



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

«TANTI DATI PER CHI? COME
UTILIZZARLI NEL WELFARE DI
COMUNITÀ E NELL'ECONOMIA
DEL WELFARE»

BIG DATA, SFIDE E PROSPETTIVE
PER LA SANITA' PUBBLICA DEL
FUTURO

Davide Gori

15 May 2023

Cosa sono i Big Data?

Non esiste una definizione singola, prendiamo Wikipedia:

- In statistica e informatica, la locuzione inglese big data ("grandi [masse di] dati") o l'italiana megadati indica genericamente una raccolta di dati informatici così estesa in termini di volume, velocità e varietà da richiedere tecnologie e metodi analitici specifici per l'estrazione di valore o conoscenza. Il termine è utilizzato dunque in riferimento alla capacità (propria della scienza dei dati) di analizzare ovvero estrapolare e mettere in relazione un'enorme mole di dati eterogenei, strutturati e non strutturati (grazie a sofisticati metodi statistici e informatici di elaborazione), al fine di scoprire i legami tra fenomeni diversi (ad esempio correlazioni) e prevedere quelli futuri. I big data possono essere utilizzati per diversi scopi tra cui quello di misurare le prestazioni di un'organizzazione nonché di un processo aziendale.



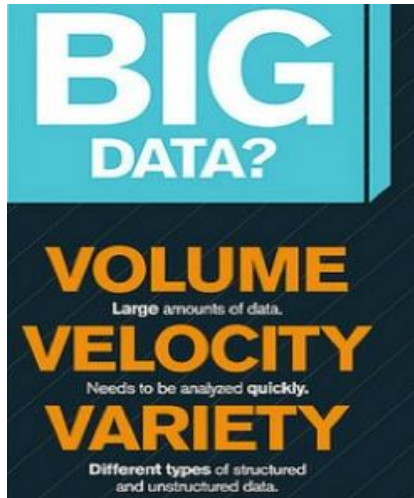
Cosa sono i Big Data?

Non esiste una definizione singola, prendiamo Wikipedia:

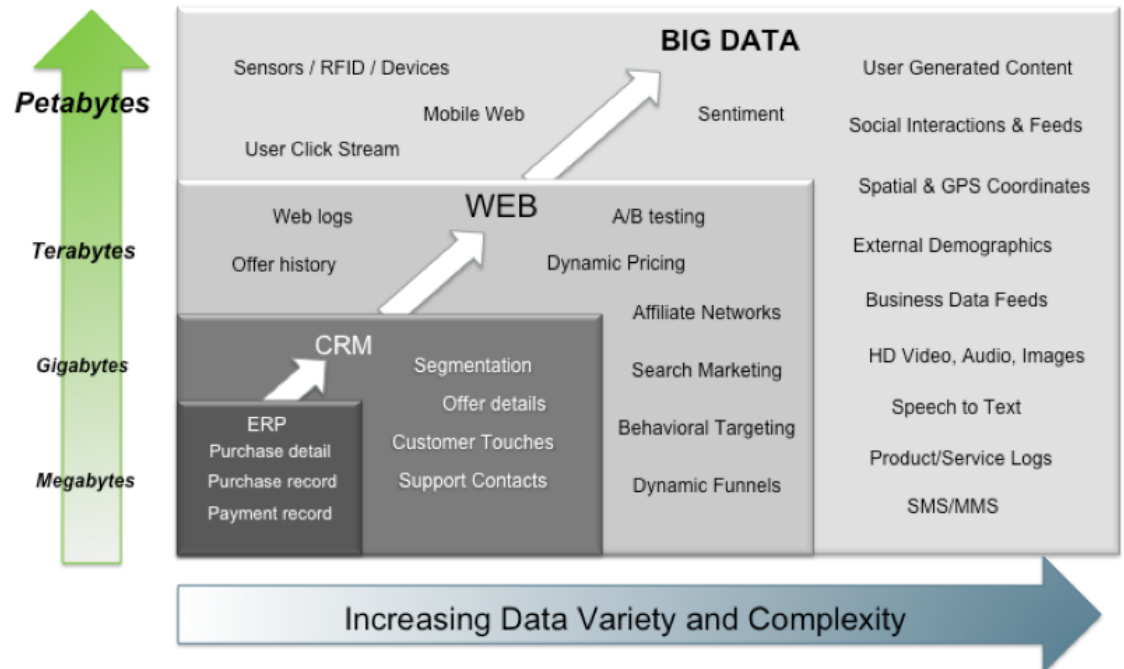
- Le sfide dei big data comprendono l'acquisizione, la cura, l'archiviazione, la ricerca, la condivisione, il trasferimento, l'analisi e la visualizzazione di essi.
- La tendenza a creare insiemi di dati più grandi è dovuta alle informazioni aggiuntive che si possono ricavare dall'analisi di un unico grande insieme di dati correlati, rispetto a insiemi separati più piccoli con la stessa quantità totale di dati, consentendo di trovare correlazioni ad esempio per individuare le tendenze di mercato, determinare la qualità della ricerca, prevenire le malattie, combattere il crimine e determinare le condizioni del traffico stradale in tempo reale.



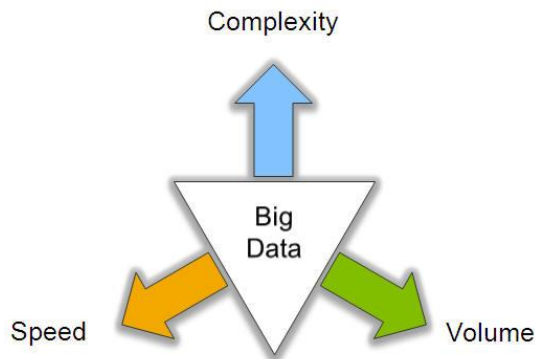
Big Data: 3V's



Big Data = Transactions + Interactions + Observations



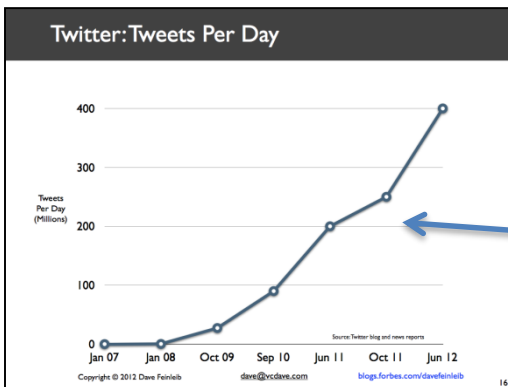
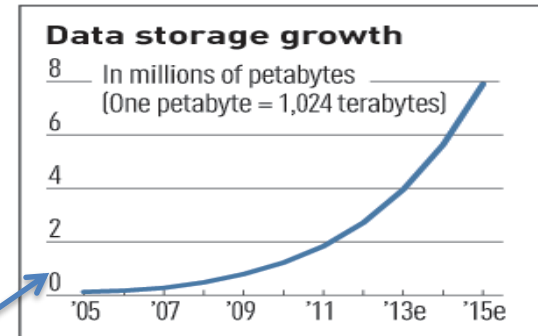
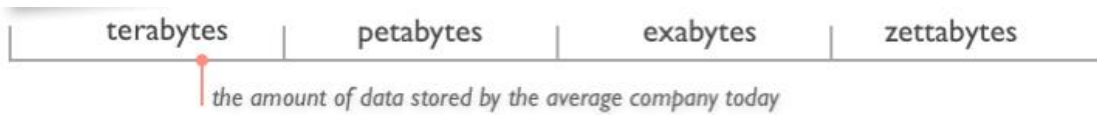
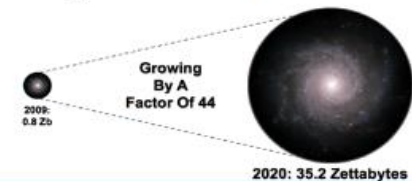
Source: Contents of above graphic created in partnership with Teradata, Inc.



Volume (Scala)

- **Volume di dati:**
 - Aumento di 44 volte dal 2009 al 2020
 - Da 0,8 zettabyte a 35zb
- **Il volume dei dati sta aumentando in modo esponenziale**

The Digital Universe 2009-2020



Exponential increase in collected/generated data



? TBs of data every day

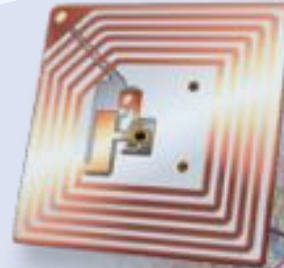
12+ TBs of tweet data every day



25+ TBs of log data every day



30 billion RFID tags today (1.3B in 2005)



4.6 billion camera phones world wide



100s of millions of GPS enabled devices sold annually



76 million smart meters in 2009... 200M by 2014

2+ billion people on the Web by end 2011





CERN's Large Hydron Collider (LHC) generates 15 PB a year



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Velocità

- I dati iniziano ad essere generati velocemente e devono essere elaborati velocemente
- Analisi dei dati online
- Decisioni tardive → opportunità mancate
- **Esempi**
 - **Promozioni elettroniche:** In base alla vostra posizione attuale, alla vostra cronologia degli acquisti, a ciò che vi piace → invia promozioni in questo momento per il negozio più vicino a voi
 - **Monitoraggio sanitario:** sensori che monitorano le vostre attività e il vostro corpo → qualsiasi misurazione anomala richiede una reazione immediata



Real-time/Fast Data



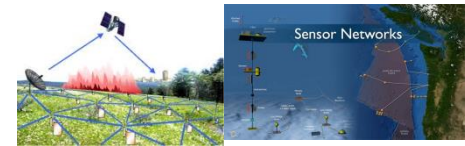
Social media and networks
(all of us are generating data)



