

ERICE 9-12 DICEMBRE 2025

LXIV Corso della Scuola Superiore di
Epidemiologia nel 62esimo anno
della Fondazione Majorana



ERICE LXIV

LXIV CORSO
Prevenzione,
preparedness
e disastri naturali:
quale ruolo
per la sanità pubblica?

RELATORE: Giordano d'Aloisio

Dichiarazione Autocertificazione Docente /Relatore

ASSENZA CONFLITTO D'INTERESSE

Il Sottoscritto Giordano d'Aloisio in qualità di relatore ai sensi *ai sensi dell'art. 76, comma 4 dell'Accordo Stato-Regioni del 2 febbraio 2017 e del paragrafo 4.5. del Manuale nazionale di accreditamento per l'erogazione di eventi ECM,*

dichiara

che negli ultimi due anni **NON** ha avuto rapporti anche di finanziamento con soggetti portatori di interessi commerciali in campo sanitario.

La Preparedness guidata dai Big Data

Giordano d'Aloisio

giordano.daloisio@univaq.it

LXIII Corso Prevenzione, Preparedness e Disastri Naturali: Quale Ruolo per la Sanità Pubblica?

9-12 Dicembre, Erice, Centro di Cultura Scientifica Ettore Majorana

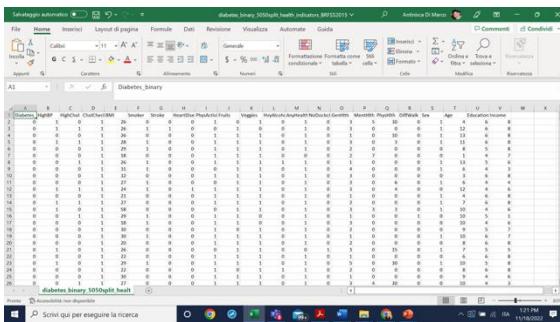
Indice

- Big Data
- Big Data per la Preparedness
- Conclusioni

BIG DATA

Dati: Dataset vs Database

- Un **dataset** è una collezione di record strutturati.
- Si può intendere come una fotografia di un sistema o fenomeno ad un certo istante di tempo.
- Un dataset può essere estratto da un database
- Un foglio Excel è un esempio di dataset
- Una **Base di Dati (Database)**, nella sua accezione più generica, è una collezione omogenea di dati strutturati e connessi tra loro
- Esistono una serie di programmi applicativi che consentono di effettuare su tali dati operazioni di accesso, modifica e interrogazione (*query* di ricerca).

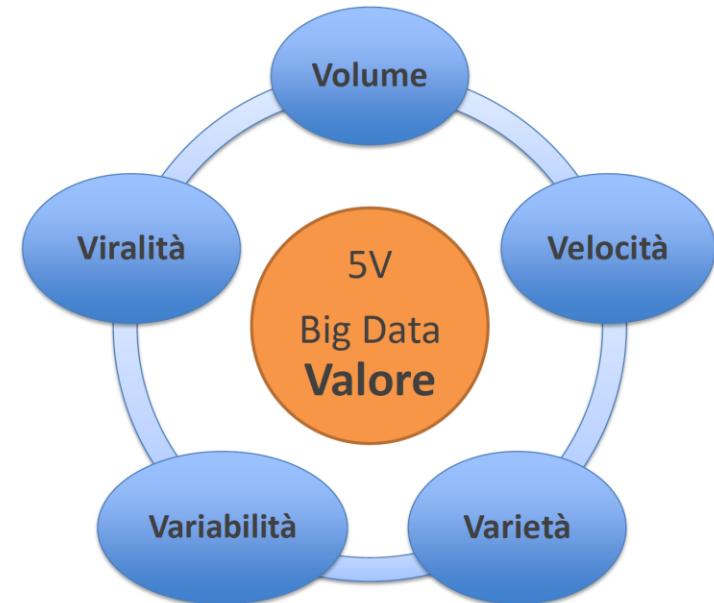


The screenshot shows a Microsoft Excel spreadsheet titled "diabetes_binary_30Stage_health". The first few columns are labeled: "Outcome", "Pregnancies", "Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI", "DiabetesPedigreeFunction", "Age", and "Outcome". The "Outcome" column contains binary values (0 or 1). The "Pregnancies" column ranges from 0 to 14. The "Glucose" column ranges from 0 to 199. The "BloodPressure" column ranges from 0 to 130. The "SkinThickness" column ranges from 0 to 9. The "Insulin" column has values like 0, 70, 160, etc. The "BMI" column ranges from 0 to 67. The "DiabetesPedigreeFunction" column ranges from 0.07 to 4.42. The "Age" column ranges from 21 to 81. The last column, "Outcome", also contains binary values (0 or 1). The entire dataset consists of 76 columns and approximately 760 rows of data.



