

Data sfera, big data, blockchain

La nuova frontiera del bibliotecario biomedico 3.0

FRANCESCA GUALTIERI

Scientific Information & Library Services
Rottapharm Biotech, Monza
francesca.gualtieri@rottapharmbiotech.com

SILVIA MOLINARI

Biblioteca Scientifica IRCCS Fondazione Mondino, Pavia
silvia.molinari@mondino.it

(Per il GIDIF-RBM – Gruppo italiano documentalisti
industria farmaceutica e istituti di ricerca biomedica)

Tra i numeri primi ce ne sono alcuni ancora più speciali. I matematici li chiamano primi gemelli: sono coppie di numeri primi che se ne stanno vicini, anzi quasi vicini, perché fra di loro vi è sempre un numero pari che gli impedisce di toccarsi per davvero.

(Paolo Giordano, *La solitudine dei numeri primi*)

Da Pitagora alla Data Sfera

1, 2, 3 ... 1021 byte... questo è il volume di dati da cui saremo inondati nella prossima decade, un numero che metterebbe soggezione anche al celebre matematico e filosofo greco Pitagora, una quantità di dati 180 milioni di volte più grande dell'ammontare di tutta la documentazione mai raccolta e conservata in una biblioteca, la più grande biblioteca del mondo. Così come abbiamo creato un'isola con lo scarto della plastica al largo dell'oceano, così creiamo un pianeta di dati con l'utilizzo delle tecnologie, la *data sfera*, un

pianeta che si auto-alimenta senza alcun intervento programmato e organizzato, senza richiedere la gestione attiva né dell'uomo né della comunità dell'informazione che lo crea. Un'entità autonoma, dunque, che arriverà a governare le infrastrutture dei servizi, dei dispositivi e forse anche le nostre vite! Ma che cos'è esattamente questa *data sfera*? L'intento di questa riflessione è quello di capirlo e soprattutto di analizzare la futura applicazione in ambito scientifico.

Nel prossimo futuro oltre il 70% della popolazione mondiale sarà connessa a un dispositivo (cellulare, tablet, computer o altro) e potrà ricevere o fornire informazioni in tempo reale grazie alla connettività 5G; questo avverrà senza prestare alcuna attenzione alla protezione e alla sicurezza del canale informativo e senza valutare l'affidabilità e bontà dei sistemi. Pensiamo a tutti i dati che ciascuno di noi ogni giorno offre più o meno consapevolmente con estrema facilità e gratuitamente attraverso il semplice utilizzo del cellulare; questi dati alimentano la data-sfera ogni attimo della nostra giornata. Tale meccanismo porta

a riprendere la riflessione sulle parole del creatore del sistema più virale degli ultimi decenni, il padre del mela-phonino, Steve Jobs: “se il servizio è gratis significa che il prodotto sei tu”. Se tutto è gratuitamente alla portata di tutti in tempo reale, vuol dire che il prodotto in vendita siamo esattamente noi (!). Questa affermazione ci pone di fronte a una significativa presa di coscienza: da utenti dobbiamo sapere qual è il valore che auto-produciamo.

Questa è una nuova frontiera socio-economica, un modello di business innovativo dove i dati sono il nuovo petrolio, il nuovo oro nero dell'era digitale. Un olio nero che sotto forma di byte alimenta un commercio tra le aziende che raccolgono dati e quelle che vorrebbero riutilizzarli; sembra che gli iscritti a Facebook possano produrre un valore economico pari a 32 miliardi di dollari all'anno, cifra di tutto interesse e rispetto.