Scientific instruments
(collecting all sorts of data)



Mobile devices
(tracking all objects all the time)



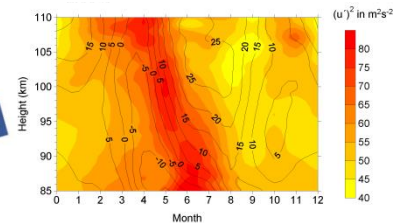
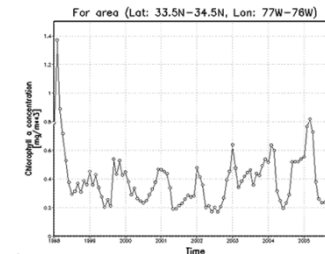
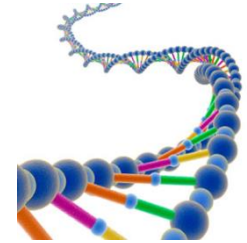
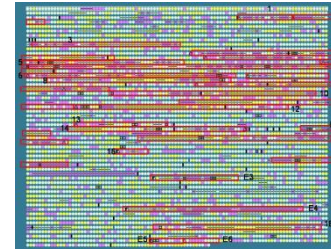
Sensor technology and networks
(measuring all kinds of data)

- Il progresso e l'innovazione non sono più ostacolati dalla capacità di raccogliere dati, ma dalla capacità di gestire, analizzare, riassumere, visualizzare e scoprire la conoscenza dai dati raccolti in modo tempestivo e scalabile.



Varietà (Complessità)

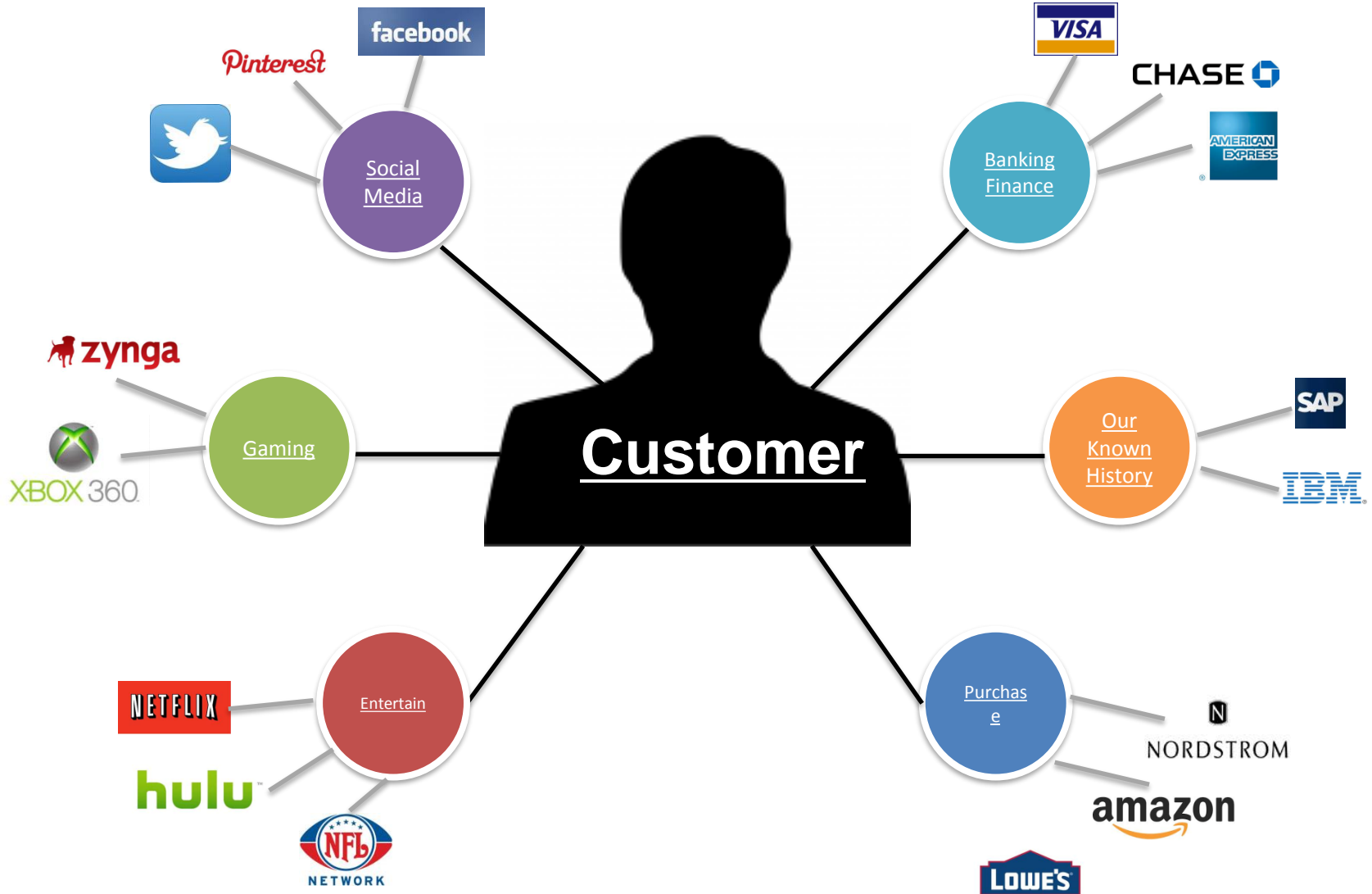
- Dati relazionali (tabelle/transazioni/dati di legacy)
- Dati testuali (Web)
- Dati semi-strutturati (XML)
- Dati grafici
 - Rete sociale, Web semantico (RDF), ...
- Dati in streaming
 - È possibile analizzare i dati una sola volta
- Una singola applicazione può generare/raccogliere molti tipi di dati
- Grandi dati pubblici (online, meteo, finanza, ecc.)



Per estrarre la conoscenza → tutti questi tipi di dati devono essere collegati tra loro.

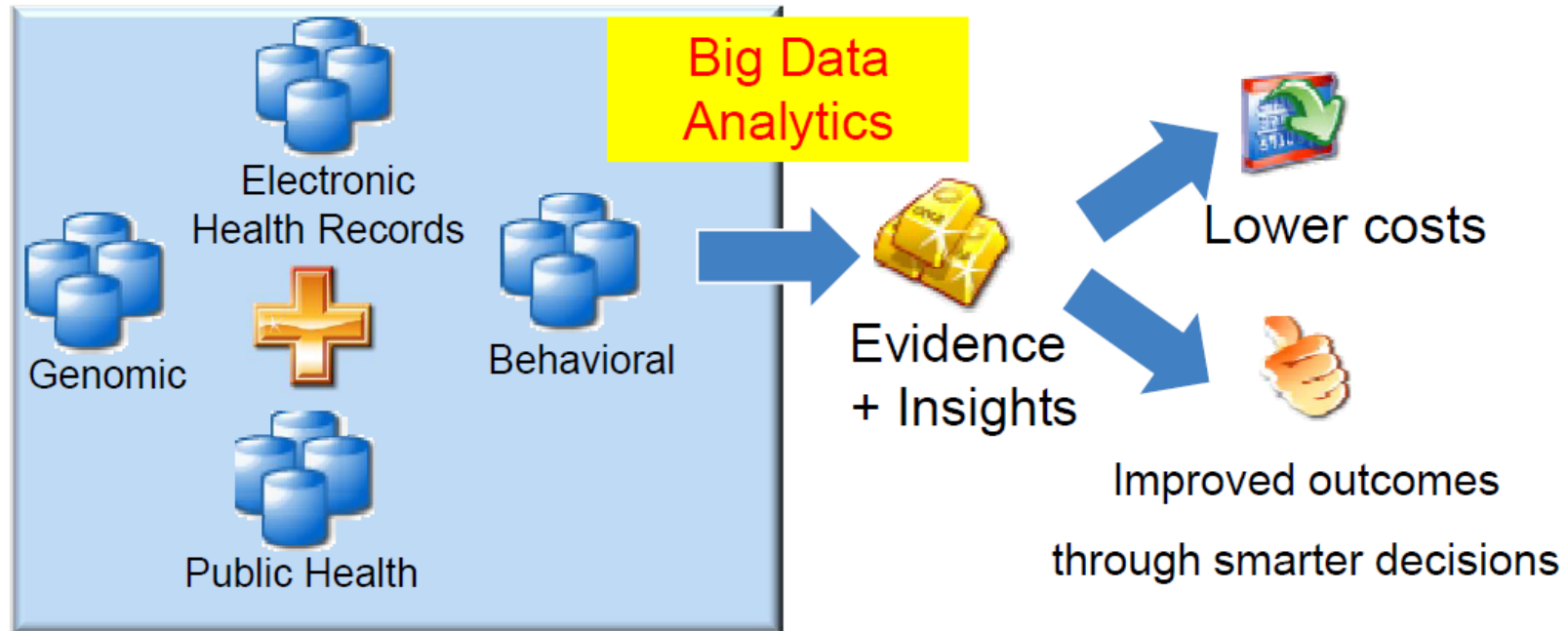


Esempio: in ambito commerciale - Un'unica visione del cliente



COSA SONO I BIG DATA NELLA SANITA' PUBBLICA?

