

L'IA per la raccolta e valorizzazione del dato

Gianmario Cinelli · CERGAS, SDA Bocconi

L'EDS, la ricerca e la scala del valore del dato nelle esperienze internazionali

L'EDS E IL DATO PER LA RICERCA

Uso secondario — ricerca: tre cataloghi, un denominatore comune — l'estrazione



IL DATO DISPONIBILE

- Dati clinici pseudonimizzati da tutte le strutture
- 14 tipologie di documenti (referti, PS, dimissioni, vaccinazioni...)
- Anagrafe Nazionale Assistiti (ANA) — identità unica
- Dati codificati: FHIR R4 · LOINC · AIC/ATC · ICD-10



SERVIZI DI CONSULTAZIONE

- Ricerca su dati pseudonimizzati
- Estrazione di dataset / coorti per la ricerca
- Cruscotti e statistiche di popolazione (anonimizzati)
- Monitoraggio alimentazione e utilizzo del dato



GOVERNANCE E ORIZZONTE

- Governance su dati anonimizzati
- Gestione terminologie e qualità del dato
- Verifica finalità e consensi (uso secondario)
- Integrazione EHDS / HealthData@EU (vincol. 2029)
- OMOP CDM — standard di fatto uso secondario UE

LA SCALA DEL VALORE DEL DATO — IN CHIAVE RICERCA



I servizi di ricerca dell'EDS si fermano a raccolta ed estrazione. ...possiamo accontentarci di un'infrastruttura che visualizza ed estrae, o serve un ambiente che permetta di analizzare e agire sul dato per la ricerca?

LE ESPERIENZE INTERNAZIONALI DI RICERCA

IL Clalit Research Institute

Clalit Health Services · Research Institute (2010)

OBIETTIVO

EHR di routine a scala nazionale (uso secondario): ~7M assistiti (>50% pop.), longitudinale ~30 anni. Case-finding predittivo + RWE, senza coorte costruita né genomica.

IMPATTO

RWE vaccino COVID (NEJM 2021, ~1,2M) · HCV: resa 1:1.000→1:10 (NEJM AI 2024) · C-Pi: >1.000 MMG · 250k+ care gap/mese

CN China Kadoorie Biobank

Oxford CTSU + Chinese Academy of Med. Sciences · 2004

OBIETTIVO

Coorte 512k adulti cinesi (5 rurali + 5 urbane), 20 anni di follow-up; validazione cross-ancestry delle scoperte genetiche.

IMPATTO

512.891 partecipanti · WGS completo dell'intera coorte (Q1 2025) · >600 pubblicazioni · landmark su rischio CV e dieta

FI FinnGen

Consorzio pubblico-privato finlandese · biobanche + industria

OBIETTIVO

Genomica di popolazione su isolato genetico finlandese, linkata ai registri nazionali; potenza statistica su varianti rare.

IMPATTO

500.348 partecipanti = ~10% pop. FI · 21,3M varianti · 2.447 endpoint · ~2.500 loci su 1.900+ patologie

La ricerca e la scala del valore del dato: le sperimentazioni del SSN e le strade di change management

SPERIMENTAZIONI SSN PER LA VALORIZZAZIONE DEL DATO NELLA RICERCA

Cinque capacità che fanno salire il dato della ricerca lungo la scala del valore



TRASFORMARE LA RICERCA CON IL DATO E L'AI: TRE STRADE DI CHANGE MANAGEMENT



RETE IRCCS

rete di aziende · ambito verticale

3-5 IRCCS in rete (es. oncologici) · biobanca/registro federato



SCENARIO

Coorte di ricerca federata multi-IRCCS: l'AI compila i registri dai referti; coorti costruite e analizzate in locale, il dato non lascia l'azienda.



USE CASE

Registro oncologico multicentrico automatizzato → RWE + reclutamento trial. Il modello Clalit nel perimetro IRCCS.



REGIONE

federazione di aziende · scala regionale

Una Regione pilota EDS (Lombardia, Veneto, ER, Lazio...)



SCENARIO

Piattaforma regionale di ricerca sull'EDS: uso secondario dei dati di routine di milioni di assistiti — analitica di popolazione e case-finding.



USE CASE

Case-finding di cronicità non diagnosticate su scala regionale (modello Clalit) · epidemiologia real-world.



AGENAS / MINISTERO

modello nazionale · asset pubblico

AGENAS / Ministero · nodo nazionale di ricerca



SCENARIO

Nodo nazionale allineato OMOP/EHDS: l'Italia espone il dato per l'uso secondario e partecipa agli studi federati europei (EHDEN — già 1° contributore).



USE CASE

Studi federati multinazionali su milioni di pazienti (modello OHDSI/LEGEND) · EHDS uso secondario (vincolante 2029).

Quale livello — rete IRCCS, Regione, Nazione — guiderà la trasformazione della ricerca attraverso il dato e l'AI?

