



Fascicolo Sanitario Elettronico e Medicina Digitale

Repository - Data Lake - Data Warehouse



www.pexels.com



Premessa

Oggi tutti noi viviamo circondati dai dati. Oggi più di sempre. I siti web, i social media, le app etc sono tutti strumenti che ci consentono di generare e gestire dati. Si parla di big data cioè di grandi masse di dati o di megadati dalla fine degli anni '90, con una crescita esponenziale delle loro dimensioni. Basti pensare che abbiamo raggiunto gli zettabyte (equivalente a 1.000×10^7 pari a 10^{21}) di dati, cioè 180 milioni di volte i dati della biblioteca nazionale USA. Tra i diversi strumenti in grado di raccogliere e generare dati si colloca il Fascicolo Sanitario Elettronico, sistema informatizzato che raccoglie, organizza e archivia in formato digitale i dati relativi alla salute della persona o del paziente, consentendo la condivisione sicura e rapida delle informazioni. Quali possono essere gli sviluppi e le potenzialità del FSE? Diversi studi hanno preso in esame le potenzialità e da quanto emerge in letteratura sembra che oltre alla riduzione dei costi della ricerca clinica il FSE possa avere ricadute positive anche nell'ambito della ricerca clinica. Sappiamo però bene che fino a oggi il FSE ha trovato sul suo percorso molteplici criticità che occorre affrontare e superare. Per arrivare a capire quale potrebbe essere la strategia migliore per un uso vincente del FSE occorre fare alcune riflessioni sulle caratteristiche dei dati, elementi alla base del FSE, e sulle strategie per una loro gestione adeguata.

Big Data - caratteristiche

Secondo il McKinsey Global Institute si definisce sistema di Big Data “un dataset la cui

dimensione è talmente grande che eccede la capacità dei sistemi di database relazionali di catturare, immagazzinare, gestire ed analizzare i dati”. Perché i big data diventino davvero lo strumento di svolta per la medicina del futuro occorre razionalizzare le modalità di raccolta delle informazioni e far sì che i dati siano raccolti in modo rigorosamente omogeneo, armonico e integrato, in una logica di interoperabilità. Rispetto al passato si stanno facendo notevoli passi in avanti, ora però occorre procedere in questa direzione senza fermarsi così da creare nuove strategie per trarre dai dati il massimo della conoscenza. Sappiamo bene infatti che i dati da soli sono spesso privi di significato.

In quest'ottica l'archiviazione e la classificazione diventano fasi operative imprescindibili. Quando si parla di dati e di classificazione non si può prescindere dal concetto di metadati cioè di dati che contengono una descrizione informativa e pertinente su altri dati. Sebbene oggi si parli sempre più spesso di metadati in realtà i metadati esistevano anche in passato e tradizionalmente erano riferiti alle metodologie impiegate per catalogare i libri in biblioteca. In epoca digitale questo concetto non solo si è ampliato ma a questo si sono aggiunti altri due concetti strettamente connessi: Data Lake e Data Warehouse. Il termine Data Lake viene usato per definire un archivio di dati di grandi dimensioni nel quale i dati sono conservati in formato grezzo e non strutturato. Un Data Warehouse (DW) invece è un repository di dati che archivia

dati strutturati, filtrati ed elaborati che sono stati raccolti e trattati per uno scopo specifico. Perché parliamo di Data Lake in relazione al FSE? Perché l'uso del Data Lake può diventare uno strumento importante per l'elaborazione dei dati del FSE. La maggior parte dei dati del FSE sono infatti in formato non strutturato o semi strutturato proprio per la natura e le caratteristiche del FSE. Con i Data Lake si possono memorizzare i dati in formato grezzo mantenendo tutti i dettagli originali dei dati, inclusi quelli non strutturati e semi-strutturati, senza doverli trasformare in uno schema rigido.