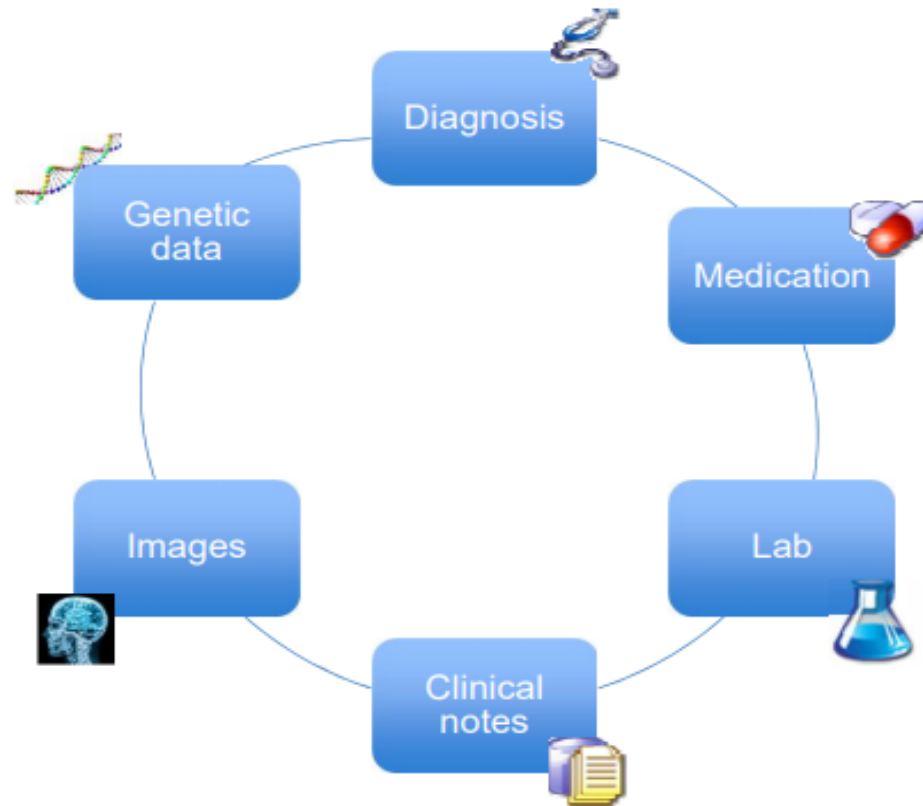
Overall Goals of Big Data Analytics in Healthcare



GOAL: Provide Personalized care through right intervention to the right patient at the right time.

L'ANALISI DEI BIG DATA

Heterogeneous Medical Data



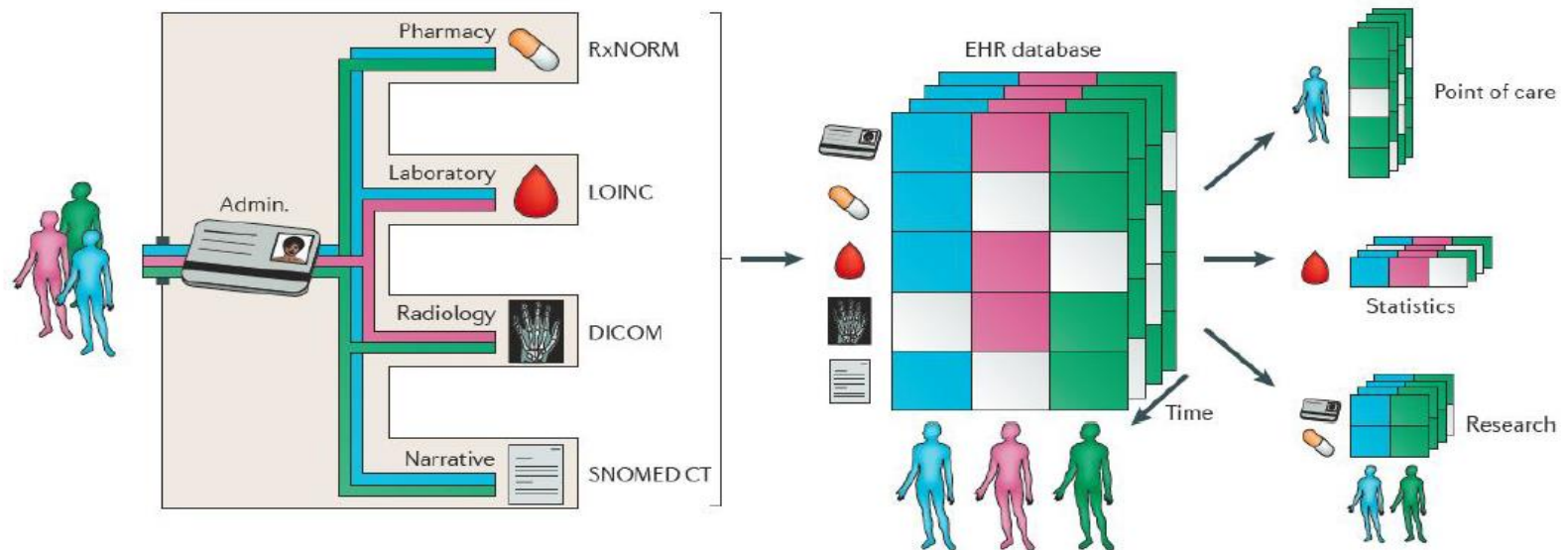
Jimeng Sun, Large-scale Healthcare Analytics



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

L'ANALISI DEI BIG DATA

Data Collection and Analysis



Effectively integrating and efficiently analyzing various forms of healthcare data over a period of time can answer many of the impending healthcare problems.

Jensen, Peter B., Lars J. Jensen, and Søren Brunak. "Mining electronic health records: towards better research applications and clinical care." *Nature Reviews Genetics* (2012).



L'ANALISI DEI BIG DATA

Different Kinds of Outcomes



Economic Cost

Continuous Outcomes



Diagnostic

Presence of
a disease

Binary Outcomes

What
disease

Categorical Outcomes

How
serious

Ordinal Outcomes



Survival analysis

Survival Outcomes



INTERNET OF THINGS

- Il settore sanitario non è stato abbastanza veloce nell'adattarsi al movimento dei big data rispetto ad altri settori. Pertanto, l'utilizzo dei big data nel settore sanitario è **ancora agli inizi**. Ad esempio, i big data sanitari e biomedici non sono ancora convergenti per migliorare i dati sanitari con la patologia molecolare. **Tale convergenza può aiutare a svelare vari meccanismi d'azione o altri aspetti della biologia predittiva**. Pertanto, per valutare lo stato di salute di un individuo, è necessario **coniugare i set di dati biomolecolari e clinici**. Una di queste fonti di dati clinici in ambito sanitario è *l'Internet of Things(IoT)*.



INTERNET OF THINGS

- Utilizzando la rete di dispositivi IoT, un medico può misurare e monitorare vari parametri dai suoi pazienti nei luoghi di vita e di lavoro. Pertanto, grazie a un intervento e a un trattamento tempestivi, un paziente potrebbe non aver bisogno di un ricovero ospedaliero o di una visita medica, con una conseguente **significativa riduzione dei costi sanitari**. Alcuni esempi di dispositivi IoT utilizzati in ambito sanitario includono dispositivi indossabili **per il fitness o il monitoraggio della salute, biosensori, dispositivi clinici** per il monitoraggio dei segni vitali e altri tipi di dispositivi o strumenti clinici. **Questi dispositivi IoT generano una grande quantità di dati sulla salute.**

Nasi G, Cucciniello M, Guerrazzi C. The role of mobile technologies in health care processes: the case of cancer supportive care. J Med Internet Res. 2015;17(2):e26.



Mobile computing and mobile health (mHealth)

- La pratica della medicina e della sanità pubblica con l'uso di dispositivi mobili, **nota come mHealth o mobile health**, pervade diversi livelli di assistenza sanitaria, soprattutto per le malattie croniche, come il diabete e il cancro. **Le organizzazioni sanitarie utilizzano sempre più spesso servizi di salute e benessere mobili per implementare modi nuovi e innovativi di fornire assistenza e coordinare la salute e il benessere.** Le piattaforme mobili possono **migliorare l'assistenza sanitaria** accelerando la comunicazione interattiva tra pazienti e operatori sanitari. Infatti, Apple e Google hanno sviluppato recentemente piattaforme dedicate come ResearchKit di Apple e Google Fit per sviluppare applicazioni di ricerca per il fitness e le statistiche sulla salute.

Nasi G, Cucciniello M, Guerrazzi C. The role of mobile technologies in health care processes: the case of cancer supportive care. J Med Internet Res. 2015;17(2):e26.

Apple, ResearchKit/ResearchKit: ResearchKit 1.5.3. 2017.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Estrarre l'informazione dalle cartelle cliniche elettroniche

- Le strategie emergenti basate su ML o AI stanno contribuendo a perfezionare le capacità di elaborazione delle informazioni del settore sanitario. **Ad esempio, il natural language processing (NLP) è un'area in rapido sviluppo dell'apprendimento automatico in grado di identificare le strutture sintattiche chiave in un testo libero**, di aiutare nel riconoscimento vocale e di estrarre il significato dietro una narrazione (i.e. diari clinici). **Gli strumenti NLP possono aiutare a generare nuovi documenti, come il riassunto di una visita clinica**, o a dettare note cliniche. Il contenuto unico e la complessità della documentazione clinica possono rappresentare una sfida per molti sviluppatori NLP. Ciononostante, dovremmo essere in grado di estrarre informazioni rilevanti dai dati sanitari utilizzando approcci come **l'NLP. AI nelle decisioni?**

Nasi G, Cucciniello M, Guerrazzi C. The role of mobile technologies in health care processes: the case of cancer supportive care. J Med Internet Res. 2015;17(2):e26.

