

Sommario

1. Caratteristiche dei Big Data	3
a. Le 5 V e alcune caratteristiche dei Big Data	3
Volume	3
Velocità	3
Transitorietà	3
Varietà delle fonti	3
Eterogeneità	4
b. Gli ambiti e i contesti di utilizzo in ambito sanitario e della PA	4
Verso l'utilizzo di dati non strutturati	5
Perché utilizzare dati esterni all'azienda	5
Esempi di utilizzo	6
2. I dati di laboratorio	7
Cosa è l'informatizzazione del laboratorio	7
a. Caratteristiche intrinseche	8
b. Caratteristiche estrinseche	8
3. Le relazioni tra Big Data e dati di Laboratorio: "Big DataLIS"	8
a. Analogie e differenze delle caratteristiche intrinseche ed estrinseche	8
b. Analogia e differenze nelle potenzialità di utilizzo	8

Hanno collaborato

Tommaso Trenti, Azienda Ospedaliera Universitaria e Azienda USL di Modena, t.trenti@ausl.mo.it

Andrea Padoan, Azienda Ospedale-Università di Padova, andrea.padoan@unipd.it

Roberto Guerranti, Università degli Studi di Siena, roberto.guerranti@unisi.it

Aldo Clerico, Scuola Superiore Sant'Anna, clerico@ftgm.it

Debora Angeletti, Azienda USL di Modena, d.angeletti@asul.mo.it,

Tommaso Pirotti, Azienda USL di Modena, t.pirotti@ausl.mo.it

Matteo Petracca, Engineering Ingegneria Informatica spa, Matteo.Petracca@eng.it

Federico Cinelli, Engineering Ingegneria Informatica spa, Federico.Cinelli@eng.it

Giovanna Carissimi, Engineering Ingegneria Informatica spa, giovanna.carissimi@eng.it

Natalia Pianesi, Engineering Ingegneria Informatica spa, natalia.pianesi@eng.it

Silvia Civolani, Dedalus, silvia.civolani@dedalus.eu

Mauro Muccioli, Dedalus, mauro.muccioli@dedalus.eu

Giancarlo Giaccone Giancarlo.Giaccone@abbott.com

Valentina Pecoraro Ausl di Modena valepecoraro@gmail.com

Laura Roli Ausl di Modena l.rolis@ausl.mo.it

Daria Debbia Ausl di Modena daria.debbia@gmail.com

1. Caratteristiche dei Big Data

a. Le 5 V e alcune caratteristiche dei Big Data

Volume

- Dati generati automaticamente da macchine (sensori, Strumenti scientifici....)
- I contenuti degli utenti (Web2.0) ovvero Social network, Blog, Wiki, social bookmarking, Podcasting.....
- Transazioni bancarie e movimenti sui mercati finanziari

Velocità

Sono necessarie tecnologie per l'acquisizione dei dati "ad alta velocità"

Le tecnologie a cui si fa riferimento sono chiamate streaming data o complex event processing (CEP)

Consentono di monitorare più fonti dati e analizzarli in modo incrementale a bassissima latenza

La velocità riguarda anche la capacità di sfruttarli con rapidità, individuando le informazioni utili per il business e minimizzando i tempi di elaborazione

Adozione di nuove tecnologie di database non relazionali (es. colonnari, database NoSQL)

Transitorietà

I dati generati da sensori, log dei web server sono persi se non sono salvati immediatamente

Documenti e pagine web possono essere rimossi

E' NECESSARIO IMMAGAZZINARE TUTTI I DATI ORIGINALI, INDIPENDENTEMENTE DAL LORO IMMEDIATO UTILIZZO

OPERAZIONI DI PULIZIA E SCARTO POSSONO SIGNIFICARE ELIMINAZIONE DI INFORMAZIONI UTILI IN FUTURO

Abbandonare i Db relazionali (RDBMS) per altre tecnologie (es Apache Hadoop)

Architetture distribuite per incrementare le performance di calcolo e la velocità di accesso ai dati

