

## **Clelia Di Serio**

Università Vita e Salute S. Raffaele di Milano

### ***Scienza e coscienza nella medicina “data driven”***

Le tecnologia del “dato” è progredita parallelamente alle tecnologie biomediche rendendo il concetto di analisi epidemiologica, analisi predittiva e surveillance parti integrate di un unico processo alla base della prevenzione e degli interventi di sanità pubblica.

Così come la medicina di precisione impiega dati sui singoli pazienti per pianificare l'intervento terapeutico più adatto, la prevenzione di precisione fa uso dei “big data” per studiare l'impatto dei fattori di rischio sullo sviluppo di determinate patologie all'interno di una popolazione e per identificare strategie atte a prevenirne l'insorgenza o per ottimizzare l'uso delle pratiche e delle risorse terapeutiche esistenti.

Il concetto di “big data” in biomedicina è differente a seconda che si parli di ricerca di base o di ricerca clinica. Nella ricerca di base i big data sono spesso legati al dato “omico” che si riferisce alla genomica, proteomica, trascrittomica ovvero quei rami della biologia molecolare che studiano la struttura, la funzione, l'evoluzione e la mappatura del genoma. Il dato viene generato da diverse piattaforme tecnologiche e richiede strumenti computazionali altamente sofisticati per integrare le diverse fonti di dati ed interpretare le malattie. E' così che sono sorte alcune tra le piattaforme bioinformatiche più sofisticate, tipo Galaxy, nate per accogliere e rendere fruibili molti strumenti di analisi in unica interfaccia apparentemente facilmente utilizzabile.

Da un punto di vista della ricerca clinica i “big data” si associano invece alla creazione degli Electronic Medical Records (EMR) o più in generale Electronic Health Records (EHR) ovvero di delle piattaforme che raccolgono in modo informatizzato i dati di cartelle cliniche di enormi campioni di pazienti provenienti da consorzi di studi su malattie croniche e infettive con il fine di acquisire informazioni sulla salute “globale” del paziente da questo processo di “federazione” dei dati, ai fini della prevenzione.

Ma mettere i “big data” a servizio della salute è tutt'altro che semplice. Per quanti sforzi si stiano facendo in queste direzioni troppo spesso i “big data” non corrispondono a “big information”. Avere molti dati a disposizione in molti contesti biomedici aumenta anche la fallacia del ragionamento “causale” creando facili associazioni che vengono interpretate in termini “predittivi” laddove la “previsione” può risultare fuorviante in mancanza di un disegno sperimentale alla base. Sempre più spesso infatti è l'interpretazione causale dei “big data” ad essere responsabile di falsi risultati, che hanno portato anche alcune linee importanti di ricerca in false direzioni anche a causa dei rapidi cambi tecnologici che rendono serie temporali di dati poco confrontabili dopo pochi anni. Infatti a fronte dell'insorgere delle nuove biotecnologie, dell'aumento della mole e della complessità dei dati, dello sviluppo dei “supercomputer” capaci di sfruttare le potenzialità offerte dai “big data” non sempre vi è un aumento di “conoscenza” dovuto alla scarsa riproducibilità dei risultati. Diventa quindi necessaria una riflessione su quanto sia importante investire non solo sugli aspetti informatici ma sul processo di generazione e di estrazione dei dati, sulle tecniche di analisi che permettano alla medicina di acquisire sempre maggior “coscienza” non solo delle “evidenze” ma della necessità di integrazione delle conoscenze.

## **Antonietta Mira**

Istituto Lombardo – Università della Svizzera Italiana - Università dell’Insubria di Varese

### ***Scienze computazionali al servizio delle scienze della vita***

Molti fenomeni relativi alle scienze della vita si possono descrivere come una serie di complesse relazioni fra elementi che, a seconda della scala di osservazione, vanno da interazioni fondamentali fra particelle fisiche e molecole chimiche fino ad arrivare alle interazioni fra organismi, individui e specie. La comprensione scientifica di questi complessi fenomeni vitali si fonda sull’interazione di tre pilastri: la teoria, l’esperienza e la simulazione.

Nelle scienze computazionali, partendo dai modelli teorici e sfruttando potenzialità di calcolo sempre crescenti, il passaggio fondamentale per la comprensione di un fenomeno è rappresentato dall’analisi dei dati mediante diverse classi di algoritmi come, ad esempio, Machine Learning, Artificial Intelligence, Monte Carlo Simulation, Approximate Bayesian Computation. L’interpretazione dei risultati prodotti dagli algoritmi affronta sfide legate alla riduzione della dimensionalità dei dati nel tentativo di svelarne strutture nascoste, l’accuratezza e la quantificazione dell’incertezza nelle previsioni, la riduzione dei tempi di analisi e l’ottima gestione delle risorse. Simili sfide si possono vincere grazie ad una interazione fra i saperi che valorizzi il tema della riproducibilità della ricerca e sia attenta alla tutela della privacy.

Gli esempi con cui illustrerò come queste sfide vengono fronteggiate dalle scienze computazionali si riferiscono a diversi contesti empirici, tutti ben rappresentati nel Laboratorio di Data Science che dirigo presso l’Università della Svizzera italiana. In ambito biologico, illustrerò studi relativi alla dinamica molecolare, alla risonanza magnetica funzionale e all’interazione fra proteine. In ambito medico, riferirò analisi relative alla gestione delle emergenze cardiovascolari mediante il posizionamento ottimale di defibrillatori sul territorio ed alla previsione della diffusione di epidemie nel tempo e nello spazio.

Per sviluppare teorie innovative, disegnare algoritmi efficienti e analizzare dati per supportare efficacemente decisioni, non si può prescindere da un approccio multidisciplinare che faccia perno su curiosità, fantasia e rigore. È quindi fondamentale che, al centro della rivoluzione digitale e computazionale che stiamo vivendo, venga posto l’uomo con le sue doti di ricercatore. Solo così, infatti, troverà compimento la sfida più ardua: estrarre informazioni dai dati sotto forma di conoscenza tesa a guidare decisioni a supporto delle scienze della vita.