Granularità del dato

A questo però va aggiunta un'altra caratteristica che non si può trascurare quando si parla di Data Lake e in generale di repository dei dati: la granularità dei dati archiviati ossia la possibilità di stabilire vari livelli di dettaglio (o di sintesi) dei dati raccolti. Facendo riferimento al FSE la granularità dei dati è un elemento particolarmente interessante anche se non è semplice definire a priori quale sia la migliore granularità per gli obiettivi che si desidera raggiungere. Sicuramente quando si fa riferimento al FSE la granularità deve essere tale da consentire di prendere decisioni appropriate. Se mancano informazioni questo diventa impossibile, ma avere una quantità elevata di informazioni è sempre utile? Per rispondere a questa domanda occorre probabilmente calare la domanda in contesti specifici tenendo in considerazione il destinatario finale e lo scopo che si vuole raggiungere:



un livello di dettaglio troppo alto infatti potrebbe in alcuni casi essere persino controproducente. Facendo una ricerca su quanto emerge dalla letteratura riguardo la granularità ottimale in medicina non è possibile trarre indicazioni conclusive, gli studi attualmente disponibili non forniscono infatti evidenze sufficientemente precise riguardo alle singole persone.

Dati e medicina delle evidenze

Queste riflessioni sulle caratteristiche dei dati e sugli strumenti per la loro gestione non possono prescindere da riflessioni più generali su come sta cambiando la medicina e in particolare la medicina basata sulle evidenze. Il concetto di EBM è cambiato nel corso degli anni e sempre più si sta passando da una medicina centrata sulla malattia a una medicina centrata sul paziente. Se da una parte gli studi clinici controllati continuano a essere il fondamento della ricerca clinica e dell'innovazione terapeutica in quanto sono lo strumento capace di fornire prove solide da cui poter trarre conclusioni sull'efficacia di un farmaco o in generale di un intervento terapeutico, i rischi e le raccomandazioni, dall'altra non è più sufficiente avere evidenze che si basino su pazienti medi rappresentativi di una condizione ideale, ma occorrono strategie per studiare le terapie nella vita reale. Lo sviluppo della genomica e delle altre scienze omiche ha dato il via a una nuova era della medicina di precisione, ma perché questo inizio abbia un seguito virtuoso occorre di nuovo sottolineare l'importanza

dell'analisi e della gestione dei dati. È grazie alla gestione efficace ed efficiente di della enorme immensa mole di dati infatti che possiamo migliorare le cure andando verso una medicina sempre più precisa e cucita ("sartorializzata") su misura per il paziente.

Dati e ricerca clinica

Più aumenta la disponibilità dei dati, maggiori sono i cambiamenti e le innovazioni anche nell'ambito della ricerca clinica. La recente esperienza con COVID-19 ha dato l'opportunità di comprendere l'importanza del lavoro sinergico per la generazione di nuove evidenze. Inoltre durante la pandemia per le limitazioni che ne sono conseguite è stato necessario rivedere tutte le fasi della ricerca portando allo sviluppo di nuovi studi in un'ottica sempre più centrata sul paziente. In particolare è in crescita l'attenzione per gli studi perpetui o perenni, caratterizzati dall'assenza in fase di progettazione di una data di fine prestabilita, proprio per tenere aperta la possibilità di confronto e per poter verificare l'efficacia o l'inefficacia di un farmaco nel lungo periodo. L'arruolamento dei pazienti in continuo consente di raccogliere un campione demografico che è molto più simile per caratteristiche cliniche ed esiti a quello che si osserva nella pratica quotidiana piuttosto a quanto si verifica con un protocollo classico.

Prospettive

Il FSE si presenta come uno strumento di grande potenzialità. Occorrono però ancora interventi strategici per rendere il FSE più snello e più facilmente fruibile in modo da creare

realmente le condizioni per una medicina diffusamente innovativa e personalizzata. La ricerca clinica in continuo potrebbe essere una strada molto interessante, non è un caso che sempre in epoca pandemica l'EMA abbia attivato la procedura della rolling review, strumento regolatorio in base al quale il comitato per i medicinali per uso umano (CHMP) dell'EMA provvede ad esaminare i dati non appena diventano disponibili dagli studi in corso. Cioè in questo caso tutti i dati sull'efficacia, la sicurezza e la qualità di un medicinale o di un vaccino e tutta la documentazione richiesta non vengono presentati all'inizio della valutazione nell'ambito di una formale domanda di autorizzazione all'immissione in commercio ma in itinere. Questa necessità di valutare le prove in continuo toccata con mano nel caso di COVID-19 in realtà è un bisogno che si ha in tutti gli ambiti dalla ricerca clinica e che quindi dovrebbe essere opportunamente applicata diffusamente e gestita diffusamente nel tempo.♥