ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

Condivisione e riuso dei dati nella ricerca clinica e sanitaria

Rita Banzi

Istituto di Ricerche Farmacologiche Mario Negri IRCCS

18 Novembre 2025

GenOA week 2025 Who Owns Our Knowledge?

Dichiarazione conflitti di interesse

- Nessun conflitto di interesse economico-finanziario rispetto ai contenuti della presentazione
  - *Progetto OSIRIS Open Science to Increase Reproducibility in Science*
  - *Progetto eCREAM enabling Clinical Research in Emergency and Acute care Medicine*
  - *Progetto EATRIS CONNECT digital transformation to accelerate translational medicine*

- Moltissimi di tipo intellettuale
  - *Penso che la scienza sia un bene comune e che la ricerca scientifica che informa le decisioni sanitarie debba essere trasparente e aperta in tutte le sue fasi, dall'ideazione alla disseminazione dei risultati*

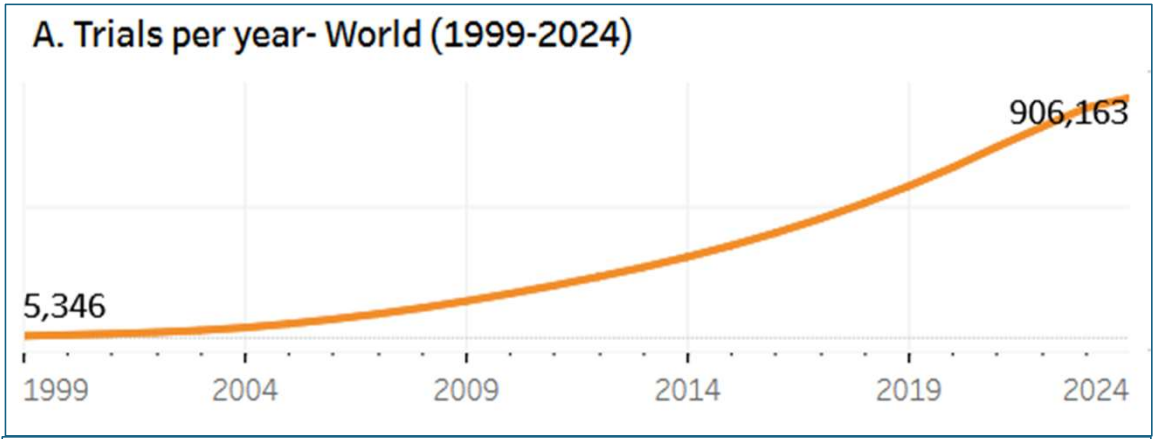
Outline

1. La condivisione dei dati (individuali) degli studi clinici

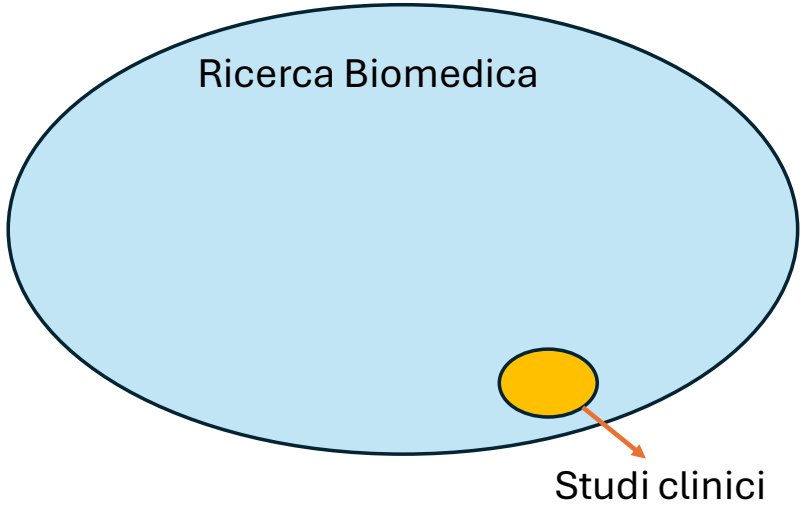
2. Riuso dei dati sanitari per scopi di ricerca

Perché parlare di condivisione dati negli studi clinici

Gli studi clinici sono «esperimenti» medici che coinvolgono persone e hanno lo scopo di stabilire se un intervento sanitario (farmaco, dispositivo, procedura diagnostica, tecnica chirurgica, ecc.) funziona ed è migliore rispetto ad alternative già disponibili

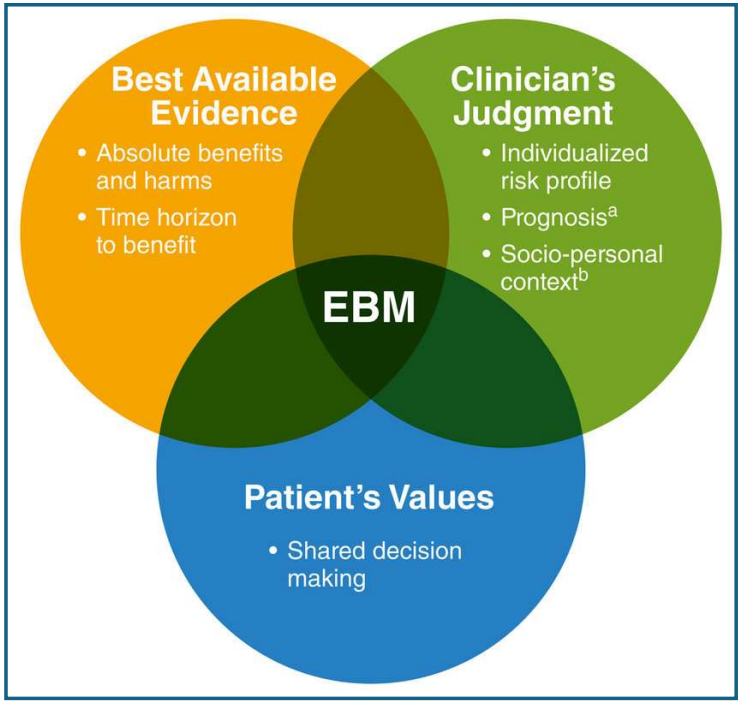


www.who.int/observatories/global-observatory-on-health-research-and-development/monitoring/number-of-trial-registrations-by-year-location-disease-and-phase-of-development



