

FOReSiGHT



SAPIENZA
UNIVERSITÀ DI ROMA

Scienza dei dati

Contorno

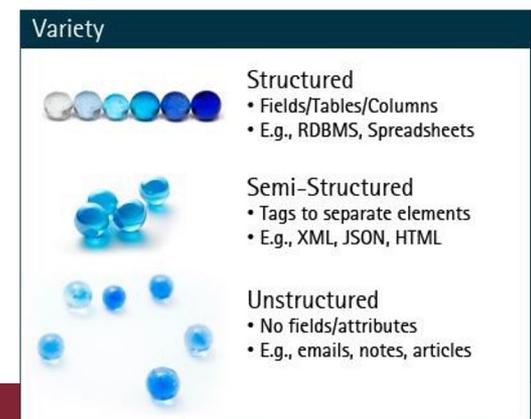
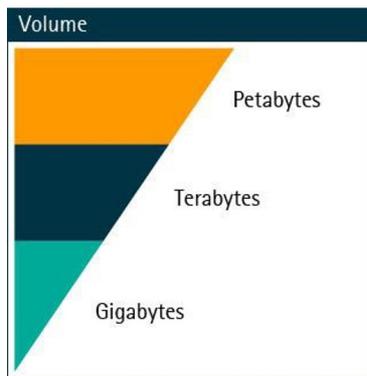
- **Big Data e sfide**
- **Cos'è la scienza dei dati**
- **Scienza dei dati e mondo accademico**
- **Scienza dei dati e altri**
- **Casi studio**
- **Punti essenziali**
- **Conclusione**

Dati ovunque

- **Molti dati vengono raccolti e archiviati**
 - Esperimenti scientifici
 - Internet delle cose
 - Dati web, commercio elettronico
 - Transazioni finanziarie, transazioni bancarie/di credito
 - Commercio e acquisti online
 - Rete sociale
 -molti altri!

Big Data

- I Big Data sono insiemi di dati così grandi o così complessi che i metodi tradizionali di archiviazione, accesso e analisi della loro suddivisione sono troppo costosi. Tuttavia, c'è molto valore potenziale nascosto in questi dati, quindi le organizzazioni sono ansiose di sfruttarlo per promuovere l'innovazione e il vantaggio competitivo.
- Le tecnologie e gli approcci dei Big Data vengono utilizzati per ricavare valore da ambienti ricchi di dati in modi che gli strumenti e i metodi di analisi tradizionali non possono fare.



Cosa fare con questi dati?

- Aggregazione e statistica
 - Archiviazione dati e OLAP
- Indicizzazione, ricerca e query
 - Ricerca basata su parole chiave
 - Corrispondenza dei modelli (XML/RDF)
- Scoperta della conoscenza
 - Estrazione dei dati
 - Modellazione statistica
- Guidato dai dati
 - Analisi predittiva
 - Apprendimento approfondito

Big Data e scienza dei dati

- "... il lavoro più interessante nei prossimi 10 anni sarà quello degli statistici", Hal Varian, capo economista di Google
- Si prevede che l'occupazione dei data scientist crescerà del 36% dal 2021 al 2031, molto più velocemente della media di tutte le occupazioni
<https://www.bls.gov/ooh/math/data-scientists.htm>
- Nuovi corsi di laurea, corsi, boot-camp:
 - ad esempio, a Berkeley: Statistiche, I-School, CS, Astronomia...
 - Una proposta (altrove) per un Master in "Big Data Science"
 - Piani per il flusso di data science presso AUST
 - RDA-CODATA Scuola di Ricerca in Scienza dei Dati

Cos'è la scienza dei dati?

- Un'area che gestisce, manipola, estrae e interpreta la conoscenza da un'enorme quantità di dati
- La scienza dei dati (DS) è un campo di studio multidisciplinare con l'obiettivo di affrontare le sfide dei big data
- I principi della scienza dei dati si applicano a tutti i dati, grandi e piccoli

Cos'è la scienza dei dati?

- Teorie e tecniche provenienti da molti campi e discipline vengono utilizzate per indagare e analizzare una grande quantità di dati per aiutare i decisori in molti settori come scienza, ingegneria, economia, politica, finanza e istruzione
 - Informatica
 - Riconoscimento di modelli, visualizzazione, data warehousing, calcolo ad alte prestazioni, database, intelligenza artificiale
 - Matematica
 - Modellazione matematica
 - Statistiche
 - Modelli statistici e stocastici, Probabilità.

Definizioni

...[DS include] matematica, statistica, ingegneria dei dati, riconoscimento e apprendimento di modelli, informatica avanzata, visualizzazione, modellazione dell'incertezza, data warehousing e calcolo ad alte prestazioni con l'obiettivo di estrarre significato dai dati e creare prodotti di dati

M
FUORI

Il campo della scienza dei dati sta emergendo all'intersezione tra i campi delle scienze sociali e statistiche, dell'informazione e dell'informatica e del design

BERKELEY S COOL OF
INFORMAZIONE _

INTERDISCIPLINARE _

NUOVI TIPI
DEI DATI _

Data
Science

I DATI COME
PRODOTTO

NUOVI METODI PER
DARE SENSO AI DATI

Estrazione di conoscenza dal grande volumi di dati strutturati o non strutturato, che è una continuazione del data mining sul campo e dell'analisi predittiva, noto anche come knowledge discovery e data mining (KDD). I "dati non strutturati" possono includere e-mail, video, foto, social media e altri contenuti generati dagli utenti.

WIKIPEDIA

Innanzitutto, la materia prima, la parte "dati" della Data Science, è sempre più eterogeneo e non strutturato. In secondo luogo, i computer interpretano i dati automaticamente, rendendoli agenti attivi nel processo della creazione di senso.