Big Data

- I Big Data sono dati che superano i limiti degli strumenti tradizionali.
- *Sono dati solitamente disponibili in grandi volumi, che si presentano in differenti formati (spesso privi di struttura) e con caratteristiche eterogenee, prodotti e diffusi generalmente con una elevata frequenza, e che cambiano spesso nel tempo.*
- Per questo motivo sono identificati con le **5V (+1)**.



Big Data in Ambito Sanitario

- L'ambito sanitario è uno dei domini più influenzati dai big data
- Secondo The European House – Ambrosetti, nel 2025 il comparto sanitario risulta la prima fonte di Big Data, contribuendo a circa il 30% del volume totale
- Esempi di utilizzo dei big data in ambito sanitario sono:
 - Creazioni di basi di dati cliniche sui pazienti (e.g., [**CancerLinQ**](#))
 - Digitalizzazione dei processi ospedalieri (*Ospedali Digitali*), in grado di monitorare tutti i dati dell'ospedale
 - Processi predittivi (e.g., Machine Learning) in grado di predire diagnosi di un paziente

Sorgenti di Big Data in Ambito Sanitario

Cartelle Cliniche Elettroniche (CCE)

- Formati digitali delle normali cartelle cliniche

Dispositivi connessi alla rete

- Qualsiasi dispositivo che effettui un monitoraggio di un paziente ed è connesso ad una rete internet

Banche dati genomiche

- Banche dati che contengono sequenziamenti genomici di vario tipo

Immagini Digitali

- Immagini come lastre o TAC in formato digitale

VOLUME

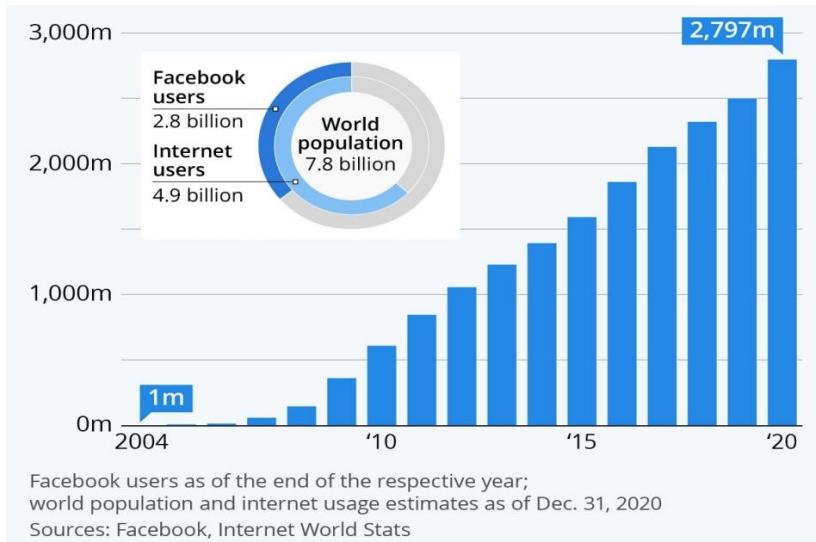
Quanto sono grandi questi dati?

Volume

In un 1 minuto:

- **100 mila tweet** trasmessi nel mondo.
- **35 mila "Like" FB** a siti ufficiali di organizzazioni.
- **160 milioni (circa) di email** inviate.
- **2 mila check-in su 4square** effettuati.

Ciò va aggiunto alle restanti **"attività digitali"**, generando una enorme mole di dati e informazioni a loro volta incrociabili.



Volume



- Il **primo passo** quando si opera con i Big Data è allora l'**immagazzinamento**. L'**analisi (e la pulizia)** avvengono in una fase successiva (per evitare di perdere potenziali informazioni).
- Ciò richiede **importanti investimenti in termini di storage e di capacità di calcolo** adatta all'analisi di grandi moli di dati.
- Tecnologia open source più diffusa e utilizzata: **Apache Hadoop**.