Varietà delle fonti

Diversità dei formati

Assenza di struttura: documenti, blog post, commenti sui social network

Varietà di fonti: generati automaticamente da macchine, generati dagli utenti

VARIETÀ DELLE FONTI



Eterogeneità

- Impossibilità di predefinire uno schema/tracciato in fase progettuale
- Necessità di scegliere i cosiddetti database NoSql o schemaless database
- Maggiore adattamento alla variabilità dei dati
- Applicazione di metodologie agile development, ovvero implementazione per piccole iterazioni successive sulla base dei requisiti di progetto

b. Gli ambiti e i contesti di utilizzo in ambito sanitario e della PA

L'evoluzione delle tecnologie digitali ha introdotto nuovi comportamenti e nuovi paradigmi di interazione e fruizione delle informazioni. La storia dell'evoluzione tecnologica, gli strumenti a cui pensiamo quando pensiamo a innovazione digitale e salute (sensori e dispositivi wireless, mobile, social network, genomica, internet, analisi evoluta delle immagini, data management, etc) hanno profondamente cambiato l'approccio al consumo di informazioni, sia dentro le aziende che fuori. Sono caduti i muri divisorii tra i ruoli e spesso anche tra le competenze. Questo si verifica anche nella gestione dell'informazione e dei servizi sanitari dove si è passati da una gestione monodirezionale medico – paziente dell'evoluzione dalla malattia, alla condivisione della diagnosi (apertura dei referti all'esterno nel caso del laboratorio), dalla gestione alla predizione (applicazione di nuove tecnologie di data analytics sul dato di laboratorio, ricerca con strumenti avanzati), fino alla prevenzione (percorsi di inclusione del cittadino).

Il progresso dei **#sensori** [wearable] verso la precisione, così come la durata della batteria e l'adozione di massa a partire dal 2009 al presente

Gli utenti **#mobile** hanno superato il numero di utenti del desktop già intorno al 2014

Il **#genoma** umano da 1,000 dollari annunciato da Illumina a metà del 2014 che ci ha permesso di scoprire il **“noi non sappiamo ciò che non sappiamo ancora”** (si prevede che sia da \$ 100 “presto”)

On-demand e potenza di elaborazione scalabile (**#cloud**)

#Intelligenza Artificiale, mentre il termine è stato inventato nel 1956, “secondo Jack Clark di Bloomberg, il 2015 è stato un anno cruciale per l’intelligenza artificiale, con il numero di progetti software che utilizzano AI all’interno soltanto di Google, passando da un “utilizzo sporadico” nel 2012 a più di 2.700 progetti” & Vinod Khosla che ha affermato con audacia che la tecnologia (AI) potrebbe sostituire l’80% di quello che fanno i medici (2012).

I **dati biometrici, l’alimentazione e lo stile di vita** assumono un ruolo decisivo per migliorare l’efficacia delle terapie, coinvolgendo anche l’aspetto sociale del paziente.

La **Medicina della persona** (o Medicina personalizzata) rappresenta il sistema di cura nell’era digitale ed è anche una possibile soluzione alle nuove problematiche del prossimo futuro, come **l’aumento dell’aspettativa di vita, la multi-morbilità, la cronicità delle malattie ed il sostenimento dei costi dell’intero sistema sanitario**

Verso l’utilizzo di dati non strutturati

I dati semi e destrutturati aumentano tanto più l’Azienda è interessata alla valutazione di dati provenienti dal mondo esterno:

- Documenti di testo (word, excel, Power point, equivalenti prodotti da software free, PDF)
- Immagini in vari formati
- E-mail
- Strumenti del web 2.0
- Blog , forum, wiki, post e commenti sui Social (Facebook, LinkedIn)
- Piattaforma di micro blogging (Twitter)
- Pagine web

dobbiamo avere la consapevolezza che non sempre le relazioni tra le informazioni che troviamo fuori e noi stessi sono relazioni forti o più semplicemente banali, anche concetti debolmente connessi con noi possono incrementare di molto l’immagine complessiva che possiamo disegnare di noi e del nostro futuro