DHAR

...usare semplicemente i dati non è proprio ciò che intendiamo con " **scienza dei dati** ". Un'applicazione dati acquisisce il suo valore dai dati stessi e di conseguenza crea più dati. Non è solo un'applicazione con dati; è un prodotto di dati.

La scienza dei dati consente la creazione di prodotti di dati

LOUKADIS (O'REILLY M EDIA)

La scienza dei dati è lo studio di da dove provengono le informazioni, cosa rappresentano e come possono essere trasformate in una risorsa preziosa nella creazione di business e Strategie informatiche

ROUSE

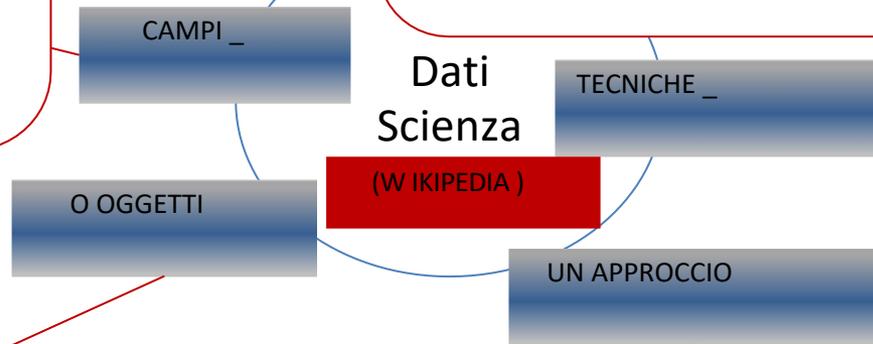
Fondamentalmente, la scienza dei dati prevede l'utilizzo di metodi automatizzati per analizzare enormi quantità di dati ed estrarre conoscenza da loro.

UNIVERSITÀ
DI

Il panorama della scienza dei dati

- Nanotecnologie
- Fisica
- Robotica
- Matematica
- Statistiche
- Teoria dell'informazione
- Tecnologie dell'informazione
- AI

- Elaborazione del segnale
- Modelli di probabilità
- Apprendimento automatico
- Apprendimento statistico
- Estrazione dei dati
- Banca dati
- Ingegneria dei dati
- Riconoscimento di modelli
- Visualizzazione
- Analisi predittiva
- Modellazione dell'incertezza
- Archiviazione dei dati
- Compressione dati
- Programmazione computer
- Calcolo ad alte prestazioni



I metodi adattabili ai Big Data sono di particolare interesse per la scienza dei dati, sebbene la disciplina non sia generalmente considerata limitata a tali dati.

Lo sviluppo dell'apprendimento automatico, un ramo dell'intelligenza artificiale utilizzato per scoprire modelli nei dati da cui è possibile sviluppare modelli predittivi, ha rafforzato la crescita e l'importanza della scienza dei dati.

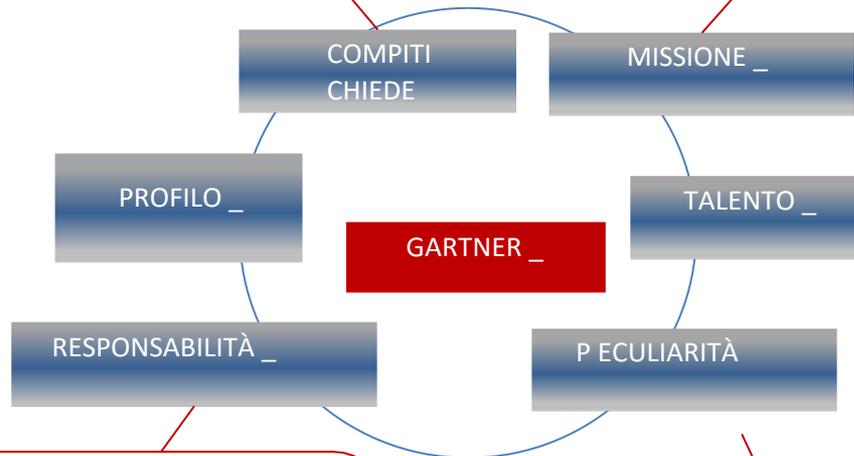
Chi è un Data Scientist?

Oltre alle capacità analitiche avanzate, questa persona è anche abile nell'integrazione e nella preparazione di set di dati ampi e diversi, nell'architettura di database specializzati e ambienti informatici e nella comunicazione dei risultati

Uno scienziato dei dati può o meno avere conoscenze specializzate del settore per aiutare nella modellazione dei problemi aziendali e nella comprensione e preparazione dei dati.

Il data scientist è emerso un nuovo ruolo, distinto da... ma

quelli degli affari analisti e statistici dell'intelligence (BI).