Data sfera e big data

L'immensa mole di dati o tracce seminate dai cibernauti in rete costituisce la nostra *data sfera* e dà l'avvio all'era dei *big data*. Era in cui le persone frammentate in tracce informatiche vengono poi ricostruite minuziosamente fino allo spasmo dei dettagli personali, in algoritmi di abitudini, percentuali di utilizzo, scelte e mode (in senso statistico). Il nostro stile di vita, quello che mangiamo e perfino sogniamo diventano dati che possono essere riutilizzati per indottrinarci o, spereremmo, farci vivere meglio e più a lungo. Pensiamo ai dati sulle nostre abitudini alimentari o più in generale ai dati sulla nostra salute: possono essere utilizzati a scopo di ricerca per sviluppare nuovi farmaci o mettere a punto nuove terapie per la cura del diabete.

In passato abbiamo visto che lo studio attento di alcuni effetti collaterali di alcune sostanze attive ha portato al riposizionamento di alcuni farmaci per la cura di malattie orfane. L'intelligenza artificiale messa al servizio della persona può prevenire o curare malattie. Nell'era del “grande fratello” dove tutti vedono tutto, possiamo incanalare questa mole di dati, questa energia-risorsa, per il benessere pubblico? Come possiamo assicurare che il progresso tecnologico sia un buon mezzo al servizio dello stesso produttore? La tecnologia riuscirà a salvare la tecnologia? Riusci-

remo a usare bene e in modo sicuro i dati? La sicurezza in tutte le sue declinazioni sarà un argomento delicato di cui occuparsi e soprattutto da gestire nel prossimo futuro. Ottenere le giuste autorizzazioni per usare bene questi dati è la sfida che devono vincere i governi e le imprese che attraverso la blockchain intendono mettere al servizio di altre società i dati prodotti dall'utilizzo della rete da ognuno di noi.

La blockchain e l'Open Science Cloud

La blockchain è una tecnologia che attraverso un'architettura (informatica) decentralizzata, facilita lo scambio in sicurezza delle informazioni dei cittadini. Nata inizialmente nel settore finanziario allo scopo di proteggere le transazioni online da attacchi di pirateria informatica, rappresenta il modo sicuro di scambiare informazioni su internet.

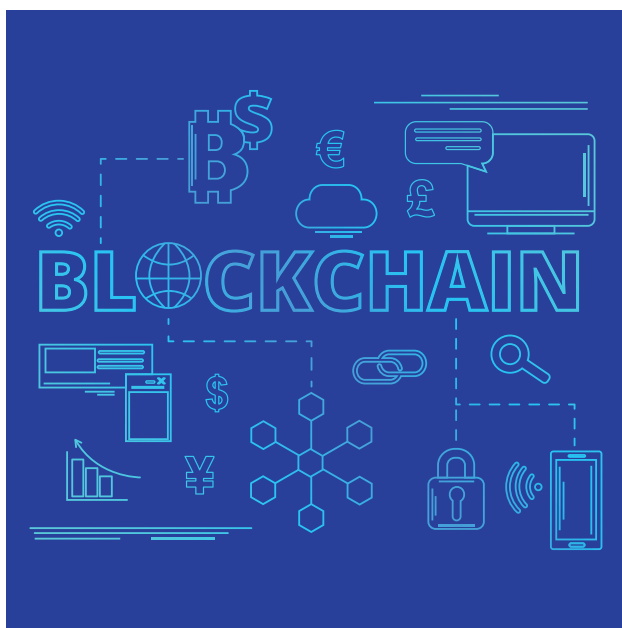
La scorsa primavera la Commissione Europea ha lanciato una serie di proposte, nell'ambito del mercato digitale in ambito sanitario, tra cui ricordiamo la creazione dell'open-science cloud, incentiva l'utilizzo di tecniche di estrazione di dati - *data mining* - e favorisce la ricerca collaborativa, dove il cittadino è il centro, il fulcro dell'attenzione degli operatori sanitari; la salute e il benessere (del cittadino) verranno garantiti attraverso l'utilizzo dei suoi stessi dati clinici condivisi.

