce in linea con l'avanzare delle capacità computazionali e IA generativa (es. tecniche avanzate quali linkage, outlier, composizione, inferenza, ecc).
- **Assenza di garanzie formali:**
la privacy dipende dalle assunzioni sul contesto.
- **Assenza di garanzie misurabili:**
le tecniche classiche sono regole euristiche: aiutano a ridurre il rischio ma non forniscono una misura matematica del massimo leakage possibile.

Anonimizzazione matematica tramite Differential Privacy

La Differential Privacy (DP) è un insieme di tecniche accomunate dall'introduzione di rumore stocastico in modo controllato e misurato rispetto alle risposte statistiche provenienti da un set di dati, per proteggere la presenza/assenza di un individuo.



GDPR

Current techniques



99.98% of "anonymized" records can be re-identified - *Nature 2019*

Differential Privacy
agora
labs

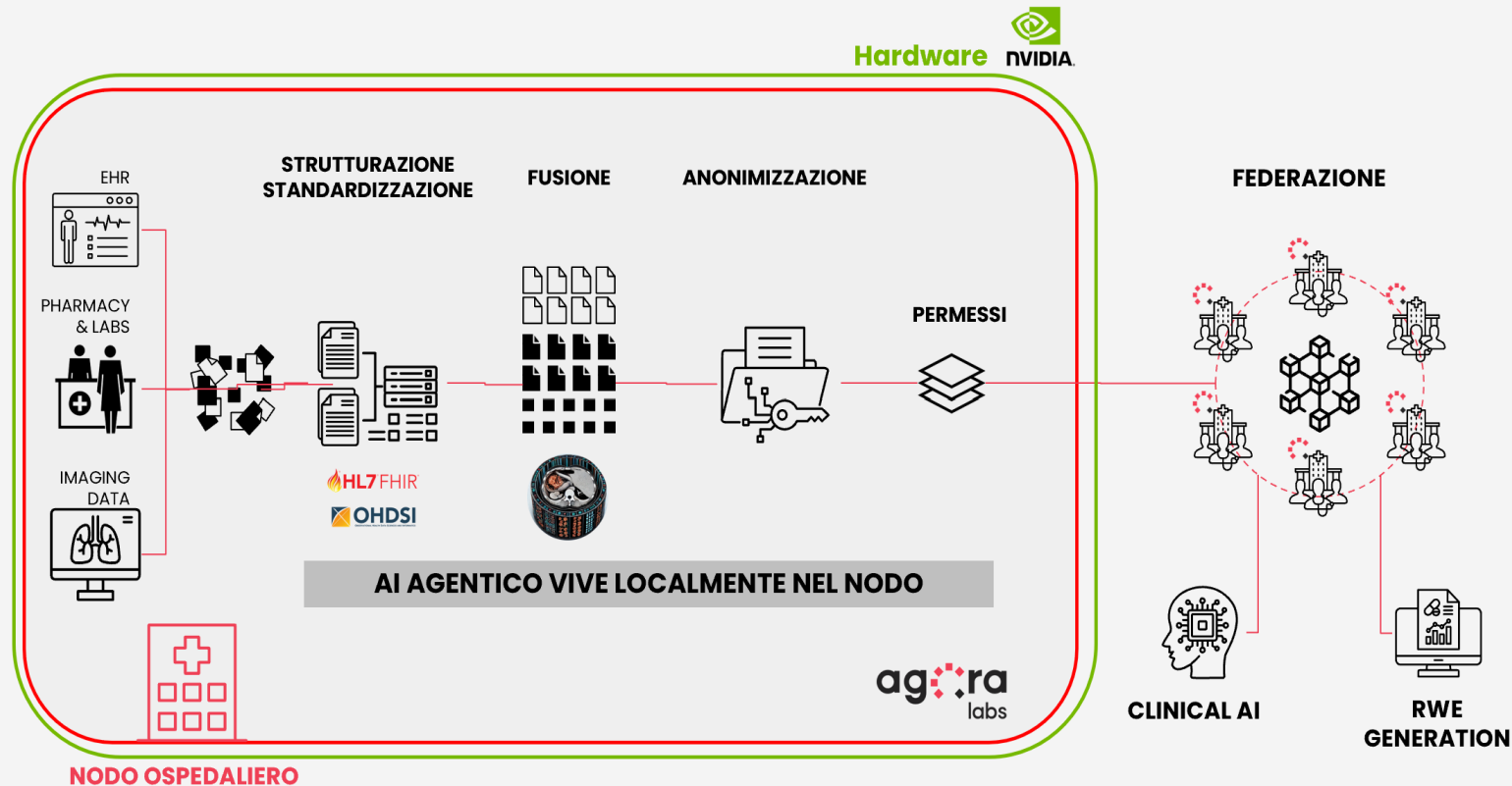


Quantifying the information actually shared in each analysis. Modify the amount each time to return only what's necessary - *Dwork 2006*



ISO 27001

Attivazione Clinica dei Dati «End To End»



1. Nodi dedicati «on-premise» presso ogni ospedale; i dati grezzi non escono mai dall'ospedale.
2. Ingestione, processamento e standardizzazione locale. Modellazione sicura e senza data lake centralizzati.
3. Output sintetici in virtù dell'applicazione della DP, per abilitare ricerca / AI fuori perimetro GDPR. La privacy è una metrica (€).
4. Query approvate dal Titolare; computazione locale; nessun trasferimento di dati ad Agora Labs o ad altri utenti.