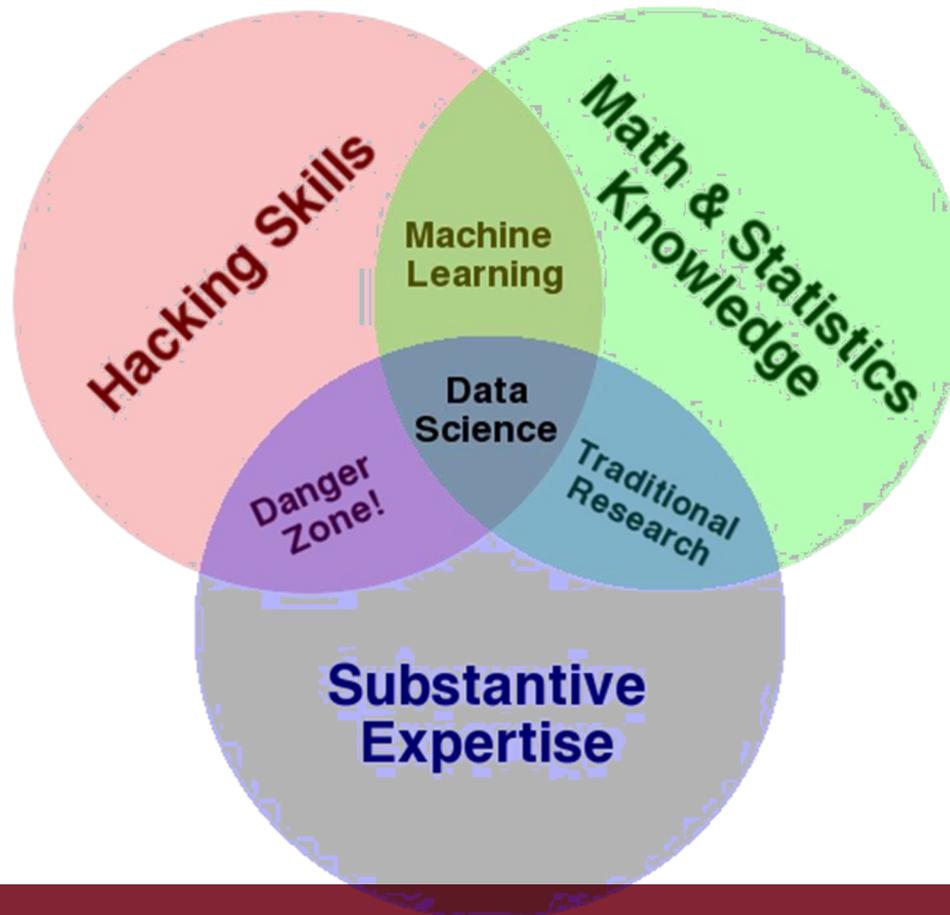
Creare valore dai dati richiede una vasta gamma di talenti: dall'integrazione e preparazione dei dati, all'architettura di ambienti informatici/database specializzati, al data mining e agli algoritmi intelligenti

Un individuo responsabile della modellazione di problemi aziendali complessi, della scoperta di approfondimenti aziendali e dell'identificazione di opportunità attraverso l'uso di statistiche, algoritmiche, di mining e di visualizzazione

I data scientist possono essere preziosi nel generare approfondimenti, soprattutto dai "big data"; ma la loro combinazione unica di competenze tecniche e commerciali, insieme alla loro crescente domanda, li rende difficili da trovare o coltivare.

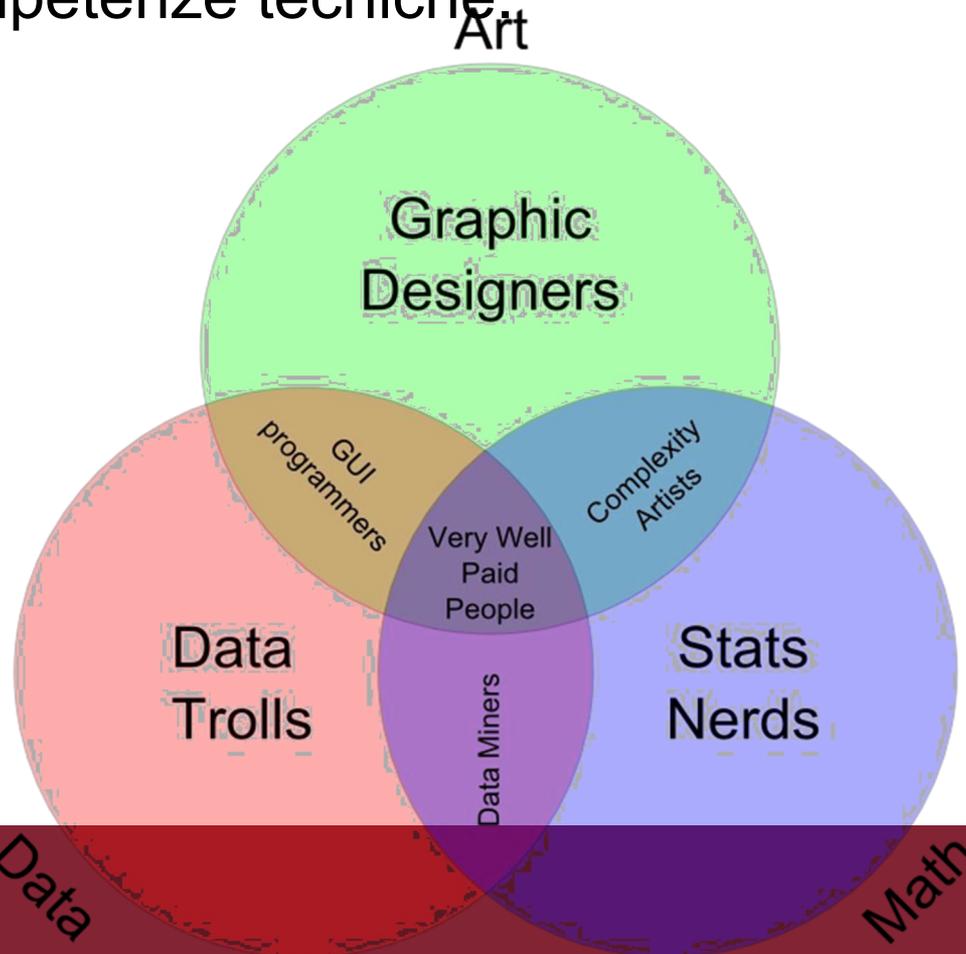
Cos'è la scienza dei dati?

- Alcune definizioni collegano competenze computazionali, statistiche e sostanziali.