VELOCITÀ

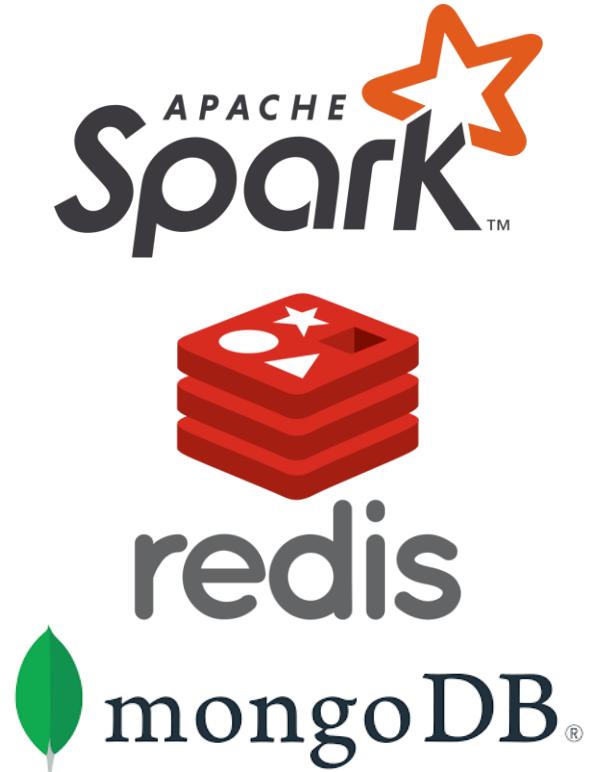
Con quale frequenza vengono prodotti e come possiamo processarli?

Velocità

- **Velocità:** è una caratteristica che ha più di un significato.
- Si riferisce in primis alla **elevata frequenza con cui i dati vengono generati** – si ripercuote sulla quantità (Volume).
- Il secondo aspetto riguarda la **velocità con cui le nuove tecnologie** permettono di **accedere e di analizzare questi dati**.
 - **Maggiore è la velocità** di accesso ai dati
 - **Maggiore sarà la velocità** in un processo decisionale
 - **Maggiore/migliore competitività** sui diversi panorami del mercato

Velocità

- Particolarmente adatte sono le **architetture distribuite** (e.g., Hadoop).
- **Gestione di strutture dati anche complesse.**
- **Accesso ai dati in tempo reale** (e.g., Redis).
- **Velocità di elaborazione** grazie a tecniche di calcolo distribuito (e.g., Apache Spark).
- Database non relazionali come i ***column DB*** e ***key/value DB*** (NoSQL).



VARIETÀ

Con quale forma si presentano questi dati?

Varietà

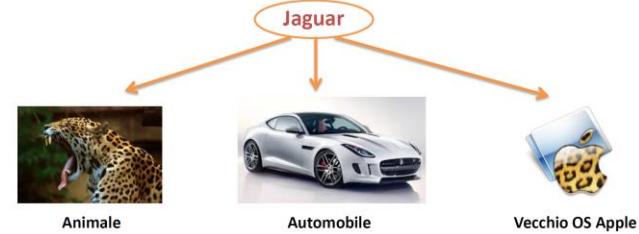
- **Varietà:** caratteristica che ha a che fare con la forma in cui i dati si presentano.
- Nel contesto **Big Data** le **informazioni** da trattare sono dati non-strutturati (o semi-strutturati). Non adatti ad essere lavorati con le tecniche tradizionali dei database relazionali.
- Dati come **email, immagini, video, audio, stringhe di testo** a cui dare un significato non si possono memorizzare in una tabella.
- Per la gestione e il salvataggio di questi dati si ricorre spesso ai **database NoSQL**. Non impongono uno schema rigido per organizzare i dati (*schema-less database*).

VARIABILITÀ

Cosa significano questi dati?

Variabilità

- **Variabilità:** caratteristica relativa alla **contestualizzazione** di un dato.
- Il significato o l'**interpretazione di uno stesso dato** può **variare** in base al contesto in cui esso viene raccolto e analizzato.
- Il **significato** di un dato **può essere differente** anche in base al **momento in cui viene fatta l'analisi**, spesso è fondamentale l'analisi in tempo reale (Velocità).
- E' importante trovare dei meccanismi che riescano a dare una **semantica ai dati** in base al contesto in cui sono espressi.



VIRALITÀ

Come si diffondono questi dati?

