



# Big Data in ambito biomedico

Venerdì 27 ottobre 2017

ore 17:00 | 20:00

Ordine degli Ingegneri della Provincia di Pavia

Via Indipendenza, 11

## Relatori

**Ing. Lucia Sacchi, Ing. Mauro Bucalo, Ing. Andrea Demartini**

Secondo la definizione adottata dal gruppo di lavoro promosso dall'International Medical Informatics Association (IMIA) su "Data Mining e Big Data Analytics", i big data sono dati la cui dimensione, eterogeneità, e variabilità sono tali da richiedere tecnologie e algoritmi del tutto nuovi, sia per la loro gestione sia per la loro analisi. È ormai ampiamente riconosciuto che i big data nascono dalla presenza congiunta di almeno quattro caratteristiche, note come le quattro V: volume, varietà, velocità e veridicità. In primo luogo, i big data possono essere caratterizzati dal loro volume, cioè per dimensioni tali da non essere mai state finora affrontate (in particolare ambito biomedico/clinico). Si prevede che la quantità totale di dati sanitari, oggi valutata pari a molto oltre 150 exabyte ( $=10^{18}$  byte), potrebbe esplodere a breve con l'utilizzo dei sequenziatori di nuova generazione per scopi diagnostici, e quando saranno integrati i dati provenienti dai dispositivi wireless di monitoraggio della salute. Il volume, tuttavia, non è l'unica proprietà dei big data: la caratteristica che rende peculiari i dati biomedici riguarda la loro estrema varietà o eterogeneità. I dati memorizzati, infatti, includono testi strutturati e non strutturati (note e referti), immagini, segnali, valori numerici. Questa varietà implica una notevole complessità dal punto di vista dell'architettura tecnologica e dell'interpretazione e dell'analisi dei dati. La terza caratteristica dei big data è la velocità di elaborazione e analisi, velocità che deve essere in linea con la necessità dell'applicativo, secondi o millisecondi per processare dati in real time, minuti o ore per le analisi in post processing. L'implicazione, dal punto di vista dell'analisi dei dati, è che l'elaborazione passi da un paradigma centralizzato ed effettuato offline, a una strategia distribuita e online. L'ultima delle 4 V è rappresentata da un elemento cruciale in campo biomedico, la veridicità (o al contrario l'incertezza dei dati). La veridicità è una problematica cruciale quando vengono combinate fonti eterogenee, di affidabilità variabile, come dati provenienti dalle cartelle cliniche, dalla letteratura, dai social networks. Questi dati sono molto diversi da quelli che si ottengono dagli studi clinici controllati.

**Coordinatore: Commissione Ingegneria dell'Informazione. Alessandro Filingeri**

È possibile partecipare al corso previa iscrizione sul portale [www.isinformazione.it](http://www.isinformazione.it) fino ad esaurimento posti. Costo di iscrizione: € 10

Per la partecipazione all'incontro verranno riconosciuti 3 CFP