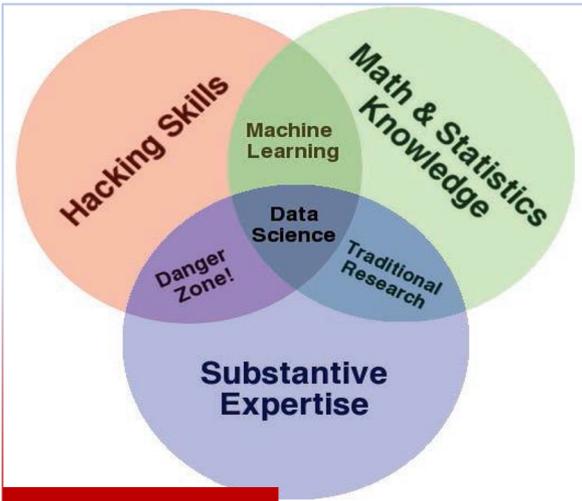


Cos'è la scienza dei dati?

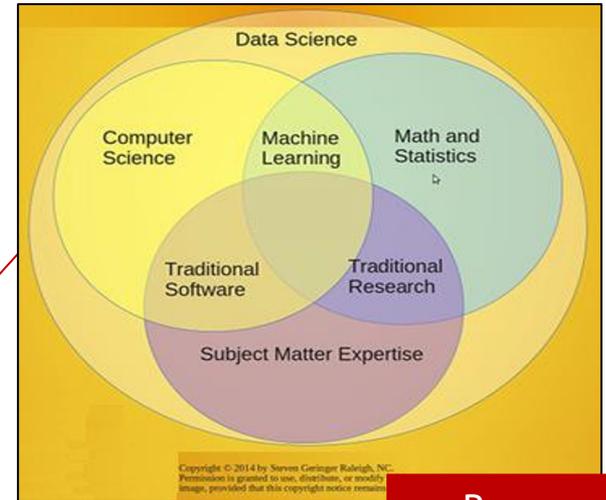
- Altre definizioni si concentrano maggiormente sulle sole competenze tecniche



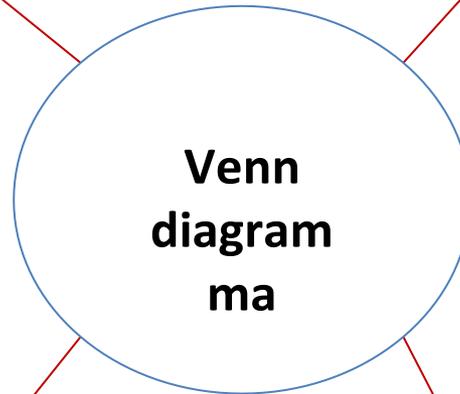
Unicorno



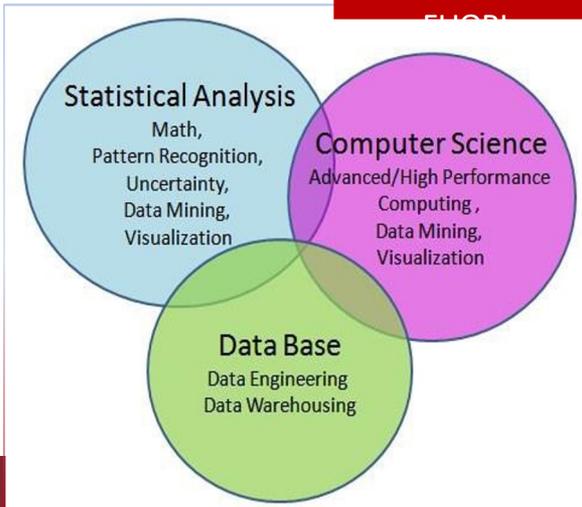
C ONVIA



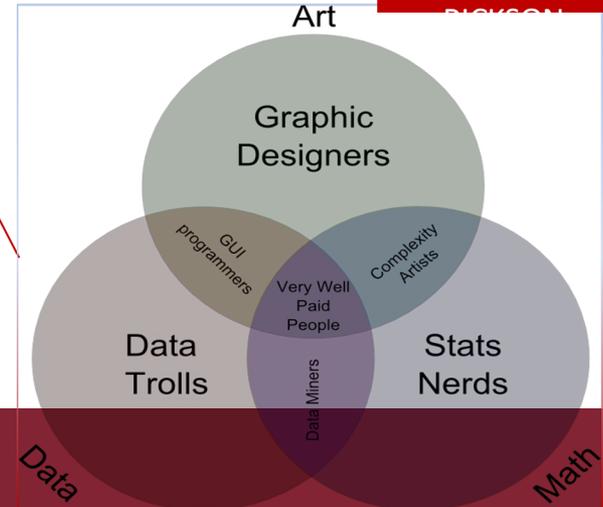
R



M



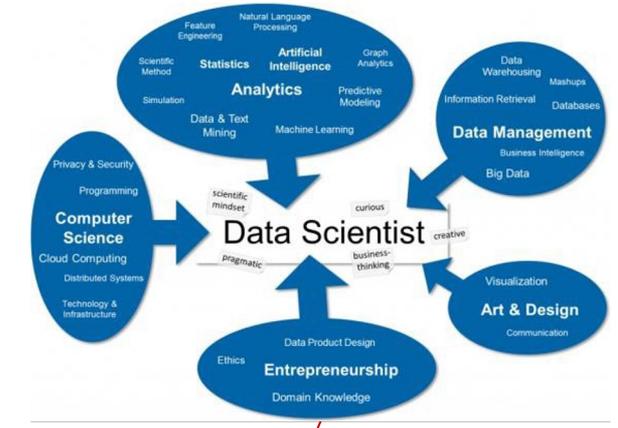
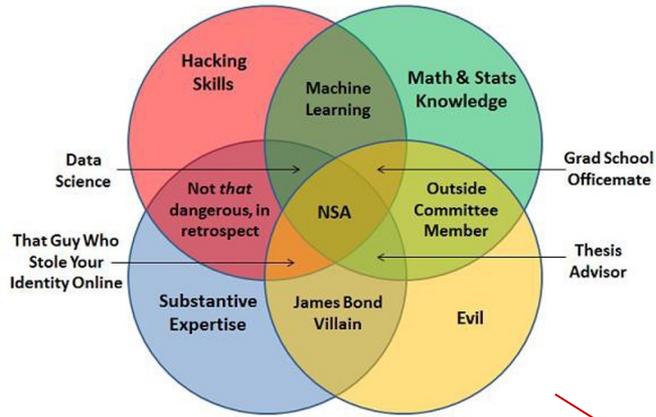
Art