I *big data* dovranno essere impegnati per facilitare e migliorare i processi di diagnosi e consentire al medico di proporre e applicare trattamenti terapeutici personalizzati per quel paziente o gruppi di pazienti. Essi potrebbero facilitare l'identificazione precoce di segnali di tossicità su piccoli gruppi di pazienti permettendo al clinico un rapido adeguamento del protocollo terapeutico. Un'altra applicazione di questi dati potrebbe essere il monitoraggio da remoto delle cartelle cliniche per facilitare le decisioni di R&D e il destino delle pipeline delle case farmaceutiche.

Watson for Drug Discovery (WDD), la suite di servizi di IBM, ad oggi ha raccolto oltre 25 milioni di abstract da Medline, oltre un milione di articoli scientifici e 4 milioni di brevetti di molecole che con la partecipazione di Pfizer intende utilizzare per selezionare nuove strategie d'intervento in immuno-oncologia per l'identificazione di nuovi target terapeutici e relativa selezione di pazienti. Un altro esempio ci è dato

da Novartis che dal giugno 2017 si avvale di questo approccio per migliorare la previsione degli esiti delle possibili terapie di cura della patologia tumorale.

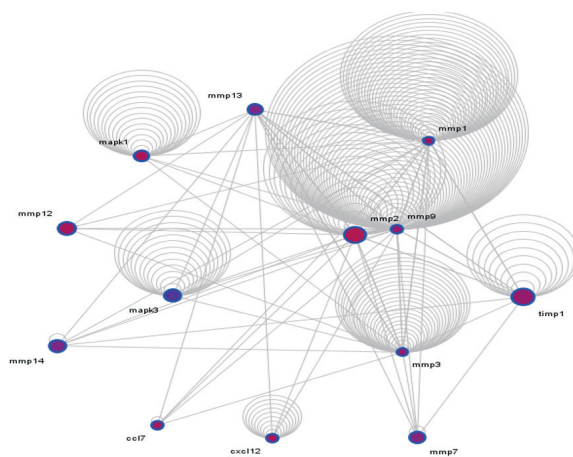
E ancora si consideri la piattaforma dell'Istituto Superiore di Sanità per lo studio delle terapie delle epatiti virali, Piter (<http://old.iss.it/piter>), dove i centri afferenti al progetto condivideranno e utilizzeranno i contenuti secondo un approccio trasparente; tutti vedranno in tempo reale le modifiche e gli aggiornamenti dei dati garantendo la salvaguardia della privacy del paziente. I clinici, gli epatologi e infettivologi potranno consultare la più grande raccolta di casi con i dati provenienti da 10.000 pazienti.



I big data e la ricerca biofarmaceutica

Ma la vera e innovativa applicazione di questa tecnologia è indirizzata alla ricerca biofarmaceutica di nuovissima generazione dove strutture smart di laboratori R&D (Research and Development) utilizzano grandi banche dati, elaborano grandi moli di dati per identificare le molecole più promettenti da avviare allo sviluppo pre-clinico. Fra gli esempi di target farmacologici si annoverano le proteine che causano infiammazione, che stimolano la crescita tumorale o le proteine utilizzate dai virus per infettare le cellule. Nella ricerca farmacologica, l'obiettivo è creare molecole, piccole-molecole, che interagiscano in maniera decisiva con i substrati proteici target. Compito

assolto da AtomNet algoritmo statistico, creato con la stessa tecnologia di Atomwise per il riconoscimento facciale, che in poco tempo è in grado di predire questi legami (molecola-proteina). Un sistema del genere può predire le migliori molecole candidate per lo sviluppo di un nuovo farmaco, offrendo ai ricercatori un metodo virtuale meno costoso, più rapido e ugualmente efficace.