Perché utilizzare dati esterni all’azienda

- Identificazione di minacce alla sicurezza, basate sull’analisi delle attività in rete

- Sistemi di lettura del testo e salvataggio delle informazioni attraverso una rappresentazione semantica
- Ricerca biologica e ambientale
- Tecniche di calcolo su dataset di enormi dimensioni
- Progetti legati a diverse aree della fisica nucleare
- Controllo e prevenzione delle malattie
- Ricerca sul cancro
- Neuroscienze (Neuroscience Information Framework)
- Dati GIS (Geographic Information System): l'analisi dei dati in un contesto geografico può far emergere pattern basati sulla dislocazione sul territorio
- RFID (Radio Frequency Identification): logistica del farmaco, logistica dei percorsi, controllo delle condizioni ambientali durante un trasporto
- CRM (Customer Relationship Management): analisi di grandi moli di dati per conoscere comportamenti, abitudini, opinioni dell'utenza al fine raggiungerla con campagne, iniziative e servizi sempre più personalizzati

Esempi di utilizzo

Autodiagnosi del paziente: grazie a dispositivi indossabili e basati su app per monitorare la salute ed eseguire ecografie, i pazienti avranno finalmente la possibilità di auto-diagnosticarsi da casa un ampio numero di condizioni di salute, senza doversi recare in ambulatorio e in ospedale.

Ospedale automatizzato: il check-in ospedaliero includerà una tecnologia di imaging in grado di valutare la frequenza cardiaca, la temperatura e la frequenza respiratoria sin dall'ingresso, i sensori potranno rilevare la pressione sanguigna ed effettuare un elettrocardiogramma entro 10 secondi, generando un triage automatico o addirittura una diagnosi al momento.

Gli operatori sanitari raddoppiano il loro tempo libero: medici e infermieri, che al momento impiegano fino al 70% del loro tempo in lavori amministrativi, potranno analizzare rapidamente ecografie o cartelle cliniche con un dispositivo mobile risparmiando tempo utile per concentrarsi sulla cura del paziente.

Archivio di dati digitali: i dispositivi si integreranno con le cartelle cliniche digitali, aggiornando automaticamente le condizioni e le terapie, fornendo agli infermieri dati più completi, in tempo reale, facilmente accessibili per prendere decisioni migliori.

Consenso AI: non appena l'intelligenza artificiale (AI) inizierà a giocare un ruolo cruciale nella diagnosi e nelle terapie, il sostegno pubblico crescerà nella misura in cui saremo favorevoli a farci rilasciare diagnosi da una macchina, purchè i servizi siano progettati e implementati intorno ai pazienti, i benefici siano spiegati e sia rilasciata l'autorizzazione.

L'Azienda USL è oggi impegnata nel rafforzamento del ruolo pionieristico di assistenza sanitaria partecipata e di **Health Population Management**

La **multicronicità**, la gestione **multi professionale**, il ruolo attivo dei pazienti grazie al potenziamento dei **canali di comunicazione virtuali (e non)**, sono i pilastri che ispirano i principali progetti in corso

Oggi più che mai c'è un bisogno diffuso di trasferire alla popolazione e a tutti gli attori coinvolti nel processo socio-sanitario **il valore della cura sul territorio**

2. I dati di laboratorio

Cosa è l'informatizzazione del laboratorio

Il principale produttore di dati strutturati aziendale: l'unica fonte, a parte quelle introdotte dai social network e da internet ad avere alcune delle caratteristiche tipiche dei Big data

La fonte informativa sulla base della quale sono prese la maggior parte delle decisioni cliniche

Uno dei principali centri di costo aziendali, generatore di costi indotti quindi primo centro nevralgico di controllo dell'appropriatezza clinica e prescrittiva: il laboratorio è stato il primo a sviluppare sistemi informatici per la valutazione dell'appropriatezza prescrittiva e di erogabilità delle prestazioni