Viralità

- **Viralità:** caratteristica che ha a che fare su **quanto e come i dati si diffondono** (Propagazione dei dati).
- La **grande quantità** di dati (spesso correlati tra loro) e l'**alta velocità con cui sono prodotti** implica una **diffusione virale** delle informazioni.
- **Esempio:** una notizia o un evento diffusi tra diversi canali. Diffusione amplificata con i collegamenti nei vari **social network**.
- **Istituzioni e alcune organizzazioni** sfruttano questa caratteristica/potenzialità per migliore attività di pronto intervento (e.g., Twitter).



VALORE

Che valore hanno i Big Data?

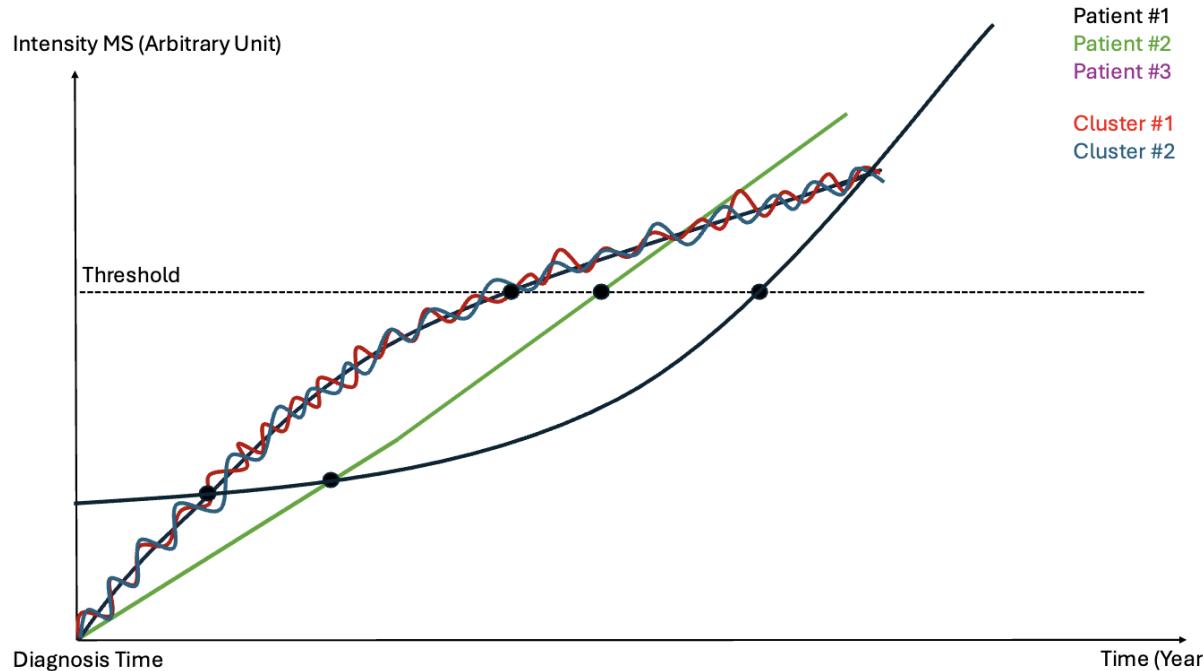
Valore

- **Valore:** è necessario comprendere e gestire in modo adeguato i dati e tutti questi aspetti ad essi legati in modo da riuscire ad **estrarre il potenziale informativo**.
- I Big Data **nascondono un grande valore**. Al primo utilizzo di solito se ne estrae soltanto una parte, il *valore rimanente rimane “dormiente”* fino ad un successivo utilizzo.
- E' quindi importante adottare metodologie e tecnologie che permettano la **continua integrazione di nuove informazioni**, in seguito ad un utilizzo reiterato, con l'obiettivo di **costruire una base di conoscenza sempre più ampia**.

Valore dei Big Data in Ambito Sanitario

- Secondo la International Data Corporation i big data in ambito sanitario sono i dati in maggiore crescita rispetto agli altri settori (e.g., finanziario o media)
- Numerosi paesi e strutture stanno investendo in cartelle cliniche elettroniche, strumenti di prevenzione e sistemi di sensoristica per aumentare la mole di dati a disposizione

Esempio



D'Aloisio, G., Di Matteo, A., Cipriani, A., Lozzi, D., Mattei, E., Zanfardino, G., ... & Placidi, G. (2024, September). Engineering a Digital Twin for Diagnosis and Treatment of Multiple Sclerosis. In *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems* (pp. 364-369).