Perché parlare di condivisione dati negli studi clinici

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS



EBM: Evidence-based medicine

Più in generale «evidence-informed decision making»

Il contesto: tre livelli di informazione

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

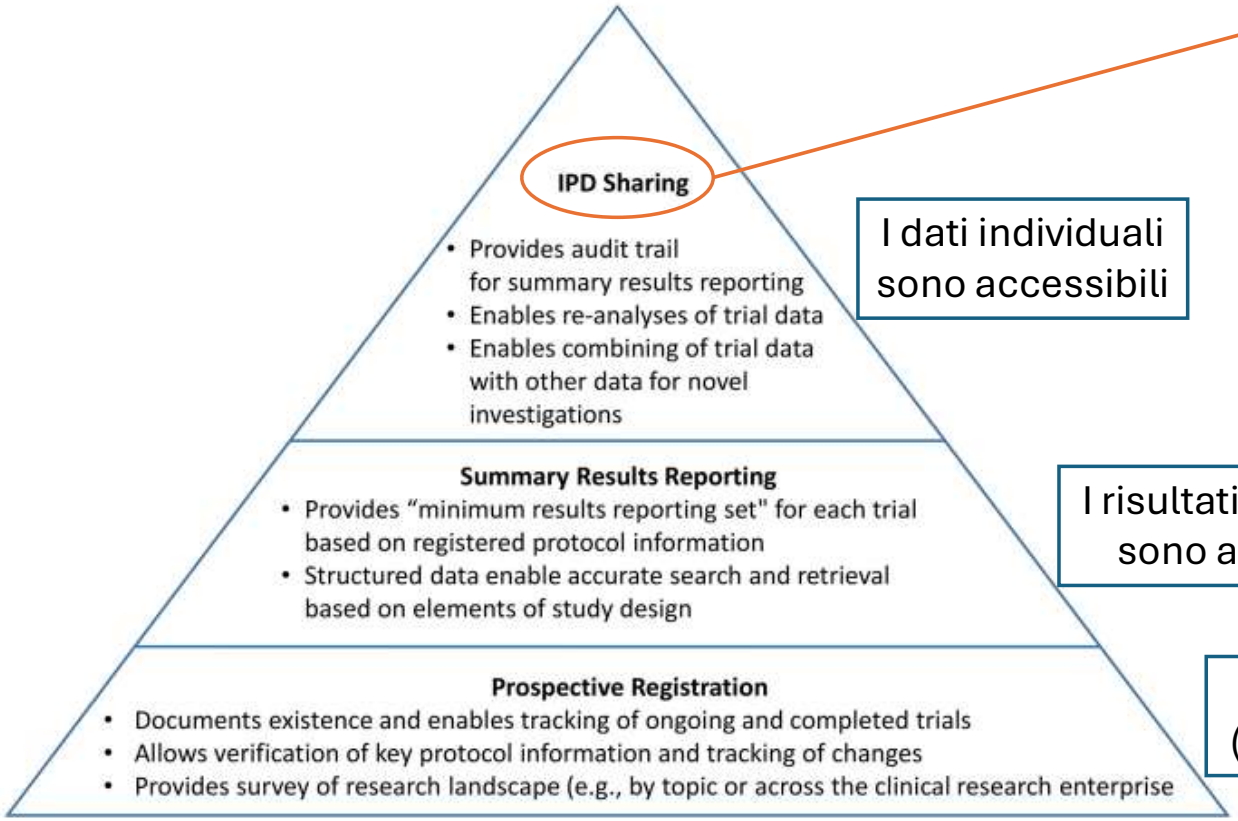
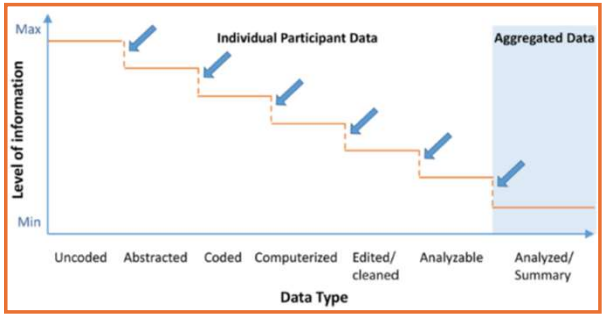


Fig 2. Schematic depicting the functions of the three key components of the TRS.

Set di dati relativi ai singoli partecipanti a uno studio clinico (dati su variabili socio-demografiche e cliniche, ecc.) alla base dei risultati aggregati



I dati individuali sono accessibili

I risultati (aggregati) sono accessibili

Lo studio esiste (info base sul protocollo)

Zarin et al PloS Med 2016
Zarin Plos Med 2013

Condivisione e riuso dati individuali degli studi clinici

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

Vantaggi e opportunità

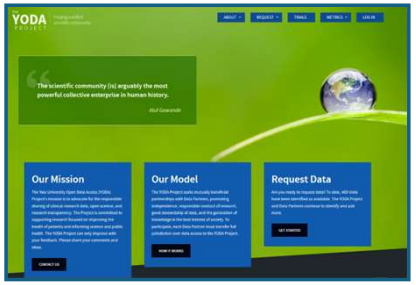
- Trasparenza e fiducia nella ricerca
- Analisi indipendenti dei risultati originali
- Analisi dei data set per esplorare ipotesi secondarie o alternative
- Supporto alla definizione di nuove ipotesi di ricerca
- Sviluppo e validazione di approcci statistici
- Contestualizzazione dei risultati (meta-analisi di dati individuali)
- Stimolo alla collaborazione
- Maggiore garanzie di qualità dei dati
- Riduzione reporting bias

Ostacoli e preoccupazioni

- Di chi sono i dati?
- Ri-uso improprio
- Interpretazione scorretta dei dati ("data-dredging“, analisi multiple)
- Preparazione dataset, archiviazione e mantenimento
- Protezione proprietà intellettuale
- Divulgazione informazioni commercialmente sensibili
- Rispetto privacy e rischio di re-identificazione
- Limiti del consenso informato

https://freebird.lshtm.ac.uk/home/

L'approccio delle piattaforme (e dell'intermediario «neutrale»)...