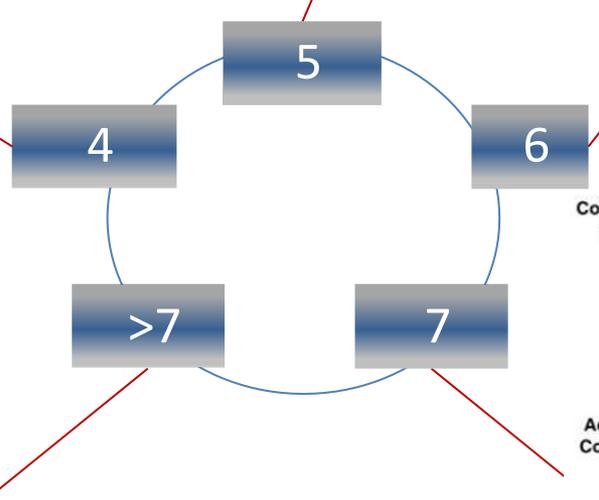
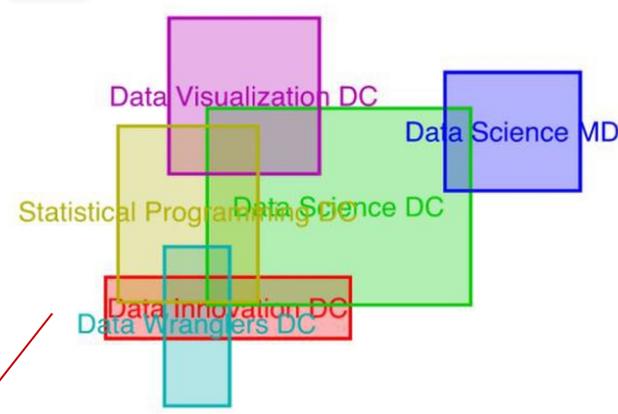
E

Data

Math



<https://youtu.be/5Zg-C8AAIGg>



Data Science Is Multidisciplinary

By Brendan Tierney, 2012

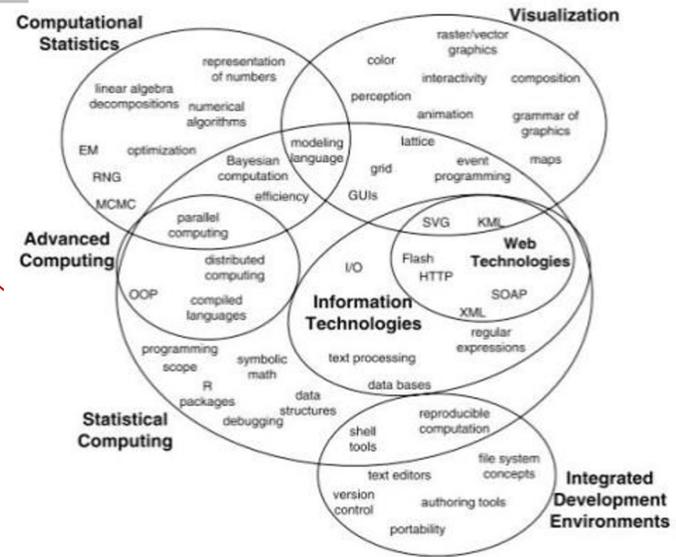
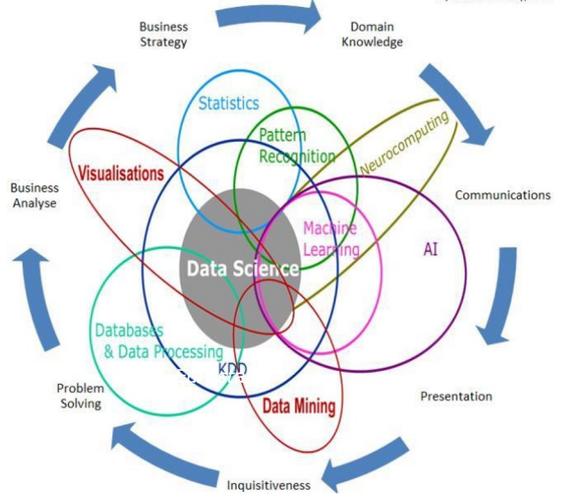
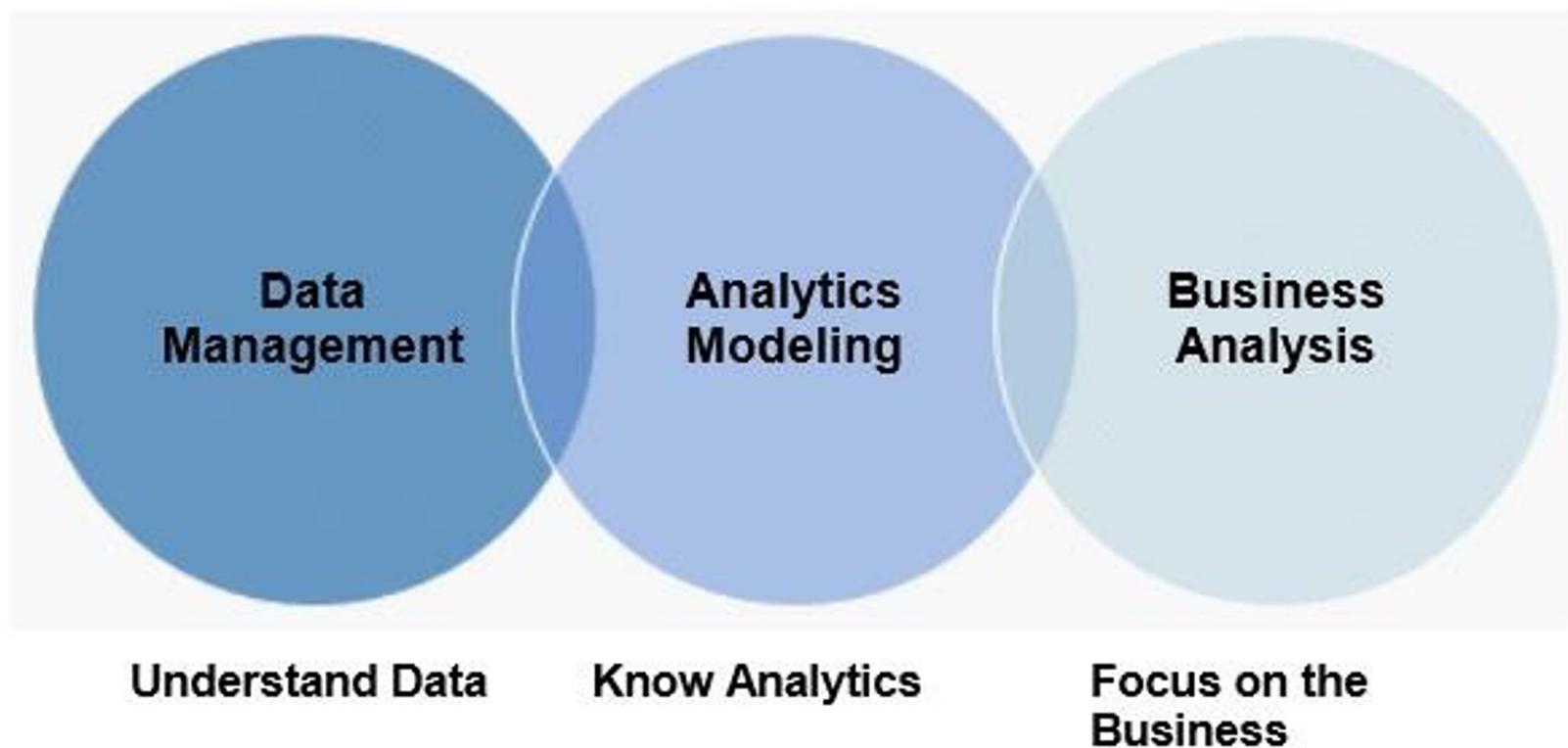