Apple, ResearchKit/ResearchKit: ResearchKit 1.5.3. 2017.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

As a library, NLM provides access to scientific literature. Inclusion in an NLM database does not imply endorsement of, or agreement with, the contents by I



EGEMS (Wash DC). 2019; 7(1): 43.
Published online 2019 Aug 20. doi: 10.53

Extracting Patient-Centered (C Records: Assessment of Urin

Davide Gori,¹ Imon Banerjee,² Benjamin
 James D. Brooks,² and Tina Hernandez-

▶ Author information ▶ Article notes ▶ C

EGEMS (Wash DC)

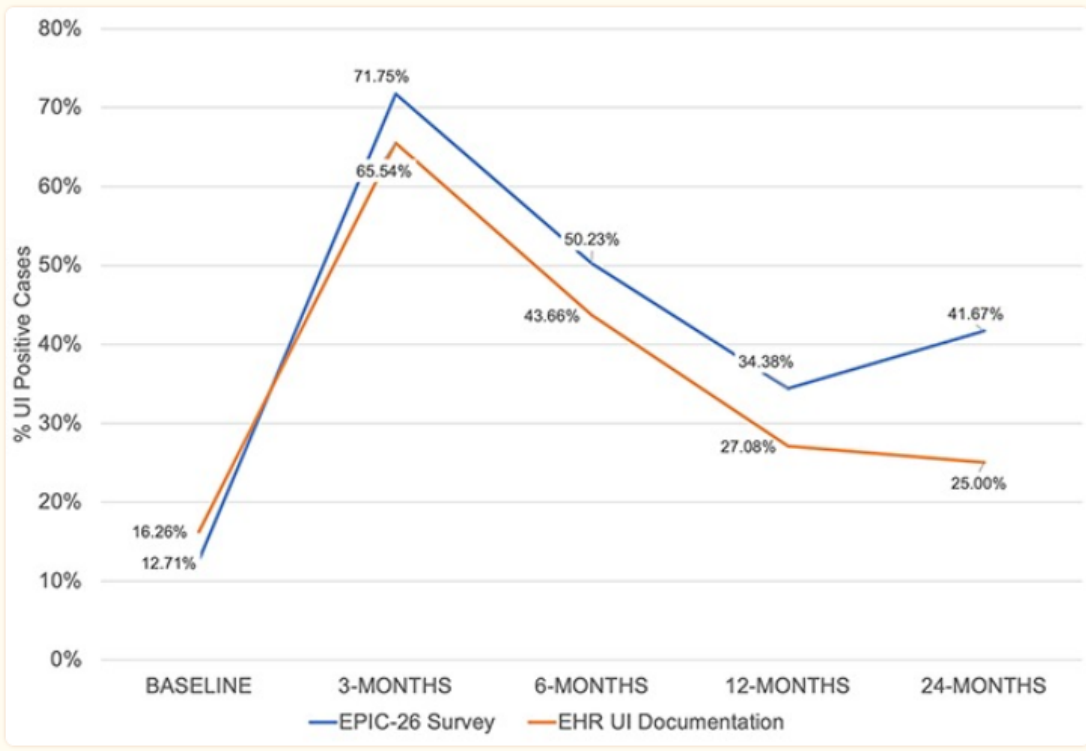


Figure 2

Urinary Incontinence Prevalence from EHR and EPIC-26 at Baseline and Post-Surgery.

Analisi delle immagini

- Alcune delle tecniche di imaging più utilizzate in ambito sanitario sono la tomografia computerizzata (TC), la risonanza magnetica (RM), i raggi X, l'imaging molecolare, gli ultrasuoni, l'imaging foto-acustico, la risonanza magnetica funzionale (fMRI), la tomografia a emissione di positroni (PET), l'elettroencefalogramma (EEG) e le mammografie. **Queste tecniche acquisiscono immagini mediche ad alta definizione (dati del paziente) di grandi dimensioni. I professionisti del settore sanitario, come radiologi, medici e altri, svolgono un lavoro eccellente nell'analisi dei dati medici sotto forma di questi file alla ricerca di anomalie mirate.**
- Sono stati sviluppati **sistemi efficienti come il Picture Archiving and Communication System (PACS) per archiviare e accedere comodamente ai dati** delle immagini e dei referti medici. Sono stati sviluppati numerosi strumenti software basati su funzionalità come la generica, la registrazione, la segmentazione, la visualizzazione, la ricostruzione, la simulazione e la diffusione per **eseguire l'analisi delle immagini mediche** al fine di scoprire le informazioni nascoste. Ad esempio, Visualization Toolkit è un software liberamente disponibile che consente una potente elaborazione e analisi di immagini 3D provenienti da test medici, mentre SPM è in grado di elaborare e analizzare 5 diversi tipi di immagini cerebrali (ad esempio MRI, fMRI, PET, TAC ed EEG).

Big data dalle omiche

I big data provenienti dagli studi "omici" rappresentano un nuovo tipo di sfida per i bioinformatici. Sono necessari algoritmi robusti per analizzare questi dati complessi provenienti da sistemi biologici. L'obiettivo finale è quello di convertire **questi enormi dati in una base di conoscenza informativa.**

L'applicazione di approcci bioinformatici per trasformare i dati biomedici e genomici in dati sanitari predittivi e preventivi è nota come bioinformatica traslazionale. È all'avanguardia nell'assistenza sanitaria guidata dai dati.

Diversi tipi di dati quantitativi in ambito sanitario, ad esempio le misurazioni di laboratorio, i dati sui farmaci e i profili genomici, possono essere combinati e utilizzati per identificare nuovi meta-dati che possono aiutare le terapie di precisione.

Per questo motivo sono necessarie nuove tecnologie emergenti che aiutino ad analizzare questo patrimonio digitale. In effetti, sono stati lanciati progetti multimilionari molto ambiziosi come "Big Data Research and Development Initiative", che mirano a migliorare la qualità degli strumenti e delle tecniche per i big data per una migliore organizzazione, un accesso efficiente e un'analisi intelligente dei big data.

Li L, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. *Sci Transl Med.* 2015;7(311):311ra174.



Servizi di big data

Table 2 List of some of big companies which provide services on big data analysis in healthcare sector

Company	Description	Web link
IBM Watson Health	Provides services on sharing clinical and health related data among hospital, researchers, and provider for advance researches	https://www.ibm.com/watson/health/index-1.html
MedeAnalytics	Provides performance management solutions, health systems and plans, and health analytics along with long track record facility of patient data	https://medeanalytics.com/
Health Fidelity	Provides management solution for risks assessment in workflows of healthcare organization and methods for optimization and adjustment	https://healthfidelity.com/
Roam Analytics	Provides platforms for digging into big unstructured healthcare data for getting meaningful information	https://roanalytics.com/
Flatiron Health	Provides applications for organizing and improving oncology data for better cancer treatment	https://flatiron.com/
Enlitic	Provides deep learning using large-scale data sets from clinical tests for healthcare diagnosis	https://www.enlitic.com/
Digital Reasoning Systems	Provides cognitive computing services and data analytic solutions for processing and organizing unstructured data into meaningful data	https://digitalreasoning.com/
Ayasdi	Provides AI accommodated platform for clinical variations, population health, risk management and other healthcare analytics	https://www.ayasdi.com/
Linguamatics	Provides text mining platform for digging important information from unstructured healthcare data	https://www.linguamatics.com/
Apixio	Provides cognitive computing platform for analyzing clinical data and pdf health records to generate deep information	https://www.apixio.com/
Roam Analytics	Provides natural language processing infrastructure for modern healthcare systems	https://roanalytics.com/
Lumiata	Provides services for analytics and risk management for efficient outcomes in healthcare	https://www.lumiata.com
OptumHealth	Provides healthcare analytics, improve modern health system's infrastructure and comprehensive and innovative solutions for the healthcare industry	https://www.optum.com/

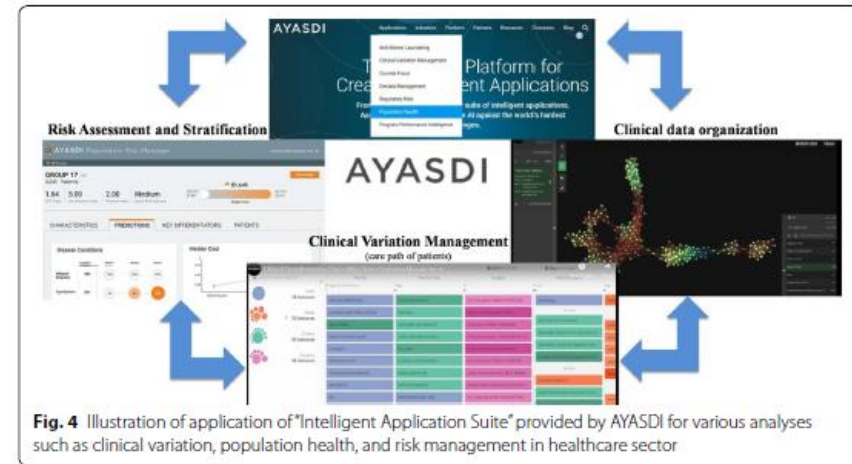


Fig. 4 Illustration of application of "Intelligent Application Suite" provided by AYASDI for various analyses such as clinical variation, population health, and risk management in healthcare sector