La principale interfaccia dell'Azienda verso la medicina del territorio e i cittadini: il referto di laboratorio è stato il primo ad essere inviato al MMG e sul FSE

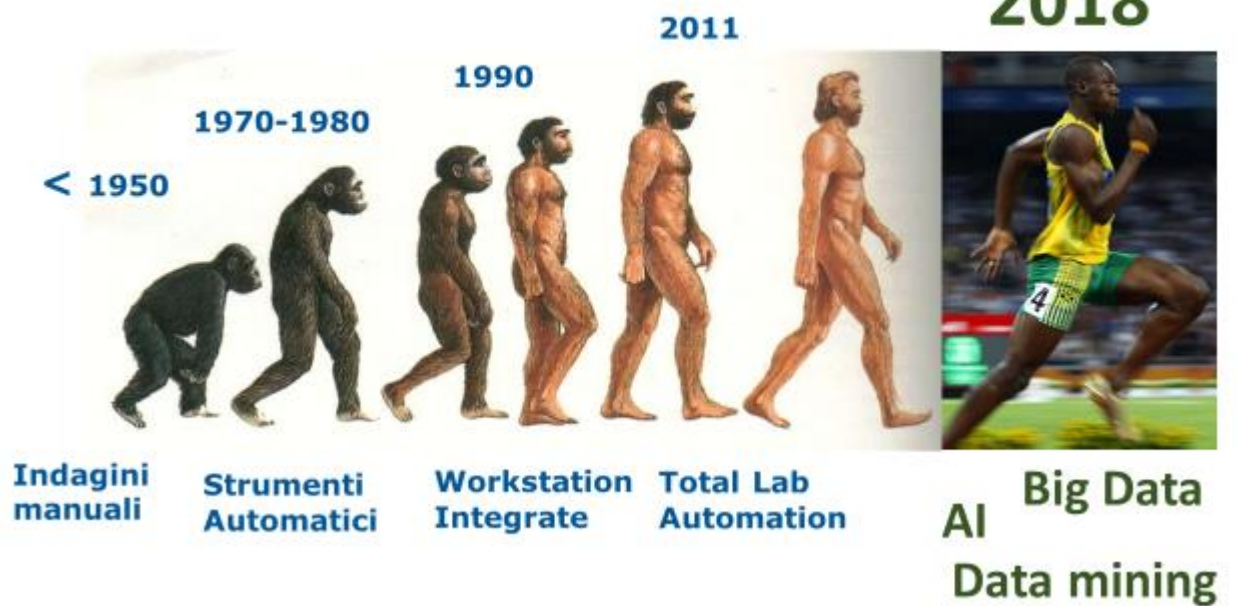
Il centro dove sono applicate le più avanzate tecnologie diagnostiche e di data mining

Il dato di laboratorio ha un ruolo di primaria importanza nei progetti di Digita Health in quanto offre un mare per lo più inesplorato di dati che possono essere analizzati e finalizzati alla conoscenza della popolazione e alla personalizzazione dei servizi.

Introducendo come fonte i dati prodotti nell'ambito del laboratorio analisi introduciamo due tipologie di fonti dati:

- Fonti dati operazionali che fanno riferimento all'applicativo gestionale di laboratorio con cui viene gestita l'attività operativa giornaliera (LIS) caratterizzati dall'elevato volume di dati prodotti
- Fonti dati che fanno riferimento ai dati prodotti dalle apparecchiature diagnostiche del laboratorio caratterizzati dal volume, dalla velocità e dalla transitorietà di produzione dei dati

L'informatizzazione del laboratorio



- a. Caratteristiche intrinseche
- b. Caratteristiche estrinseche

3. Le relazioni tra Big Data e dati di Laboratorio: "Big DataLIS"

- a. Analogie e differenze delle caratteristiche intrinseche ed estrinseche
- b. Analogia e differenze nelle potenzialità di utilizzo