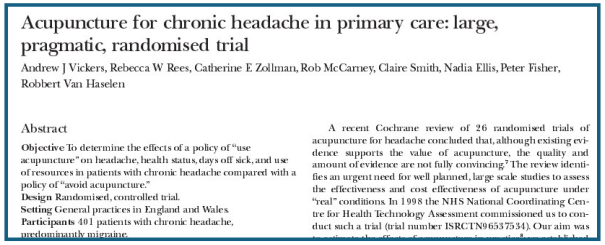


https://yoda.yale.edu/

https://vivli.org/

L'approccio collaborativo...

L'approccio «naive»...



Acupuncture for chronic headache in primary care: large, pragmatic, randomised trial

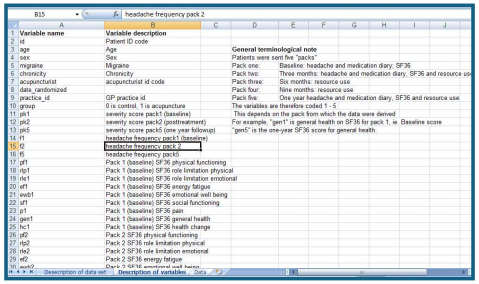
Andrew J Vickers, Rebecca W Rees, Catherine E Zollman, Rob McCarney, Claire Smith, Nadia Ellis, Peter Fisher, Robbert Van Haselen

Abstract

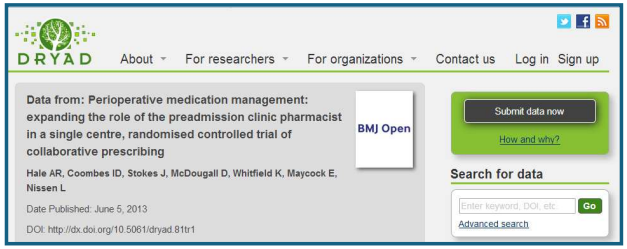
Objective To determine the effects of a policy of "use acupuncture" on headache, health status, days off sick, and use of resources in patients with chronic headache compared with a policy of "avoid acupuncture."
Design Randomised, controlled trial.
Setting General practices in England and Wales.
Participants 401 patients with chronic headache, predominantly migraine.

A recent Cochrane review of 26 randomised trials of acupuncture for headache concluded that, although existing evidence supports the value of acupuncture, the quality and amount of evidence are not fully convincing.[1] The review identifies an urgent need for well planned, large scale studies to assess the effectiveness and cost effectiveness of acupuncture under "real" conditions. In 1998 the NHS National Coordinating Centre for Health Technology Assessment commissioned us to conduct such a trial (trial number ISRCTN96537534). Our aim was

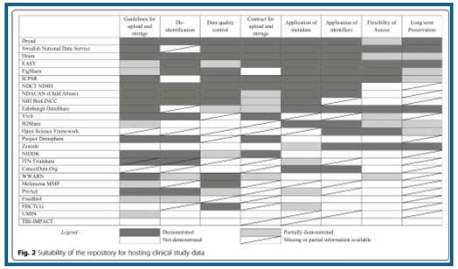
Vickers BMJ 2004



L'approccio archivistico...

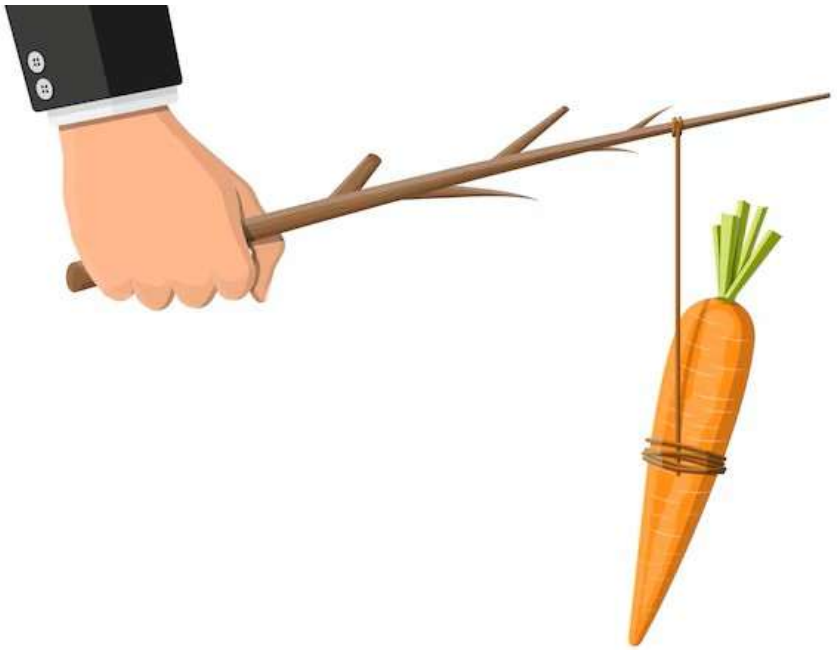


Hale BMJ Open 2013



Banzi Trials 2019

Clinical trial data sharing ecosystem



Finanziatori della ricerca

Core funders of medical research commit to strengthening clinical trials worldwide

World Health Organization

Joint statement on strengthening clinical trials 2025

25 September 2025

III. Clinical trials that are supported to be conducted to meet best-practice standards, by including the following elements in trial policies or conditions of funding:

- clinical trials are conducted in line with the WHO Guidance for Best Practices for Clinical Trials;
- proportionate engagement of communities, to be conducted throughout the trial lifecycle, as an essential component of ethical trials;
- registration and update on trial progress in a publicly available, free to access, searchable, clinical trial registry complying with WHO's international agreed standards and updating of registries to include trial results, working towards a timeframe of 12 months from primary trial completion (in line with the WHO Joint Statement on the Public Disclosure of Results);
- encourage the use of standardized data protocols where available and Core Outcome Sets;
- open-access publication of clinical trial materials (such as trial protocols and statistical analysis plans) at the earliest opportunity, and preferably through trial registries;
- timely publication of results (working towards a timeframe of 12 months from primary study completion) including reporting of outcome and adverse event data disaggregated by sex/gender and age, preferably in an open access peer-reviewed journal, with a trial registration ID and data availability statement detailing how the data underlying the publication can be accessed;
- encourage researchers during a public health emergency to rapidly and responsibly share interpretable results, including negative results, with relevant authorities for clinical guideline development and emergency use listing; and
- encourage sharing of de-identified data (or meta-data where required), complying to international data standards, in a suitable repository with a persistent identifier.