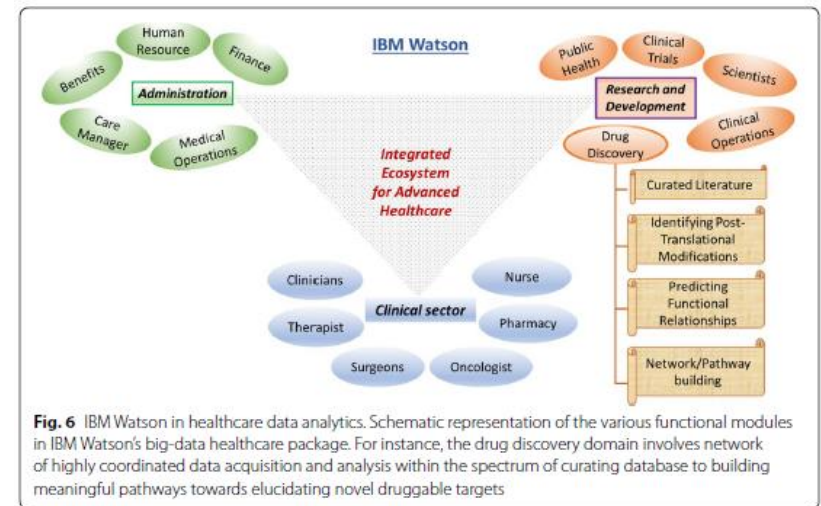


Fig. 6 IBM Watson in healthcare data analytics. Schematic representation of the various functional modules in IBM Watson's big-data healthcare package. For instance, the drug discovery domain involves network of highly coordinated data acquisition and analysis within the spectrum of curating database to building meaningful pathways towards elucidating novel druggable targets



Sfide associate ai big data sanitari

I metodi per la gestione e l'analisi dei big data sono in continuo sviluppo, soprattutto per quanto riguarda lo streaming dei dati in tempo reale, l'acquisizione, l'aggregazione, l'analisi e le soluzioni di visualizzazione che possono aiutare a integrare un migliore utilizzo degli EHR con l'assistenza sanitaria.

Ad esempio, il tasso di adozione dei programmi EHR testati e certificati a livello federale nel settore sanitario degli Stati Uniti è quasi completo. In Italia, possiamo migliorare...

La disponibilità di centinaia di prodotti EHR certificati, ognuno con terminologie cliniche, specifiche tecniche e capacità funzionali diverse, ha portato a difficoltà nell'interoperabilità e nella condivisione dei dati.

Tuttavia, possiamo affermare con certezza che il settore sanitario è entrato in una fase di implementazione "post-EHR". L'obiettivo principale è ora quello di ottenere informazioni utili da queste grandi quantità di dati raccolti con le EHR.

Reisman M. EHRs: the challenge of making electronic data usable and interoperable. Pharm Ther. 2017;42(9):572-5.



Sfide associate ai big data sanitari

- *Capacità di stoccaggio*
- L'archiviazione di grandi volumi di dati è una delle sfide principali e molte organizzazioni si sentono a proprio agio con l'archiviazione dei dati presso la propria sede. Questo comporta diversi vantaggi, come il controllo della sicurezza, dell'accesso e del tempo di attività. Tuttavia, una rete di server in loco può essere costosa e complessa da mantenere. L'archiviazione basata sul cloud è un'opzione migliore, per la quale la maggior parte delle organizzazioni sanitarie sta recentemente aderendo. Le organizzazioni devono scegliere partner cloud che comprendano l'importanza delle problematiche di conformità e sicurezza specifiche del settore sanitario. Inoltre, lo storage in cloud offre costi iniziali più bassi, un disaster recovery agile e un'espansione più semplice. Le organizzazioni possono anche adottare un approccio ibrido ai loro programmi di archiviazione dei dati, che può essere l'approccio più flessibile e praticabile per i fornitori con esigenze diverse di accesso e archiviazione dei dati.

Reisman M. EHRs: the challenge of making electronic data usable and interoperable. Pharm Ther. 2017;42(9):572–5.



Sfide associate ai big data sanitari

- *Pulizia*
- Dopo l'acquisizione, i dati devono essere purificati o ripuliti per garantirne l'accuratezza, la correttezza, la coerenza, la pertinenza e la purezza. Questo processo di pulizia può essere manuale o automatizzato utilizzando regole logiche per garantire alti livelli di accuratezza e integrità. Gli strumenti più sofisticati e precisi utilizzano tecniche di apprendimento automatico per ridurre i tempi e le spese e per impedire che i dati errati facciano deragliare i progetti Big Data.
- *Formati unificati*
- I pazienti producono un'enorme mole di dati che non è facile catturare con il formato EHR tradizionale, in quanto è complessa e non facilmente gestibile. È troppo difficile gestire i big data, soprattutto quando questi vengono forniti agli operatori sanitari senza una perfetta organizzazione dei dati. Pertanto, sono stati sviluppati sistemi di codifica medica come la Current Procedural Terminology (CPT) e la International Classification of Diseases (ICD) per rappresentare i concetti clinici fondamentali. Tuttavia, questi set di codici hanno i loro limiti.

Sfide associate ai big data sanitari

- *Accuratezza*
- Alcuni studi hanno osservato che il reporting dei dati dei pazienti nei sistemi elettronici non è ancora del tutto accurato, probabilmente a causa dello scarso utilizzo dell'EHR, dei flussi di lavoro complessi e della mancata della migliore acquisizione degli EHR per le scelte cliniche. Tutti questi fattori possono contribuire ai problemi di qualità dei big data lungo tutto il loro ciclo di vita. I sistemi EHR intendono migliorare la qualità e la comunicazione dei dati nei flussi di lavoro clinici, anche se i rapporti indicano discrepanze in questi contesti. La qualità della documentazione potrebbe migliorare utilizzando questionari auto-riportati dai pazienti per i loro sintomi.
- *Sicurezza*
- Le numerose violazioni della sicurezza, gli hacking, gli attacchi di phishing e gli episodi di ransomware hanno reso la sicurezza dei dati una priorità per le organizzazioni sanitarie. Dopo aver notato una serie di vulnerabilità, è stato elaborato un elenco di salvaguardie tecniche per le informazioni sanitarie protette (PHI). Queste regole, denominate Regole di sicurezza HIPAA, aiutano le organizzazioni a gestire l'archiviazione, la trasmissione, i protocolli di autenticazione e i controlli sull'accesso, l'integrità e la verifica. Misure di sicurezza comuni come l'utilizzo di software antivirus aggiornati, firewall, crittografia dei dati sensibili e autenticazione a più fattori possono evitare molti problemi.

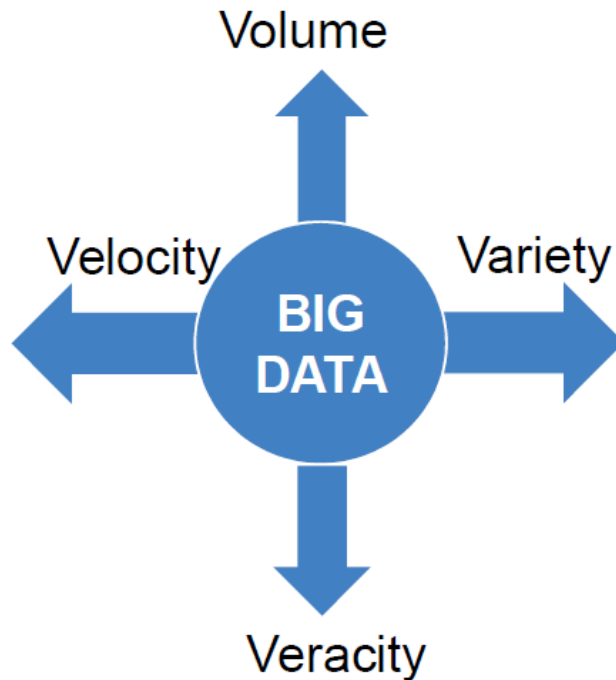
Sfide associate ai big data sanitari

- *Condivisione dei dati*
- I pazienti possono o meno ricevere le cure in più sedi. Nel primo caso, la condivisione dei dati con altre organizzazioni sanitarie sarebbe essenziale. Durante questa condivisione, se i dati non sono interoperabili, la circolazione dei dati tra organizzazioni diverse potrebbe essere fortemente limitata. Ciò potrebbe essere dovuto a barriere tecniche e organizzative. Ciò potrebbe lasciare i medici privi di informazioni chiave per prendere decisioni in merito ai follow-up e alle strategie di trattamento dei pazienti. Il più grande ostacolo alla condivisione dei dati è il trattamento dei dati come una merce che può fornire un vantaggio competitivo su vari aspetti.
- I fornitori di servizi sanitari dovranno superare tutte le sfide di questo elenco e altre ancora per sviluppare un ecosistema di scambio di big data che fornisca informazioni affidabili, tempestive e significative collegando tutti i membri del continuum di cura. Prima di superare queste sfide saranno necessari tempo, impegno, finanziamenti e comunicazione.

COSA SONO I BIG DATA NELLA SANITA' PUBBLICA?

What is Big Data

- Large and complex data sets which are difficult to process using traditional database technology.