Figure 3. Core Data Scientist Skills



Source: Gartner (March 2012)

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

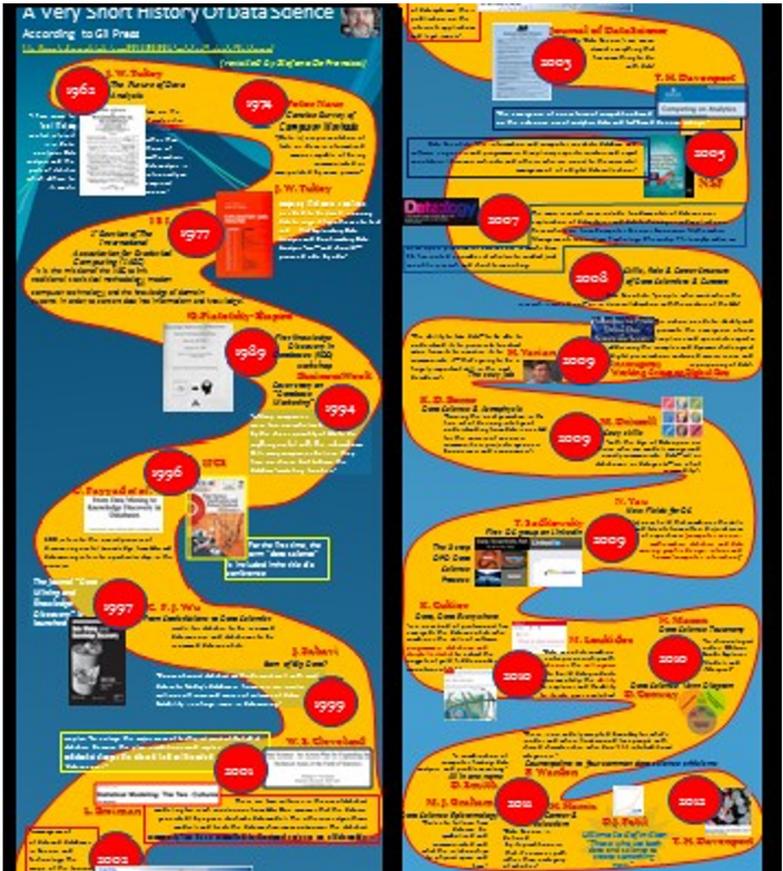
- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

MarketingDistillery.com is a group of practitioners in the area of e-commerce marketing. Our fields of expertise include: marketing strategy and optimization; customer tracking and on-site analytics; predictive analytics and econometrics; data warehousing and big data systems; marketing channel insights in Paid Search, SEO, Social, CRM and brand.