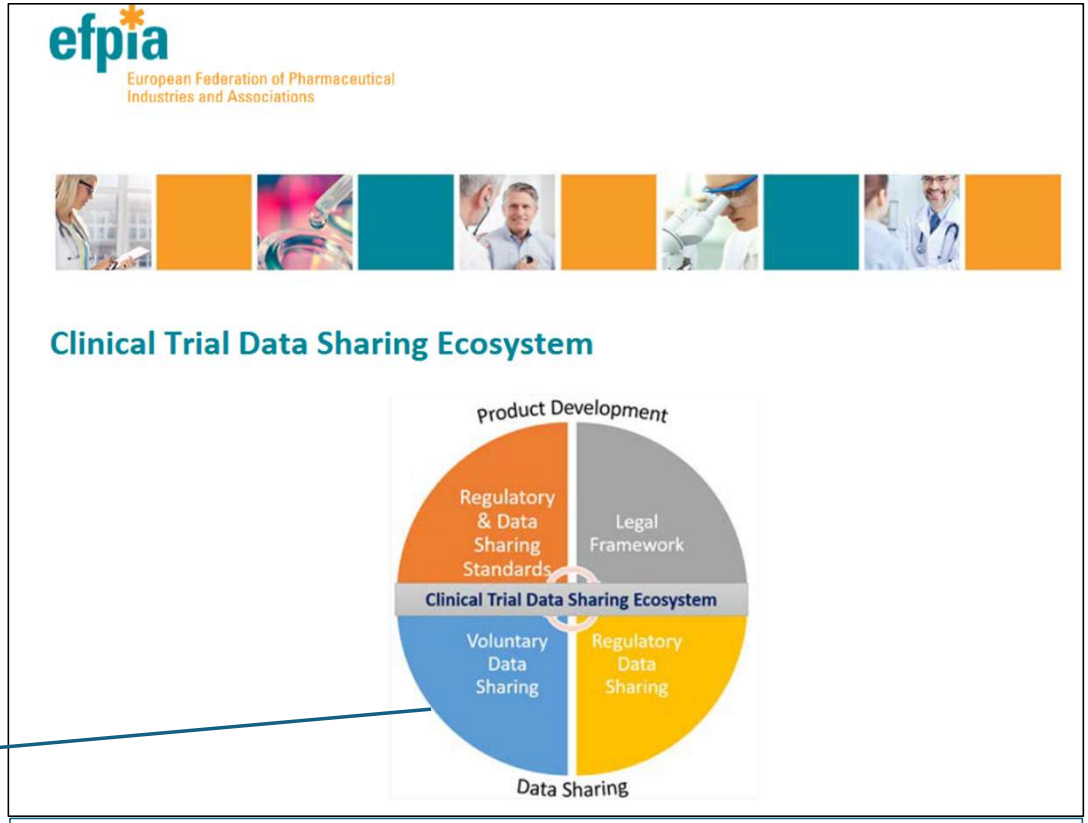
Industrie

- Molte industrie farmaceutiche hanno policy per la condivisione di dati individuali da studi clinici

- Alcune analisi hanno evidenziato che circa la metà dei dataset di studi che supportano la commercializzazione di nuovi farmaci non sono accessibili
  (definizione di fine studio, embargo, valutazione da parte di EMA e FDA, problemi con il consenso, ecc.)

Modi et al. BMC Medicine 2023

Voluntary!



https://www.efpia.eu/media/636913/clinical-trial-data-sharing-ecosystem.pdf

Editori medico scientifici

ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

2. Data Sharing

The ICMJE's data sharing statement policy is detailed in an editorial (see Updates and Editorials).

1. As of 1 July 2018 manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data sharing statement as described below.
2. Clinical trials that begin enrolling participants on or after 1 January 2019 must include a data sharing plan in the trial's registration. The ICMJE's policy regarding trial registration is explained above. If the data sharing plan changes after registration this should be reflected in the statement submitted and published with the manuscript, and updated in the registry record.

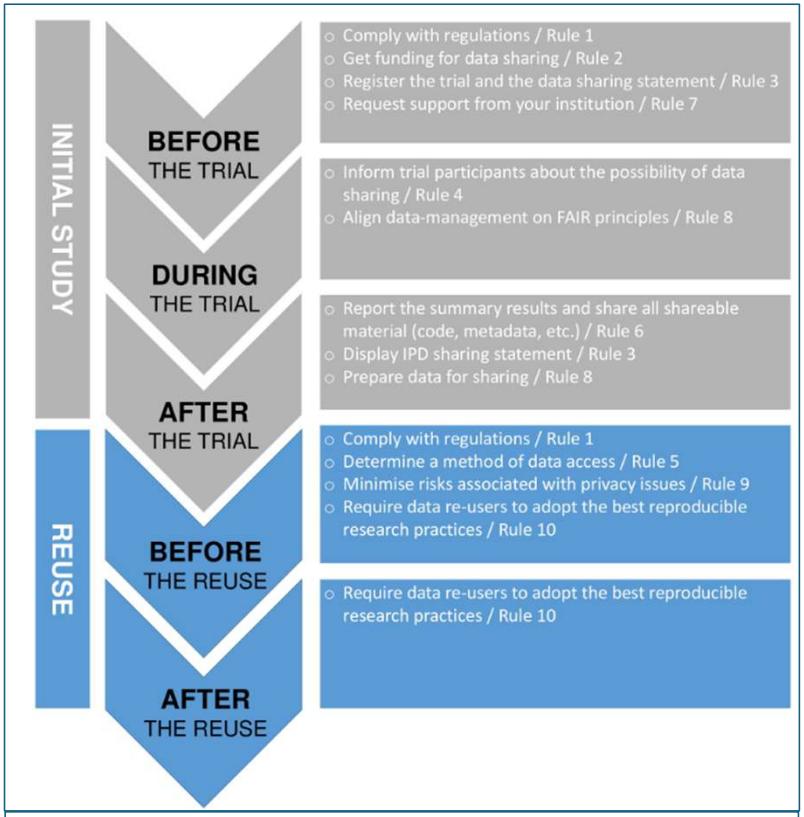
Data sharing statements must indicate the following: whether individual deidentified participant data (including data dictionaries) will be shared ("undecided" is not an acceptable answer); what data in particular will be shared; whether additional, related documents will be available (e.g., study protocol, statistical analysis plan, etc.); when the data will become available and for how long; by what access criteria data will be shared (including with whom, for what types of analyses, and by what mechanism). Illustrative examples of data sharing statements that would meet these requirements are provided in the Table.