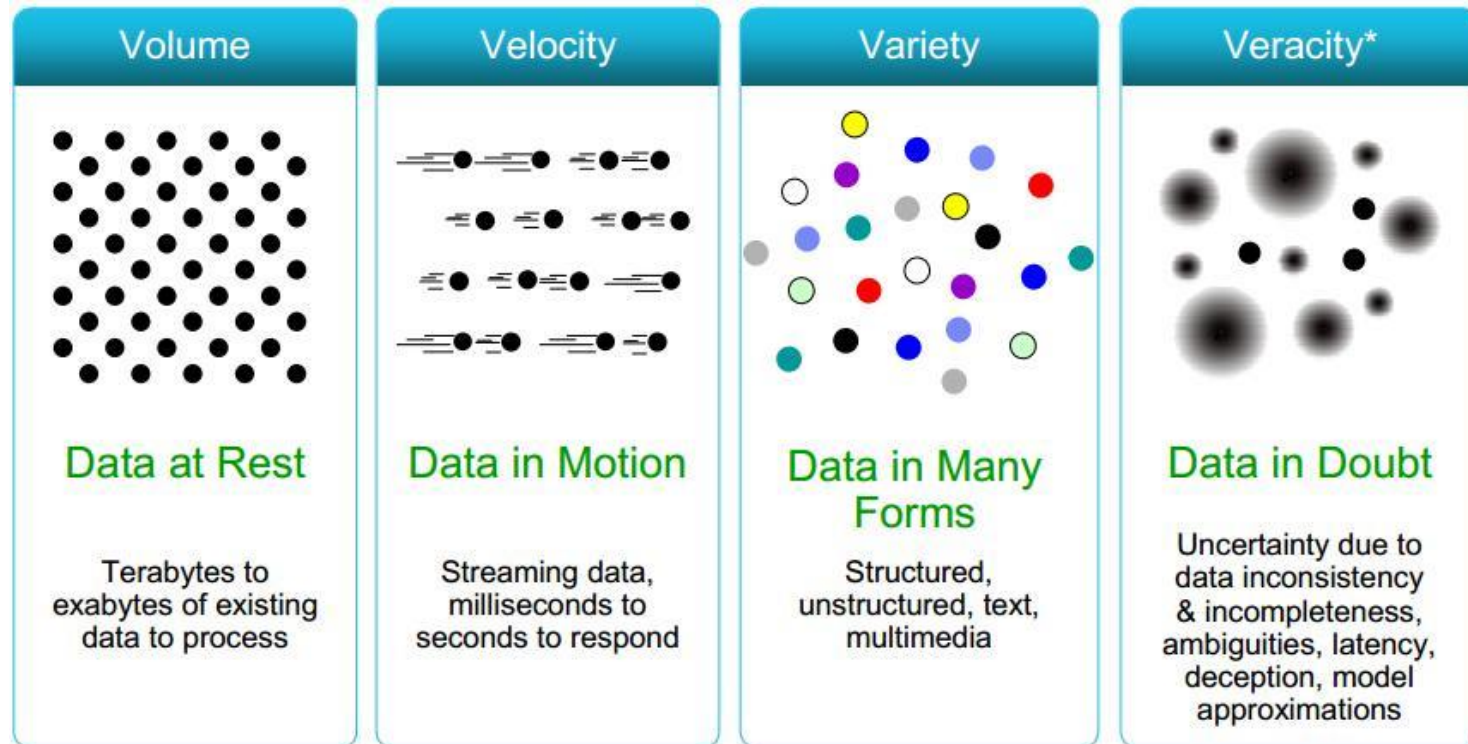
The four dimensions (V's) of Big Data

Big data is not just about size.

- Finds insights from complex, noisy, heterogeneous, longitudinal, and voluminous data.
- It aims to answer questions that were previously unanswered.



Alcuni parlano di 4V



Grandi dati o dati corretti?

Big Data or Right Data?

Ricardo Baeza-Yates

Yahoo! Labs Barcelona &
Web Research Group, DTIC, Univ. Pompeu Fabra
Barcelona, Spain
rbaeza@acm.org

Abstract. Big data nowadays is a fashionable topic, independently of what people mean when they use this term. The challenges include how to capture, transfer, store, clean, analyze, filter, search, share, and visualize such data. But being big is just a matter of volume, although there is no clear agreement in the size threshold where *big* starts. Indeed, it is easy to capture large amounts of data using a brute force approach. So the real goal should not be big data but to ask ourselves, for a given problem, what is the right data and how much of it is needed.¹ For some problems this would imply big data, but for the majority of the problems much less data is necessary. In this position paper we explore the trade-offs involved and the main problems that come with big data: scalability, redundancy, bias, noise, spam, and privacy.

Keywords: Scalability, redundancy, bias, sparsity, noise, spam, privacy.



Importanza dell'approccio critico per prendere decisioni

- La capacità di elaborare efficacemente enormi insiemi di dati è diventata parte integrante di un'ampia gamma di discipline scientifiche e accademiche. Tuttavia, ciò non rende superflua una profonda comprensione delle basi teoriche di un dominio. I big data consentono agli scienziati di superare i problemi associati ai piccoli campioni di dati in modi che includono l'allentamento delle assunzioni teoriche dei modelli, l'evitare l'overfitting dei modelli ai dati di training, una migliore gestione dei dati di training disturbati e la fornitura di ampi dati di test per convalidare i modelli.
- I problemi legati ai big data richiedono un compromesso tra scalabilità, disponibilità, prestazioni e sicurezza. Per alcuni problemi, le soluzioni precise sono talvolta impraticabili e possono richiedere algoritmi più veloci e approssimati che rischiano di ridurre la qualità della soluzione.



Importanza dell'approccio critico per prendere decisioni

- Naturalmente, quando si tratta di un argomento di tendenza, è d'obbligo un po' di cautela. La maggior parte dei problemi non ha bisogno di big data: ha bisogno dei dati giusti. Spesso i dati possono essere contraddittori, incompleti, imprecisi, soggettivi, ridondanti, parziali e confusi. Questi dati hanno il potenziale di creare confusione e disinformazione piuttosto che fornire spunti di riflessione utili. Dobbiamo evitare la tentazione di seguire un approccio basato sui dati anziché sui problemi. Come descritto in "Big Data o Right Data?", dobbiamo porre il giusto tipo di domande:
 - Come elaboriamo, filtriamo e campioniamo i dati di partenza per ottenere i dati giusti?
 - Come si determina l'attendibilità di tali dati?
 - Quanto rumore di fondo c'è?
 - Come distinguere i dati validi dallo spam? Filtrare lo spam è un problema non banale e una possibile fonte di distorsione per qualsiasi dato.
 - La distribuzione dei dati è valida o c'è un bias nascosto che deve essere corretto? Come possiamo correggere le distorsioni?
 - Come determinare ed eliminare i duplicati?



Medicina guidata dai dati

La ricerca biomedica rappresenta la base della moderna medicina basata sull'evidenza che, oggi più che mai, si basa sui dati.

Tuttavia, fino al 30% dei ricercatori ha assistito a colleghi che hanno alterato intenzionalmente i dati durante i loro studi.

Manipolazione significa "aggiustare, escludere, alterare o ricreare i dati in modo inappropriato".

OPEN ACCESS Freely available online



How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data

Daniele Fanelli*

INNOGEN and ISSTI-Institute for the Study of Science, Technology & Innovation, The University of Edinburgh, Edinburgh, United Kingdom

The screenshot shows the top navigation bar of the BMJ website with links for 'thebmj', 'covid-19', 'Research', 'Education', 'News & Views', 'Campaigns', and 'Jobs'. Below the navigation bar, the article title 'Scientific misconduct is worryingly prevalent in the UK, shows BMJ survey' is displayed. The article is from BMJ 2012;344, with a DOI of https://doi.org/10.1136/bmj.e3777, published on 12 January 2012. The citation is given as 'Cite this as: BMJ 2012;344:e3777'. There are tabs for 'Article', 'Related content', 'Metrics', and 'Responses'. The author is listed as 'Aniket Taware' with a dropdown for 'Author affiliations'. The abstract text reads: 'One in seven UK based scientists or doctors has witnessed colleagues intentionally altering or fabricating data during their research or for the purposes of publication, found a survey of more than 2700 researchers conducted by the BMJ. The survey, which was emailed to 9036 academics and clinicians who had submitted articles to the BMJ or acted as peer reviewers for the journal (response rate 31%), found that 13% of these researchers admitted knowledge of colleagues "inappropriately adjusting, excluding, altering, or fabricating data" for the purpose of publication. Just

È probabile che, se in media il 2% degli scienziati ammette di aver falsificato la ricerca almeno una volta e fino al 34% ammette altre pratiche di ricerca discutibili, le frequenze effettive di cattiva condotta potrebbero essere persino più elevate!



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Data e big data

nature

www.nature.com/nature

Vol 455 | Issue no. 7209 | 4 September 2008

Community cleverness required

Researchers need to adapt their institutions and practices in response to torrents of new data — and need to complement smart science with smart searching.

TECHNOLOGY FEATURE

THE BIG CHALLENGES OF BIG DATA

*As they grapple with increasingly large data sets,
biologists and computer scientists uncork new bottlenecks.*

Big hopes for big data

Large-scale multi-modal information on patients' health is ever increasing, providing an opportunity to use big data for taking individualized medicine to a global scale.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

1961 – 2019: Stealth research e Healthcare Unicorns



theranos

https://en.wikipedia.org/wiki/The_Inventor:_Out_for_Blood_in_Silicon_Valley

Received: 22 January 2019 | Accepted: 22 January 2019
DOI: 10.1111/med.13072



WILEY

PERSPECTIVE

Stealth research: Lack of peer-reviewed evidence from healthcare unicorns

Ioana A. Cristea^{1,2} | Eli M. Cahan^{3,4} | John P. A. Ioannidis^{1,5,6,7,8}

¹Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, Calif

²Department of Clinical Psychology and Psychotherapy, Babes-Bolyai University, Cluj-Napoca,

³New York University School of Medicine, New York, New York

⁴Department of Orthopaedic Surgery, Stanford University School of Medicine, Stanford, Califor

⁵Department of Medicine, Stanford University, Stanford, California

⁶Department of Health Research and Policy, Stanford University, Stanford, California

⁷Department of Biomedical Data Science, Stanford University, Stanford, California

⁸Department of Statistics, Stanford University, Stanford, California

Key messages

- Start-ups are widely accepted as key vehicles of innovation and disruption in healthcare, positioned to make revolutionary discoveries.
- Most of the highest-valued start-ups in healthcare have a limited or non-existent participation and impact in the publicly available scientific literature.
- The system of peer-reviewed publishing, while imperfect, is indispensable for validating innovative products and technologies in biomedicine.
- Healthcare products not subjected to peer-review but based on internal data generation alone may be problematic and non-trustworthy.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

La cattiva ricerca e le sue conseguenze



1998: Wakefield pubblica su The Lancet uno studio che riporta un legame tra la vaccinazione contro morbillo, parotite e rosolia e una sindrome di autismo e malattie intestinali in 12 bambini.