La figura illustra il metodo network per individuare possibili target farmacologici di interazione tra proteine. In inglese: PP-Interactiortarget discovery PP-Interaction network

Il bibliotecario 3.0

In questo scenario da supereroi informatici continua a offrire la sua attenta e peculiare professionalità il bibliotecario biomedico, l'information specialist che dà e darà ai ricercatori accesso alle più sofisticate tassonomie creando e condividendo i vocabolari tecnici sulle piattaforme delle banche dati. Un team collaborativo composto da biologi, chimici, statistici, informatici e bio-informatico-bibliotecari farà quasi sicuramente in futuro la differenza nel *drug discovery*. Questo ruolo necessiterà di una preparazione superiore per tutti, non solo per il bibliotecario 3.0; le associazioni dovranno interrogarsi sui punti focali della nuova educazione nel dare competenze trasversali in stretta collaborazione con le università; nell'ambito degli studi di Digital Humanities sarà possibile formare competenze trasversali dell'informatica e delle scienze biblioteconomiche. Sfruttare l'intelligenza artificiale per rispondere alle nuove domande degli utenti esperti e specializzati, domande sempre



più complesse che vanno oltre l'interrogazione di una banca dati e che intersecano saperi diversi al servizio del benessere e della salute dei cittadini, sarà la nuova frontiera del bibliotecario del futuro che, del tutto verosimilmente, saprà creare connessioni e dialogherà con la data-sfera per restituire informazioni di qualità e di sicuro valore; orienterà l'utente a estrapolare le informazioni di qualità e saprà essere giusto mediatore tra l'intelligenza artificiale e quella umana, creatrice quest'ultima di connessioni e relazioni uniche; non solo, saprà conservarle e saprà garantirne la continua fruizione.

BIBLIOGRAFIA

- JING Y., BIAN Y., HU Z., WANG L., XIE XQ., *Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era*, "The AAPS Journal", 20 (2018), 3, p. 58.
- BROWN N., CAMBRUZZI J., COX PJ., ET AL., *Big data in drug discovery*, "Progress in medical chemistry", 57 (2018), 1, p. 277-356.
- BUTTE AJ., *Big data opens a window onto wellness*, Nature Biotechnology, 35 (2017), 8, p. 720-721.
- CAPASSO A., *Blockchain in sanità. Nuova frontiera della digitalizzazione*, "About Pharma Medical Device 2018-2019" n. 164, p. 42-45.
- KONDILI L., QUARANTA MG., ET AL., *PIPER-HCV cohort study as part of the Italian platform for the study of viral hepatitis therapies*, Roma, Istituto superiore di sanità, 2015.
- PANI L., *I big data rivoluzioneranno anche il mondo del farmaco?*, <https://www.aifa.gov.it/-/i-big-data-rivoluzioneranno-anche-il-mondo-del-farmaco>
- WALLACH I., DZAMBA M., HEIFETS A., *Atomnet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery*, <https://arxiv.org/abs/1510.02855>
- BELLINI M., *Blockchain: cos'è, come funziona e gli ambiti applicativi in Italia*, "Blockchain4innovation.it", <https://www.blockchain4innovation.it/esperti/blockchain-perche-e-così-importante>.
- Come la tecnologia cambierà il mondo della medicina*, "Fastweb.it", <https://www.fastweb.it/web-e-digital/come-la-tecnologia-cambierà-il-mondo-della-medicina>.

Ultimi accessi online 19/07/2019

ABSTRACT

In future, over 70% of the world population will be connected to a device and will be able to receive or provide information in real time. Think about data that each of us offers every day, through the simple use of mobile phones. All these data feed the data-sphere every moment of our day life. Users produce value, the new socio-economic frontier, a new business model, the black oil of digital era. This paper analyzes what the data-sphere will represent, how it works and what are the possible applications into the bio-medical research. Big Data will be committed to facilitate and improve the diagnostic processes and therapeutic treatments for patients. This innovative technology is addressed to the latest generation of biopharmaceutical researchers. In computer super-hero era, the bio-medical librarians or information specialists, give and will continue to give access to the most sophisticated taxonomies. Exploiting artificial intelligence to answer new questions, they will be able to create connections with the data-sphere and knowledge to return patrons good quality of information service. Bio-medical librarian will be the right cultural mediator between artificial and human intelligence.

DOI: 10.3302/0392-8586-201906-038-1