www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html

Taichman Annals Int Med 2017

Danchev Jama Network Open 2021

Ma come si fa?



Pellen et al. PLoS Comput Biol 2023

BOX 1

Opportunities for clinical trial data-sharing experts

1. Each clinical trial data-sharing platform needs experts for a complex, thorough and efficient review process. Experts will have their own specific field, such as medicine, biostatistics, interdisciplinary data science, intellectual property, privacy protection, and ethical, legal and social issues, but also need an overarching understanding of the disciplinary nature of this topic.
2. The peer review of research papers that come from clinical trial data sharing should reflect the interdisciplinary approach and complexity. Editors or editorial staff should boost clinical trial data sharing within their journals' activities. Requiring data and analysis scripts for reviews will initiate basic processes to share clinical trial data. Reproducibility editors can steer the verification of published results based on the data used.
3. Academic institutions, such as trial centers at academic hospitals, need well-trained staff, including data managers and data engineers. These staff can set up and contribute to clinical trial data-sharing activities, especially for anonymization requirements.
4. Research funders, as well as research management, in academia, industry, governments and regulatory bodies must be trained in the complex issues of clinical trial data sharing.
5. Future medical researchers, as well as data scientists in medicine, will need basic training on clinical trial data sharing so that they can conduct their own research using shared datasets.
6. Training in meta-research for academics, professional institutions and funders will maximize the impact of data sharing.

Mansmann Nature Medicine 2023

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

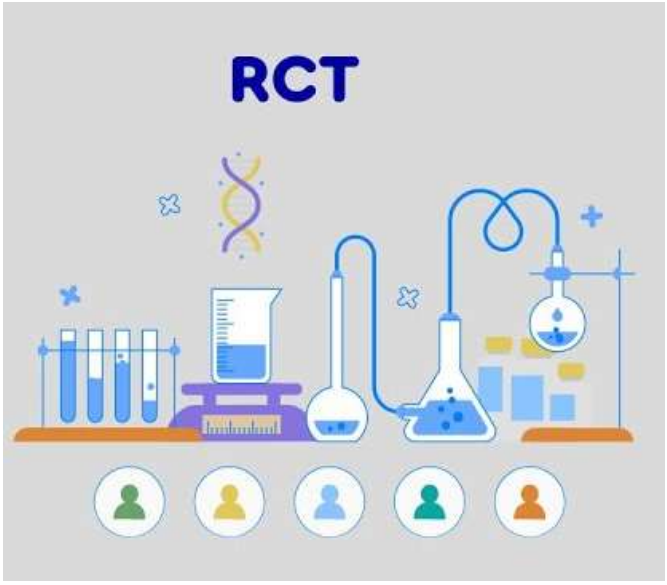
Outline

1. La condivisione dei dati (individuali) degli studi clinici

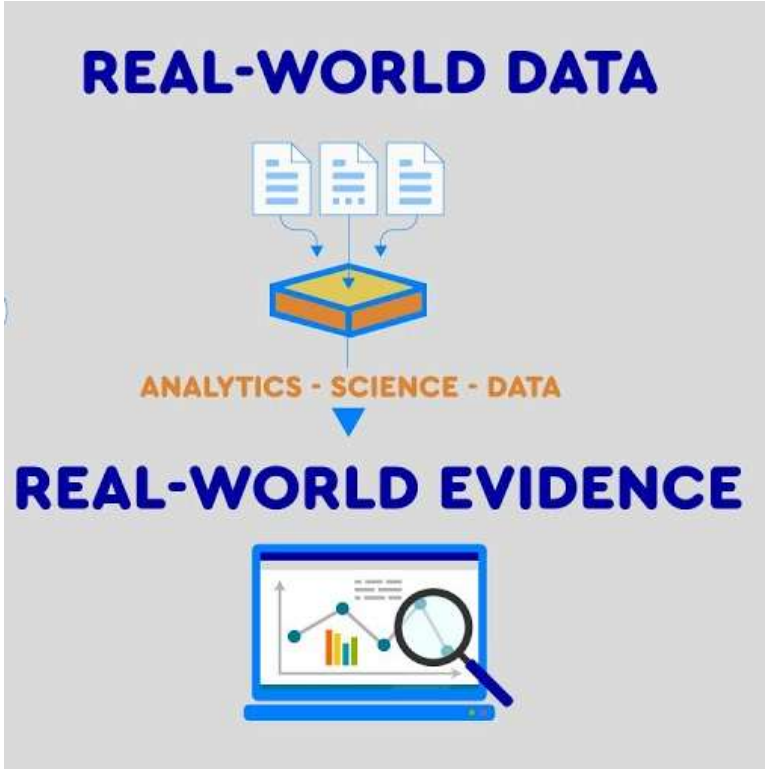
2. Riuso dei dati sanitari per scopi di ricerca

Ricerca complementare agli studi clinici

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS



insufficienti per guidare il processo decisionale poiché sono intrinsecamente incapaci di valutare l'impatto dei trattamenti nella «reale pratica clinica»



crescente interesse per lo sviluppo di metodi in grado di generare prove affidabili sull'impatto dei percorsi di cura usando dati raccolti nel «mondo reale»