2010: lo studio viene ritrattato solo 12 anni dopo: "le affermazioni contenute nell'articolo originale, secondo le quali i bambini erano stati inviati consecutivamente e le indagini erano state approvate dal comitato etico locale, si sono rivelate false".

Tra il 1998 e il 2010, la frode di Wakefield ha causato un danno incommensurabile e duraturo, attraverso la diminuzione dei tassi di vaccinazione e la sfiducia generale nelle autorità sanitarie, sentimenti che perdurano ancora oggi, nonostante la pandemia.

Gli eventi recenti dimostrano che, anche dopo più di 20 anni, la ricerca scientifica è ancora soggetta a questo tipo di cattiva condotta, a spese della comunità scientifica e della nostra società nel suo complesso.

THE LANCET

The Lancet, Volume 375, Issue 9502, Pages 627-641, 28 February 2010
doi:10.1016/S0140-6736(09)11096-8

This article was retracted

RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children

Dr AJ Wakefield FRCS, Dr SH Ashwin MB BSc, A Anthony MB BSc, J Linnell PhD, S Zhi Cao MD MRCP, S H Shah MD MRCP, M Sordani MD FRCPsych, A P Dhillon MD FRCS, M A Thomson FRCP, P Harvey FRCP, & S Venter FRCS, Dr S Davies MRCPsych, Dr J Walker-Smith FRCP

Summary

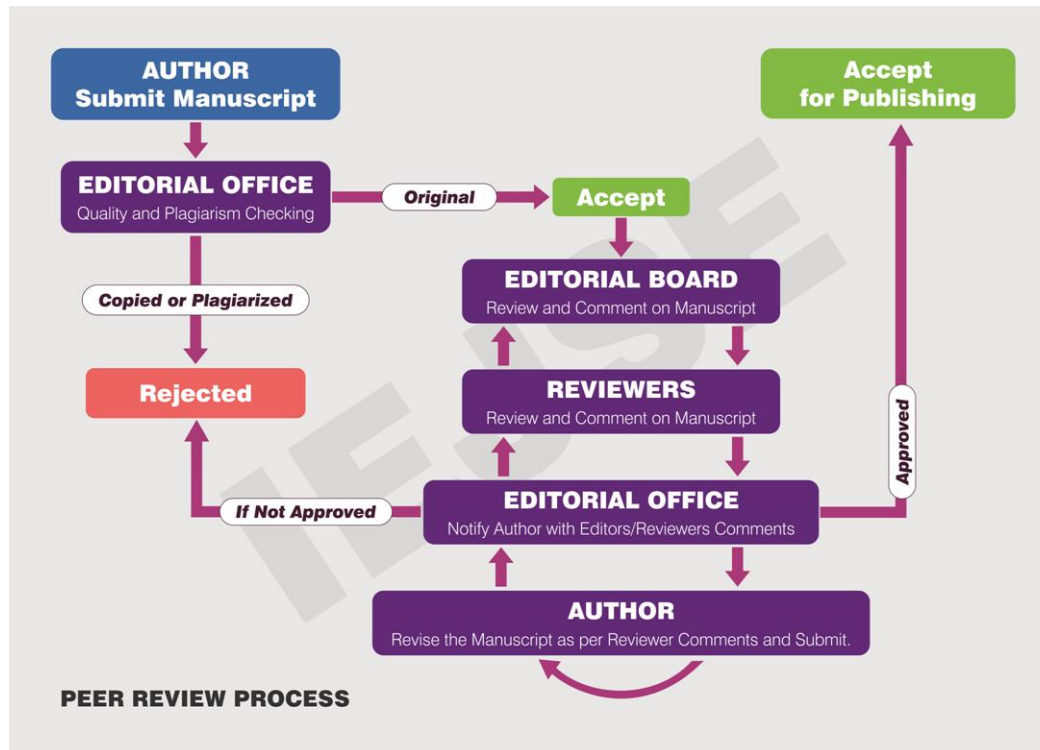
Background

We investigated a case of the ~~syndrome~~ ~~of~~ ~~children~~ ~~with~~ ~~ileal-lymphoid-nodular~~ ~~and~~ ~~non-specific~~ ~~colitis~~ ~~and~~ ~~pervasive~~ ~~developmental~~ ~~disorder~~ ~~in~~ ~~children~~.

Methods

11 children (mean age 5 years [range 2–10], 11 boys) were referred to a paediatric gastroenterology unit with a history of normal development followed by loss of acquired skills, including language, together with diarrhoea and abdominal pain. Children underwent gastroenterological, neurological, and developmental assessment and review of developmental records. Colonoscopy and biopsy sampling, magnetic resonance imaging (MRI), electroencephalography (EEG), and lumbar puncture were done under sedation. Barium follow-through radiography was done where possible. Biochemical, haematological, and immunological profiles were examined.

La Peer review nella ricerca scientifica



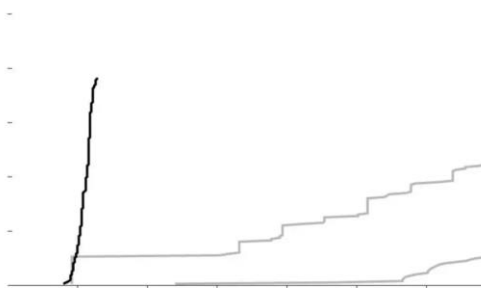
I rischi della «speed science»



Coronavirus and the risks of 'speed science'



Number of papers published

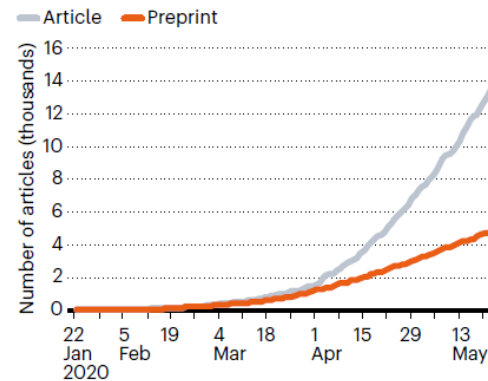


The numbers of papers published about COVID-19, SARS and MERS.

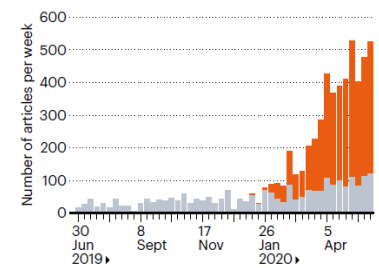
Image: Reuters Graphics

TORRENT OF PREPRINTS

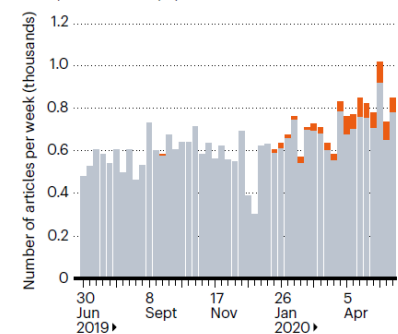
Around one-quarter of the scientific articles that relate to COVID-19 are preprints, by one estimate. Many appeared at medRxiv, a site for medical preprints, which has grown hugely as a result of the pandemic. Some were posted on bioRxiv and other preprint servers.