Marketing
DISTILLERY

Breve storia della scienza dei dati

(Vagamente basato sulla versione Gil Press)



<http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science>

1962

J. W. Tukey *The Future of Data Analysis*

"I have come to feel that my central interest is in *data analysis*... Data analysis, and the parts of statistics which adhere to it, must...



take on the characteristics of science rather than those of mathematics... data analysis is intrinsically an empirical science"

1974

Peter Naur
Concise Survey of Computer Methods
"[Data is] a representation of facts or ideas in a formalized manner capable of being communicated or manipulated by some process."

J. W. Tukey

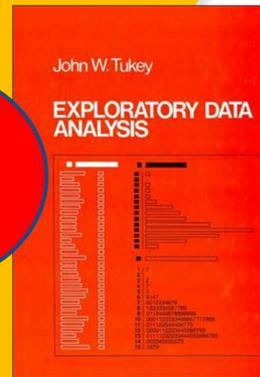
...arguing that more emphasis needed to be placed on using data to suggest hypotheses to test and that Exploratory Data Analysis and Confirmatory Data Analysis "can—and should—proceed side by side."

ISI

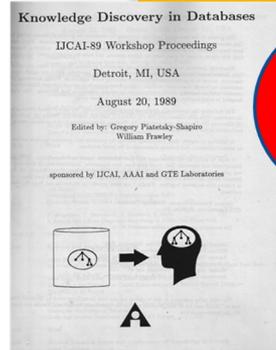
1^o Section of The International Association for Statistical Computing (IASC)

"It is the mission of the IASC to link traditional statistical methodology, modern computer technology, and the knowledge of domain experts in order to convert data into information and knowledge."

1977



G.Piatetsky-Shapiro



1989

*First Knowledge
Discovery in
Databases (KDD)
workshop*

BusinessWeek

*Cover story on
"Database
Marketing"*

1994

"...Many companies were too overwhelmed by the sheer quantity of data to do anything useful with the information... Still, many companies believe they have no choice but to brave the database-marketing frontier."

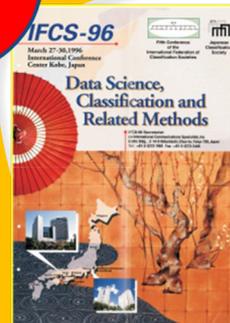
1996

IFCS

U. Fayyad *et al.*

From Data Mining to
Knowledge Discovery in
Databases

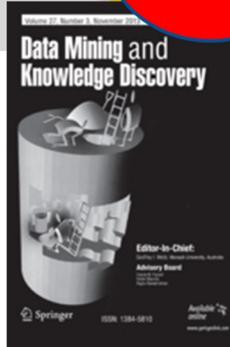
Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth



KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process.

For the first time, the term "**data science**" is included in the title of a conference

The journal "Data Mining and Knowledge Discovery" is launched



1997

C. F. J. Wu

From Statisticians to Data Scientist
...calls for statistics to be renamed data science and statisticians to be renamed data scientists

J. Zahavi

"Conventional statistical methods work well with small data sets. Today's databases, however, can involve millions of rows and scores of columns of data... Scalability is a huge issue in data mining."

Born of Big Data?

1999

...a plan "to enlarge the major areas of technical work of the field of statistics. Because the plan is ambitious and implies substantial change, the altered field will be called 'data science.'"

W. S. Cleveland

Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics

William S. Cleveland
Statistics Research, Bell Labs
wsc@bell-labs.com

2001

Statistical Science
2001, Vol. 16, No. 3, 199-231
Statistical Modeling: The Two Cultures
Leo Breiman

L. Breiman

There are two cultures in the use of statistical modeling to reach conclusions from data. One assumes that the data are generated by a given stochastic data model. The other uses algorithmic models and treats the data mechanism as unknown. The statistical community has been committed to the almost exclusive use of data models.

“...management of data and databases in Science and Technology. The scope of the Journal includes descriptions of data systems, their publication on the internet, applications and legal issues.”

2002

DATA
SCIENCE
Journal



CODATA



Journal of Data Science

2003

“By "Data Science", we mean almost everything that has something to do with data”

T. H. Davenport

“the emergence of a new form of competition based on the extensive use of analytics, data, and fact-based decision making...”



Data Scientists: “the information and computer scientists, database and software engineers and programmers, disciplinary experts, curators and expert annotators, librarians, archivists, and others, who are crucial to the successful management of a digital data collection.”

2005

NSF





2007

The main research areas include fundamentals of data science, exploration of datanature, and data technologies and applications. Researchers are from Computer Science, Economics, Mathematics, Management, Journalism, Psychology, Chemistry, Philosophy, and so on.

As an open platform for data science research, Area 96 has invited a number of scholars to conduct joint scientific research and short term visiting.

2008

Skills, Role & Career Structure of Data Scientists & Curators

Data Scientists: “people who work where the research is carried out—or in close collaboration with the creators of the data”



“The nation needs to identify and promote the emergence of new disciplines and specialists expert in addressing the complex and dynamic challenges of digital preservation, sustained access, reuse and repurposing of data”.

2009

“The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that’s going to be a hugely important skill in the next decades...”. *The sexy job*

H. Varian



Interagency Working Group on Digital Data

K. D. Borne

Data Science & Astrophysic

“Training the next generation in the fine art of deriving intelligent understanding from data is needed for the success of sciences, communities, projects, agencies, businesses, and economies.”

2009

M. Driscoll

Sexy skills

“with the Age of Data upon us, those who can model, munge, and visually communicate data—call us statisticians or data geeks—are a hot commodity.”