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

10 novembre 2025

Ricerca in ambiti particolarmente difficili

il caso dell'emergenza-urgenza

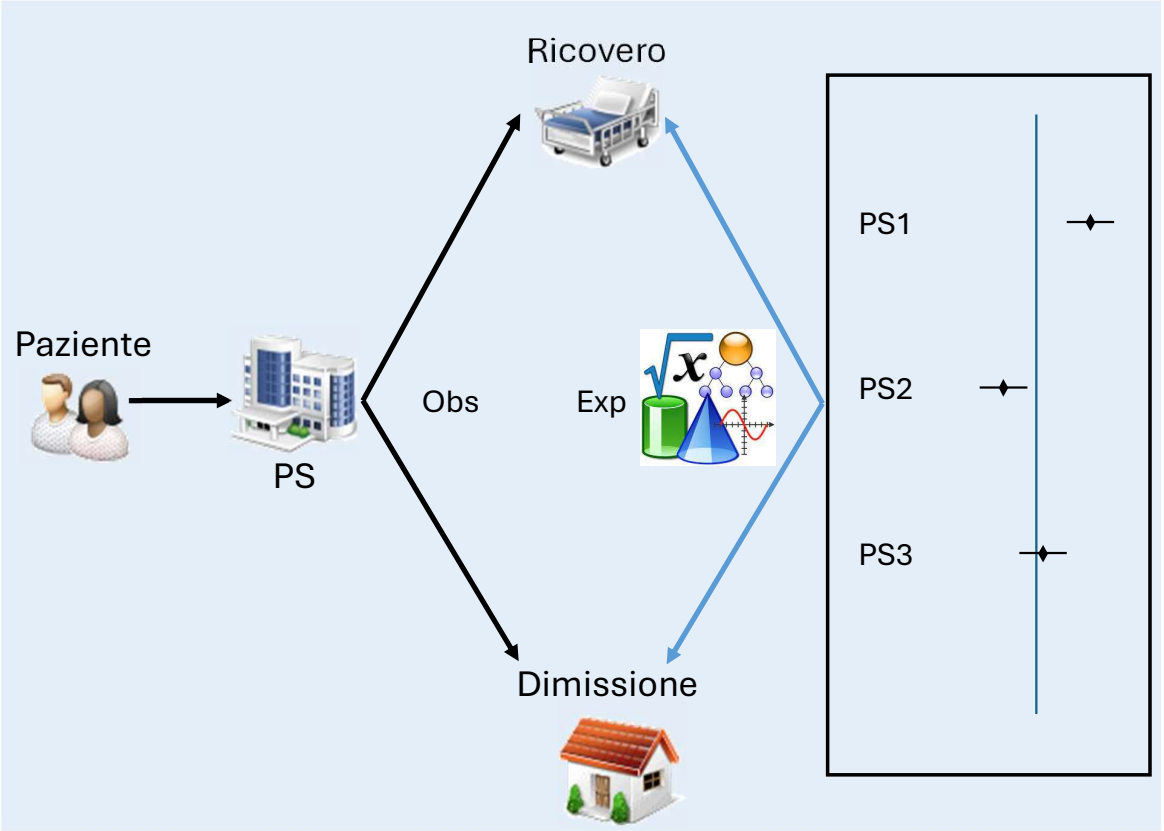
ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

Studio osservazionale in pronto soccorso per raccogliere variabili cliniche o organizzative che determinano la propensione al ricovero

- → tempo medio raccolta e gestione consenso **15 minuti**
- → tempo medio compilazione scheda raccolta dati **20 minuti**

- → tempo medio che i medici di PS dedicano a ciascun paziente è di **42 minuti** (attività clinica + amministrativa)

Fare questo studio significherebbe raddoppiare il tempo dedicato a ciascun paziente



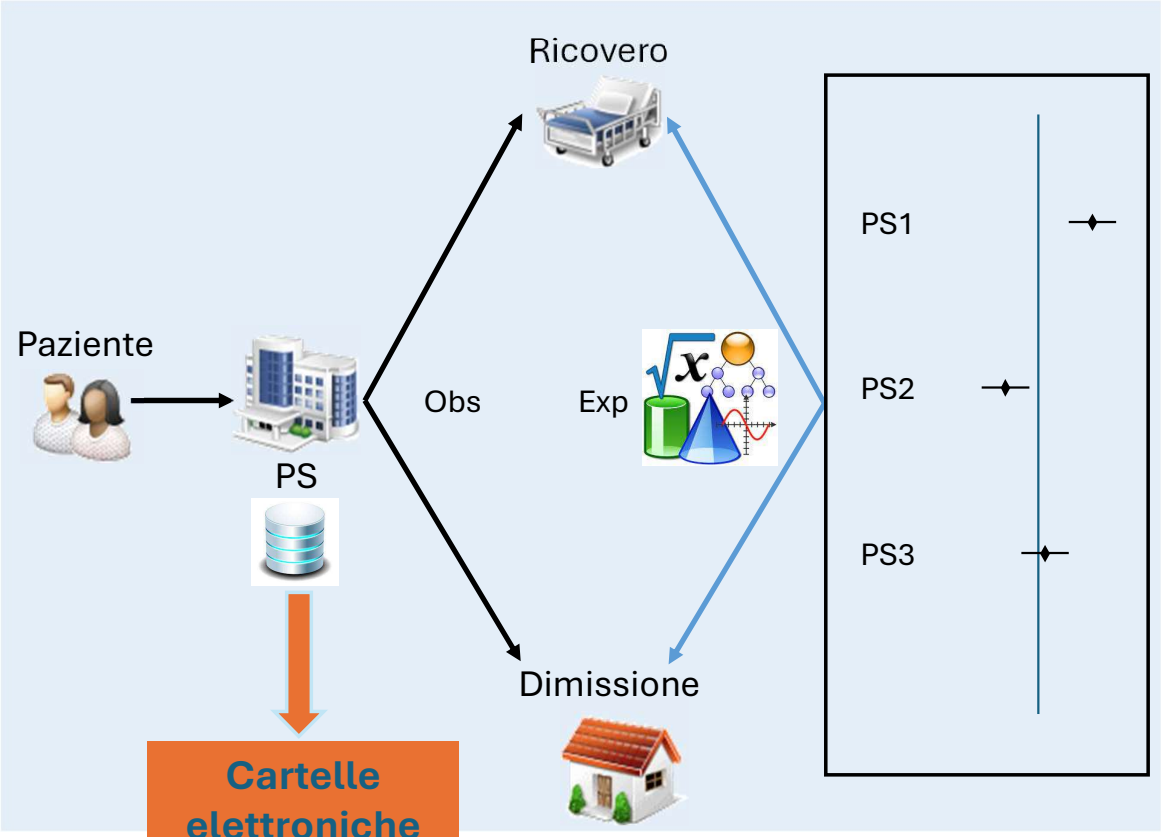
eCREAM enabling Clinical Research in Emergency and Acute care Medicine

Ricerca in ambiti particolarmente difficili
il caso dell'emergenza-urgenza

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

I dati che ci servono li abbiamo già!