medRxiv submissions
Papers COVID papers



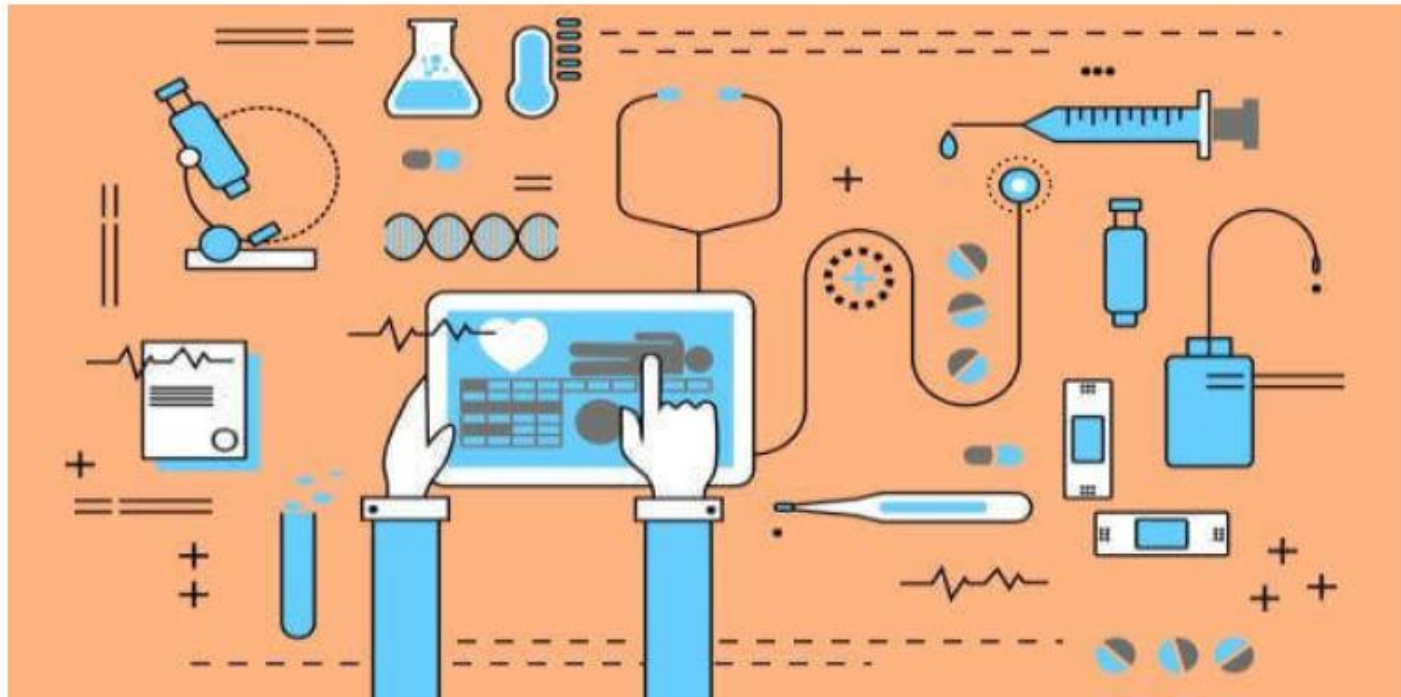
bioRxiv submissions
Papers COVID papers



24 MOTIVI PER CUI ABBIAMO BISOGNO DEI BIG DATA NELLA SANITA' PUBBLICA

24 Examples Of Big Data Analytics In Healthcare That Can Save People

By Bernardita Calzon in Business Intelligence, Jun 16th 2023



24 MOTIVI PER CUI ABBIAMO BISOGNO DEI BIG DATA NELLA SANITA' PUBBLICA

Big Data In Healthcare Applications

1. Improved patients predictions
2. Use Electronic Health Records (EHRs)
3. Real-time alerting for instant care
4. Enhance patient engagement
5. Prevent opioid abuse in the US
6. Informed strategic planning
7. Cure cancer with health data
8. Use predictive analytics in healthcare
9. Reduce fraud and enhance data security
10. Practice telemedicine
11. Integrate medical imaging
12. Prevent unnecessary ER visits
13. Smart staffing & personnel management
14. Learning & development
15. Advanced risk & disease control
16. Suicide and self-harm prevention
17. Improved supply chain management
18. Financial facility management
19. Develop new therapies & innovations
20. Manage & track mass diseases
21. Improve drug prescription processes
22. Prevent human error
23. Alerting heart issues with mobile devices
24. Bluetooth helps asthma patients



24 MOTIVI PER CUI ABBIAMO BISOGNO DEI BIG DATA NELLA SANITA' PUBBLICA

È l'applicazione più diffusa dei big data in medicina. Ogni persona ha una propria cartella clinica digitale, che comprende dati demografici, anamnesi, allergie, risultati di esami di laboratorio, ecc. **I record sono condivisi attraverso sistemi informatici sicuri e sono disponibili per i fornitori del settore pubblico e privato.** Ogni cartella è costituita da un unico file modificabile, il che significa che i medici possono apportare modifiche nel corso del tempo senza bisogno di documenti e senza il pericolo di replicare i dati.



I sistemi EHR possono anche attivare avvisi e promemoria quando un paziente deve sottoporsi a un nuovo esame di laboratorio o monitorare le prescrizioni per verificare se ha seguito gli ordini del medico.

Sebbene l'EHR sia un'ottima idea, **molti Paesi fanno ancora fatica a implementarlo completamente.** Gli Stati Uniti hanno fatto un grande passo avanti, con il 94% degli ospedali che adottano gli EHR secondo questa ricerca HITECH, ma l'UE è ancora indietro. Tuttavia un'ambiziosa direttiva elaborata dalla Commissione europea dovrebbe cambiare le cose.



Apptivvivo

24 MOTIVI PER CUI ABBIAMO BISOGNO DEI BIG

DATA NELLA SANITA' PUBBLICA

Altri esempi di analisi dei dati in ambito sanitario condividono una funzionalità cruciale: **l'avviso in tempo reale (real time alerting)**. Negli ospedali, i software di supporto alle decisioni cliniche (CDS) analizzano i dati medici in tempo reale, fornendo consigli agli operatori sanitari mentre prendono decisioni prescrittive. I dispositivi indossabili raccoglieranno continuamente i dati sanitari dei pazienti e li invieranno al cloud.

Inoltre, queste informazioni saranno accessibili al database sullo stato di salute della popolazione generale, che consentirà ai medici di confrontare questi dati in un contesto socio-economico e di modificare di conseguenza le strategie di erogazione. Le istituzioni e i gestori delle cure utilizzeranno strumenti sofisticati per monitorare questo enorme flusso di dati e reagire ogni volta che i risultati saranno preoccupanti.

Ad esempio, se la pressione sanguigna di un paziente aumenta in modo allarmante, il sistema invierà un avviso in tempo reale al medico, che si attiverà per raggiungere il paziente e somministrare misure per abbassare la pressione.

Un altro esempio è quello di Asthmapolis, che ha iniziato a utilizzare inalatori dotati di localizzatori GPS per identificare le tendenze dell'asma sia a livello individuale sia guardando a livello di salute pubblica. Questi dati vengono utilizzati insieme a quelli del CDC per sviluppare migliori piani di trattamento per gli asmatici nei diversi contesti.



24 MOTIVI PER CUI ABBIAMO BISOGNO DEI BIG DATA NELLA SANITA' PUBBLICA

Molti potenziali pazienti sono **già interessati a dispositivi smart** che registrano in modo permanente ogni passo compiuto, la frequenza cardiaca, le abitudini di sonno, ecc. **Tutte queste informazioni vitali possono essere abbinare ad altri dati tracciabili per identificare potenziali rischi per la salute in agguato.** L'insonnia cronica e la frequenza cardiaca elevata, ad esempio, possono segnalare il rischio di future malattie cardiache. I pazienti sono **direttamente coinvolti** nel monitoraggio della propria salute e **possono spingerli a condurre uno stile di vita sano.**

Un altro modo per farlo è rappresentato dai nuovi dispositivi indossabili in fase di sviluppo, che tracciano specifici trend di salute e li trasmettono al cloud dove i medici possono monitorarli. I pazienti che soffrono di diabete, asma o di ipertensione potrebbero trarne beneficio, diventare un po' più indipendenti e riducendo le visite inutili dal medico.



24 MOTIVI PER CUI ABBIAMO BISOGNO DEI BIG DATA NELLA SANITA' PUBBLICA

L'uso dei big data nell'assistenza sanitaria consente **una pianificazione strategica** grazie a una migliore comprensione delle motivazioni delle persone. I care manager possono analizzare i risultati dei controlli tra le persone appartenenti a diversi gruppi demografici e identificare quali fattori scoraggiano le persone dall'intraprendere le cure.

L'Università della Florida si è avvalsa di Google Maps e di dati gratuiti sulla salute pubblica per preparare heatmaps mirate a diversi temi, come la crescita della popolazione e le malattie croniche. Successivamente, gli studiosi hanno confrontato questi dati con la disponibilità di servizi medici nelle aree più «calde». I dati raccolti hanno permesso di rivedere la strategia di erogazione dei servizi e di aggiungere più unità di cura nelle aree più problematiche.



24 MOTIVI PER CUI ABBIAMO BISOGNO DEI BIG DATA NELLA SANITA' PUBBLICA

La telemedicina è presente sul mercato da oltre 40 anni, ma solo oggi è riuscita ad affermarsi pienamente. **Il termine si riferisce all'erogazione di servizi clinici a distanza utilizzando la tecnologia.**

Viene utilizzato per consultazioni primarie e diagnosi iniziali, per il monitoraggio remoto dei pazienti e per l'educazione medica degli operatori sanitari.

I medici utilizzano la telemedicina per fornire **piani di trattamento personalizzati e prevenire l'ospedalizzazione o la riammissione.**

Questo uso dell'analisi dei dati sanitari può essere collegato all'uso dell'analisi predittiva.

Tenendo i pazienti lontani dagli ospedali, la telemedicina contribuisce a ridurre i costi e a migliorare la qualità del servizio. I pazienti possono evitare le file e i medici non perdono tempo in consultazioni inutili.

La telemedicina migliora anche la disponibilità delle cure, poiché lo stato dei pazienti può essere monitorato e consultato ovunque e in qualsiasi momento.



24 MOTIVI PER CUI ABBIAMO BISOGNO DEI BIG DATA NELLA SANITA' PUBBLICA

Da quando è iniziata, all'inizio del 2020, la pandemia COVID-19 ha colpito milioni e milioni di persone in tutto il mondo. **La natura diffusa di questo virus ha rappresentato una sfida per il settore sanitario, che si è trovato a cercare di imparare da esso e di controllarlo allo stesso tempo.**

In questo contesto, **i big data hanno svolto un ruolo fondamentale nella risposta data a questa malattia** in crescita che ha fatto fermare il mondo per anni. Con il supporto di tecnologie avanzate di gestione dei dati, gli esperti sanitari sono stati in grado di monitorare in tempo reale la diffusione del COVID, la velocità di mutazione in condizioni diverse e l'effetto che stava avendo sulle diverse economie mondiali. Ciò è stato possibile analizzando enormi serie di dati provenienti da fonti diverse, come le cartelle cliniche e i comportamenti umani individuali.

Altre **strategie di prevenzione e preparedness** agli eventi catastrofici (ondate di calore, alluvioni) possono risultare indispensabili nel pensare la sanità pubblica del futuro.



GRAZIE!