CRITICITÀ

Da grandi poteri derivano grandi responsabilità

Criticità e rischi dei Big Data

I Big Data offrono grandi prospettive e potenzialità, ma non presentano esclusivamente caratteristiche positive. Vi sono alcuni **aspetti critici** che è bene prendere in considerazione:

- Problematiche legate alla **qualità e all'affidabilità dei dati**.
- Problematiche relative alla **privacy e alla proprietà dei dati**.

Qualità dei dati

La **qualità dei dati** è determinata da un insieme di caratteristiche:

- **Completezza:** la presenza di **tutte le informazioni necessarie** a descrivere un oggetto, entità o evento. Ad esempio, una cartella clinica deve contenere tutti i dati essenziali del paziente.
- **Consistenza:** i dati **non devono essere in contraddizione**. Ad esempio i valori di un esame di laboratorio devono essere compatibili con quelli riportati in precedenti controlli.
- **Accuratezza:** i dati devono essere corretti, cioè **conformi a dei valori reali**. Ad esempio la registrazione della pressione arteriosa deve riportare misurazioni effettivamente rilevate e non valori stimati.

Qualità dei dati

- **Assenza di duplicazione:** I dati devono essere memorizzati **una sola volta**, evitando la presenza di copie.
- **Assenza di bias:** I dati devono essere privi di discriminazione. Ad esempio, non posso avere informazioni solo per pazienti di sesso maschile.

NB: *La qualità dei dati è però legata anche al contesto in cui essi sono analizzati. Operazioni di filtraggio e pulizia devono essere fatte procedendo per gradi per evitare di eliminare dati potenzialmente utili*

Privacy nei Big Data

- Il tema Big Data si apre a problemi di Privacy, proprietà e utilizzo dei dati da parte di terzi.
- **Dati sensibili:** i dati, soprattutto in ambito sanitario, devono essere specificamente protetti per evitare che informazioni sensibili (tutte le informazioni che possano portare all'identificazione di un paziente) siano diffusi all'esterno

ATTACCO ALLA SANITÀ

**Il data breach dell'ASL Abruzzo e il
risarcimento del danno da violazione
del GDPR**

BIG DATA PER LA PREPAREDNESS

31

Applicazioni dei Big Data

Disaster Medicine and Public Health Preparedness

www.cambridge.org/dmp

Systematic Review

Cite this article: Lu S, Christie GA, Nguyen TT, Freeman JD, Hsu EB (2022) Applications of artificial intelligence and machine learning in disasters and public health emergencies.

Disaster Med Public Health Prep 16: 1674–1681

Machine Learning in Disaster Management: Recent Developments in Methods and Applications

[Vasileios Linardos¹](#), [Maria Drakaki^{2,*}](#), [Panagiotis Tzionas³](#) and [Yannis L. Karnavas⁴](#)

¹ Archeiothiki S.A., GR-19400 Athens, Greece

² Department of Science and Technology, University Center of International Programmes of Studies, International Hellenic University, 14th Km Thessaloniki-N. Moudania, GR-57001 Thermi, Greece

³ Department of Industrial Engineering and Management, International Hellenic University, P.O. Box 141, GR-57400 Thessaloniki, Greece

⁴ Electrical Machines Laboratory, Department of Electrical & Computer Engineering, Democritus University of Thrace, GR-67100 Xanthi, Greece

Show more ▾

Applications of Artificial Intelligence and Machine Learning in Disasters and Public Health Emergencies

Sally Lu¹, Gordon A. Christie PhD², Thanh T. Nguyen PhD³, Jeffrey D. Freeman PhD,
MPH⁴ and Edbert B. Hsu MD, MPH¹

Artificial intelligence applications in healthcare supply chain networks under disaster conditions

[Vikas Kumar, Fariba Goodarzian, Peiman Ghasemi, Felix T. S. Chan & Narain](#)

Cite this article: Vikas Kumar, Fariba Goodarzian, Peiman Ghasemi, Felix T. S. Chan & Narain (2024) Artificial intelligence applications in healthcare supply chain networks under disaster conditions, *International Journal of Production Research*, 63:2, 395–403, DOI: [07543.2024.2444150](https://doi.org/10.1080/00207543.2024.2444150)

Read this article: <https://doi.org/10.1080/00207543.2024.2444150>

Applicazioni dei Big Data

- Sistemi preventivi:
 - Sistemi di sorveglianza epidemiologica in tempo reale
- Pianificazione delle risorse:
 - Modelli pianificare il consumo di risorse
- Analisi del rischio e pianificazione delle vulnerabilità:
 - Utilizzo di Digital Twin per analisi territoriali
- Modelli predittivi:
 - Addestramento di modelli di intelligenza artificiale predittivi