Dobbiamo **«solo»** cercare un modo per
- Trovarli
- Capirli
- Utilizzarli nel rispetto delle normative
- Integrarli
- Renderli disponibili per la ricerca (nostra e di altri)



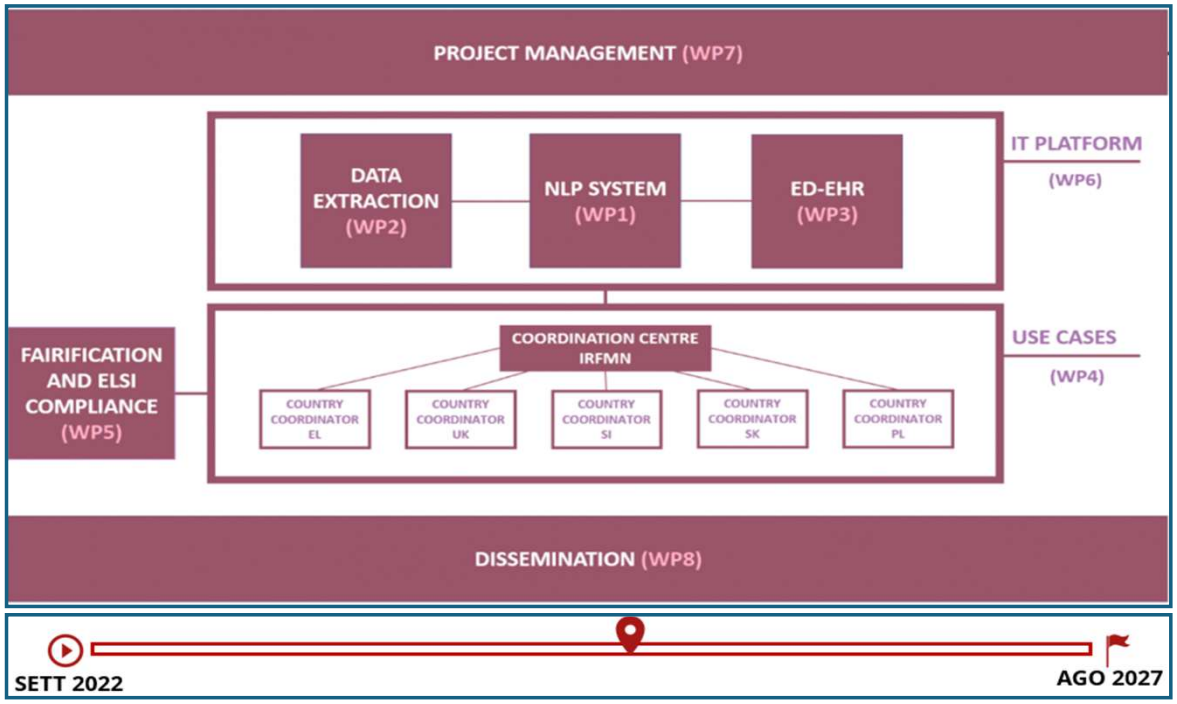
eCREAM enabling Clinical Research in Emergency and Acute care Medicine

Progetto eCREAM

enabling Clinical Research in Emergency and Acute care Medicine

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

- Sviluppare nuove soluzioni tecniche per estrarre informazioni cliniche affidabili da dati strutturati e non strutturati contenuti in diverse cartelle cliniche elettroniche dei pazienti
- Testare l'utilizzo di questi dataset su case studies
- Rendere FAIR le banche dati create nel rispetto delle legislazioni europee e nazionali



https://ecreamproject.eu/

eCREAM project
N. 1010557726

Funded by
the European Union

eCREAM enabling Clinical Research in Emergency and Acute care Medicine

Le sfide etico-legali per l'utilizzo di questi dati

- Classificazione degli studi – metodologica vs legale/regolatoria
- Base legale per il trattamento dati (GDPR + regolamenti nazionali)
- Autorizzazioni regolatorie, comitati etici e autorità per la protezione dati
- Anonimizzazione dati
- Trasferimento dati tra paesi

Necessità di approfondire gli aspetti etici, legali, sociali e trovare soluzioni che preservino l'utilità del dato e siano rispettose delle normative

Matei et al. Frontiers in Disaster and Emergency Medicine Submitted

Un esempio: eCREAM NLP-DeVal

Sviluppo e validazione di un modello linguistico di grandi dimensioni per estrarre le informazioni dai testi liberi delle cartelle cliniche

- Estrazione di milioni di testi (non annotati), tra i 90000 e 300000 pazienti per PS
- Nessuna estrazione di altri dati identificativi di pazienti e PS, dati clinici
- Disaccoppiamento dei testi relativi allo stesso paziente (anamnesi, valutazione clinica, esami…)