Quadro giuridico



Prima dell'
anonimizzazione

Base giuridica

- GDPR Art. 6, comma 1 lett. e) (*compito di interesse pubblico*) e lett. f) (*legittimo interesse*).
- GDPR Art. 9, comma 2 lett. g) (*rilevante interesse pubblico*) e/o lett. j) (*ricerca scientifica in conformità dell'art. 89*), secondo diritto UE/ Statale applicabile.
- D. Lgs 196/2003 Art. 110 dati sanitari per ricerca senza consenso, con parere Comitato Etico + DPIA, per trattamento locale pre-anonimizzazione.



Dopo
l'anonimizzazione

Fuori GDPR

- Gli output sono dati anonimi; le attività successive si svolgono fuori dal perimetro del GDPR.
- Le analisi avvengono sempre sotto consenso del Data Controller.
- Output dell'anonimizzazione riutilizzabile entro simili scopi di ricerca ai sensi Art. 8 D. Lgs. 132/2018.



Dati operativi
separati

Gestione Data User

- Dati operativi separati (es. registrazione, accessi, sicurezza): utenti piattaforma trattati con finalità e basi distinte, con logging immutabile per prevenire abusi.

Il quadrante del dato clinico



SOLVING PRIVACY

- I dati restano sotto il pieno controllo dell'ospedale
- Anonimato dei pazienti irreversibile
- Approvazioni etiche una tantum Ethical approvals



SOLVING QUALITY

- Utilizzo di LLM per estrarre e standardizzare i dati
- Data curation con interfaccia conversazionale (chat-driven) pensata per utenti non esperti



SOLVING SCALABILITY

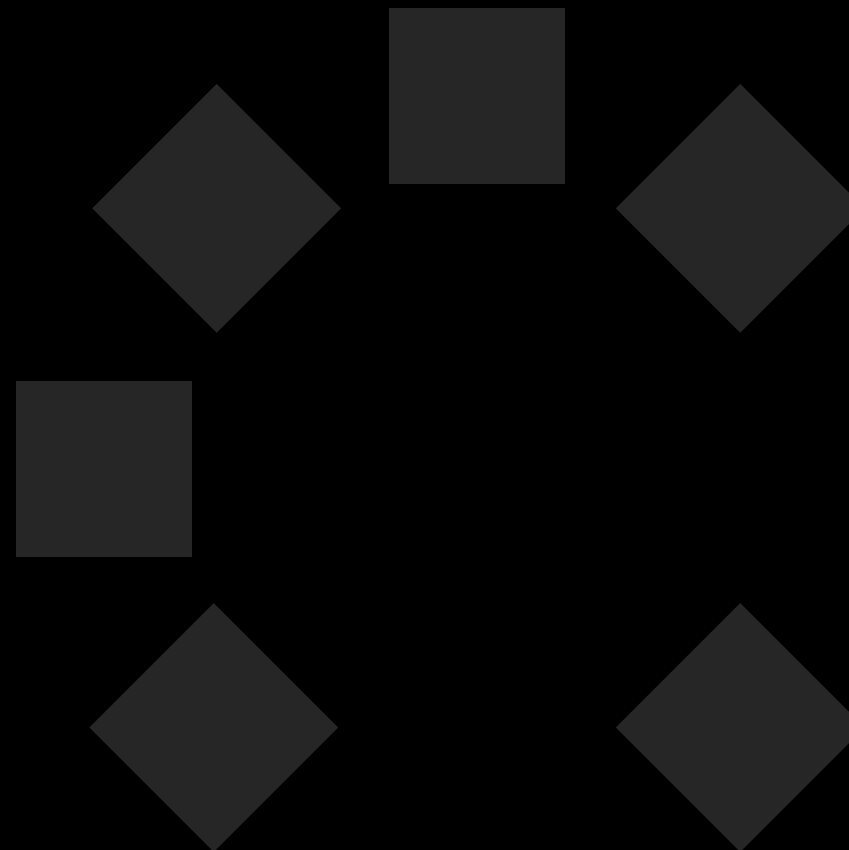
- Approvazione rapida, una sola volta
- Dati attivati in poche ore
- Hardware dedicato
- Studi in scala in una frazione del tempo e dei costi



SOLVING VALUE

- Facilitazione di partnership tra ospedali e aziende di AI e Pharma
- Incentivi, riconoscimento e monetizzazione per gli ospedali

Gra zie



Aldo Cingolani
VP of Business Operations
aldo.cingolani@agoralabs.tech