Digital Twin

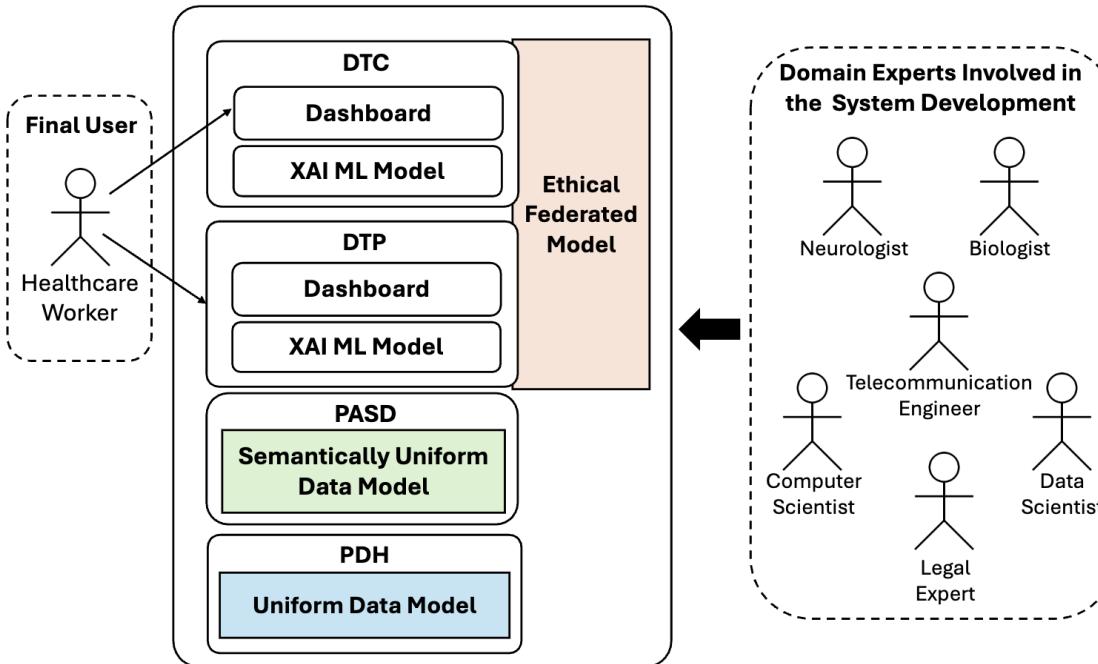
- I Digital Twin sono dei «Gemelli Digitali» creati tramite i dati
- «*A digital twin is a dynamic virtual replica of a physical system, process, or asset that uses real-time data, simulation, and analytics to mirror, predict, and optimize its real-world performance.*»

Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2017). Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11), 1016-1022. <https://doi.org/10.1016/j.ifacol.2018.08.474>

Digital Twin in Ambito Sanitario

- Nel contesto sanitario, i DT sono stati utilizzati:
 - per la manutenzione predittiva dei dispositivi medici e per la loro ottimizzazione in termini di velocità e consumo
 - l'ottimizzazione del ciclo di vita ospedaliero sviluppando piattaforme di analisi predittive, applicazioni AI sulla trasformazione di dati di grandi dimensioni di pazienti in informazioni fruibili.
 - Per rappresentare organi (ad esempio il cuore) o parti del corpo umano (ad esempio il sistema delle vie aeree)

Digital Twin in Ambito Sanitario



D'Aloisio, G., Di Matteo, A., Cipriani, A., Lozzi, D., Mattei, E., Zanfardino, G., ... & Placidi, G. (2024, September). Engineering a Digital Twin for Diagnosis and Treatment of Multiple Sclerosis. In *Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems* (pp. 364-369).

CONCLUSIONI

Signature



Conclusioni

- Una moltitudine di dati disponibile per natura, sorgente, semantica
- Questi dati, se correttamente manipolati, contengono un grande **valore**
- Diverse applicazioni dei Big Data per la preparedness in ambito sanitario
- **Competenze di analisi necessarie** per tradurre le domande in modelli e processi e per selezionare i dati

NB: *Aspetti etici dello scopo delle analisi, dell'uso dei dati e comunicazione dei risultati*