Piccola parte di testi annotati

https://zenodo.org/records/10996542

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS

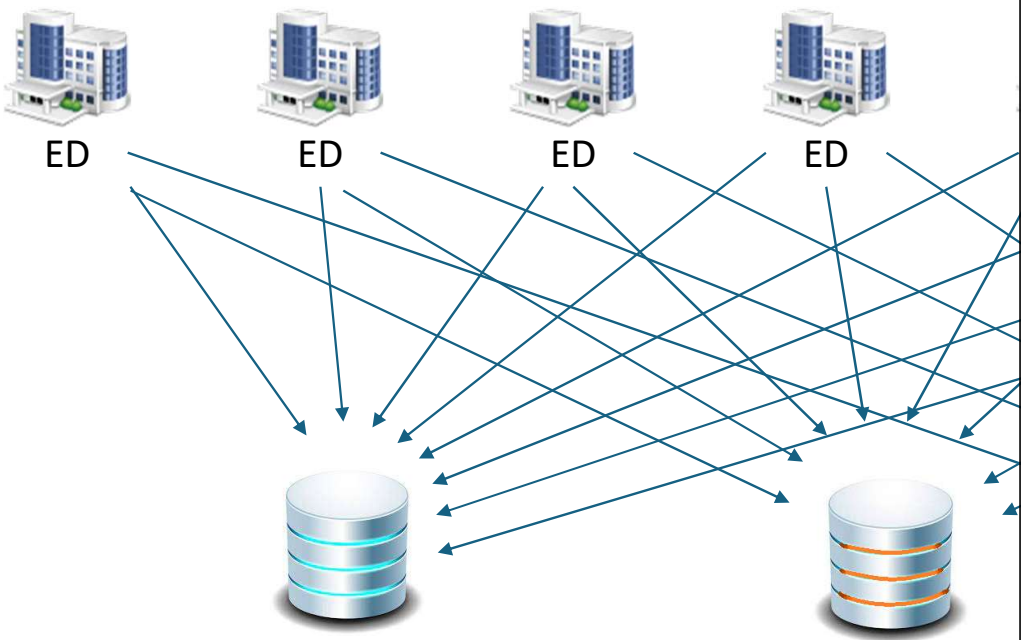
Anamnesis

Patient comes to the ER for dyspnea and oliguria. No fever. Patient with known advanced pulmonary neoplasia without further indication for active therapy. Pt asks to inform his **sister**, **Anna Lemena**, at **334556778**;- permanent AF;- arthrosis; HT lasix, metformin, prednisone, lanoxin, arcoxia. Allergy to amoxicillin and FANS. Not vaccinated antiSarsCov2, two prior infections

eCREAM enabling Clinical Research in Emergency and Acute care Medicine

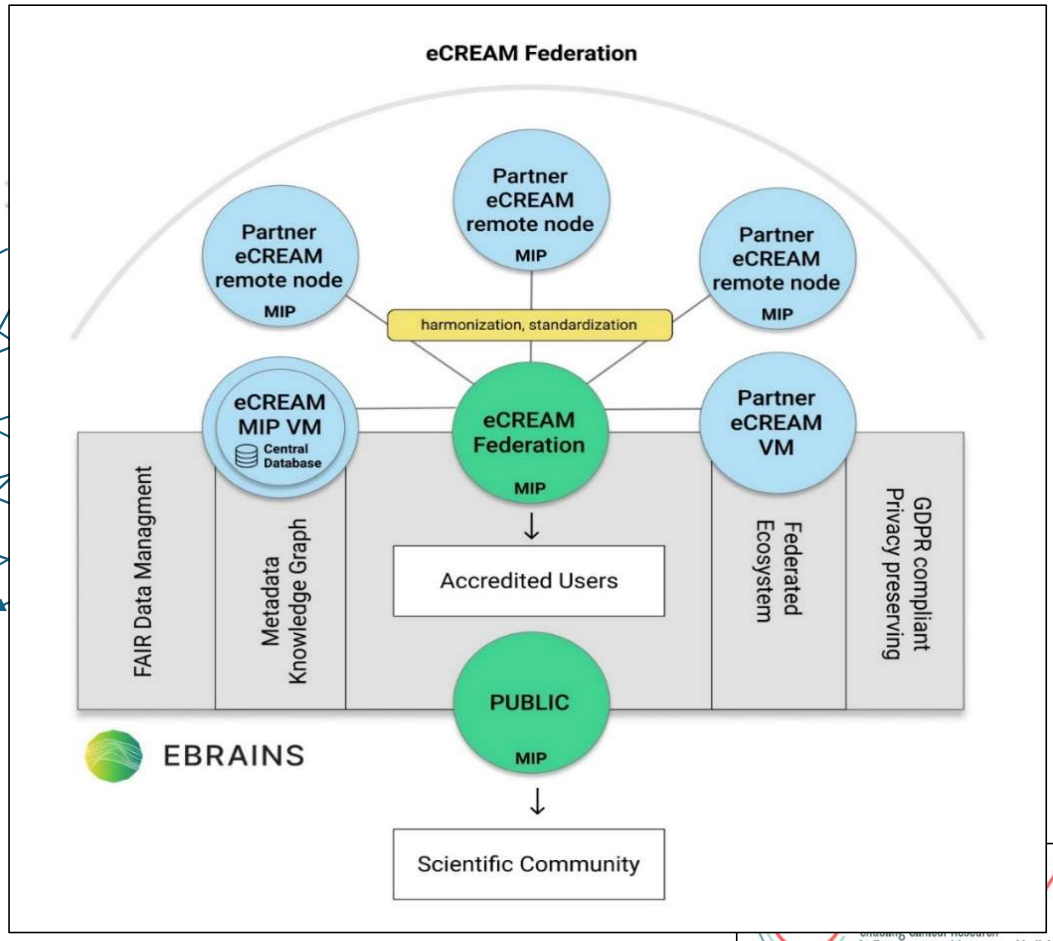
Pianificazione di una strategia per la FAIRificazione

ISTITUTO DI RICERCHE FARMACOLOGICHE MARIO NEGRI · IRCCS



Annotated and non-annotated clinical notes

Patients' characteristics



https://mip.ebrains.eu/

in Emergency and Acute care Medicine

Due semplici considerazioni conclusive

- Condivisione e riuso dei dati degli studi clinici: finalmente abbiamo smesso di discutere se farlo, parliamo di come farlo in modo rispettoso e sicuro

(anche se la strada perché diventi lo standard è ancora lunga)

- Riuso dei dati sanitari per scopi di ricerca: «what a time to be alive!»

(prossima implementazione del regolamento sullo spazio europeo dei dati sanitari)

ISTITUTO DI RICERCHE
FARMACOLOGICHE
MARIO NEGRI · IRCCS

Grazie!

rita.banzi@marionegri.it