N. Yau

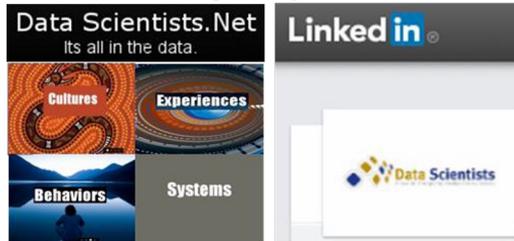
New Fields for DS

“ [a] new field that combines the skills and talents from often disjoint areas of expertise... [computer science; mathematics, statistics, and data mining; graphic design; infovis and human-computer interaction]”

2009

T. Sadkowsky

First DS group on LinkedIn



***The 3 step
OPD Data
Science
Process***

K. Cukier

Data, Data Everywhere

"... a new kind of professional has emerged, the data scientist, who combines the skills of software programmer, statistician and storyteller/artist to extract the nuggets of gold hidden under mountains of data"



2010



M. Loukides

"Data scientists combine entrepreneurship with patience, the willingness to build data products incrementally, the ability to explore, and the ability to iterate over a solution"

H. Mason

Data Science Taxonomy

"In chronological order: Obtain, Scrub, Explore, Model, and iNterpret"

2010

D. Conway

Data Science Venn Diagram



“a combination of computer hacking, data analysis, and problem solving”

All in one name

D. Smith

M. J. Graham

Data Science Epistemology

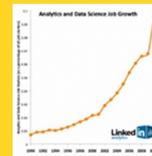
“Rules to follow. how data can be symbolized and communicated and what the relationships to physical space and time”

2011

H. Harris

Career & eclecticism

“Data Science is defined by its practitioners, that it’s a career path rather than a category of activities”



D.J. Patil

Ultimate definition

“Those who use both data and science to create something new.”

2012



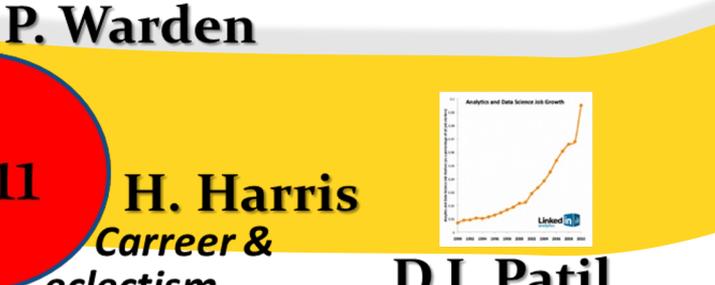
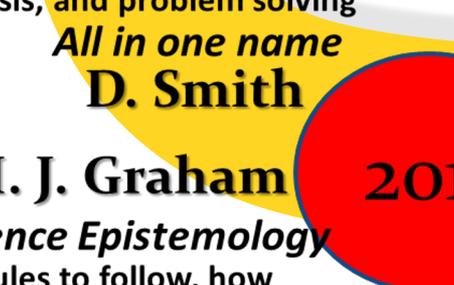
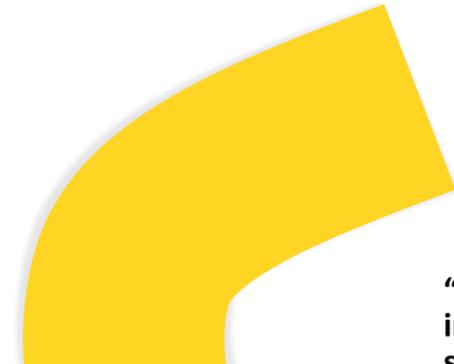
Data Scientist: The Sexiest Job of the 21st Century

T. H. Davenport

“There is no widely accepted boundary for what’s inside. and when I look around I see people with shared characteristics who don’t fit into traditional categories.”

Counterpoints to four common data science criticisms

P. Warden



Passi verso una metafisica della scienza dei dati

- Come si inserisce la Data Science nel contesto della Knowledge Organization?
- Quali sono i suoi rapporti con gli altri campi della conoscenza scientifica?
- La DS può essere spiegata come parte della filosofia della scienza?

	Dati	Informazioni	Conoscenza
Scientifico contesto	Dati Scienza	Informazione Scienza	Conoscenza Scienza
Filosofico contesto	Filosofia dei dati	Filosofia dell'informazione	Filosofia della conoscenza (epistemologia, gnoseologia)

<https://podcasts.ox.ac.uk/luciano-floridi-ethics-open-data>

La Data Science è una scienza della maturità?

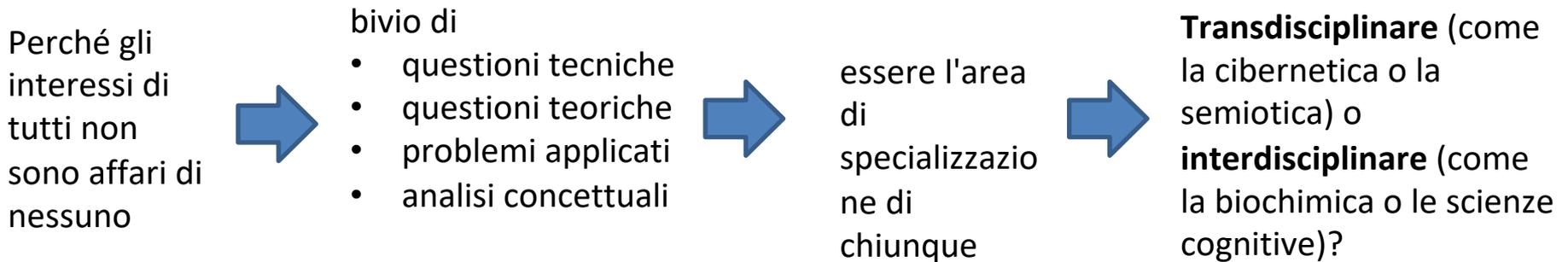
Tipi di dominio trattati da un'impresa intellettuale:

- (a) argomenti (fatti, dati, problemi, fenomeni, osservazioni e simili)
- (b) metodi (tecniche, approcci e Presto)
- (c) teorie (ipotesi, spiegazioni e così via)

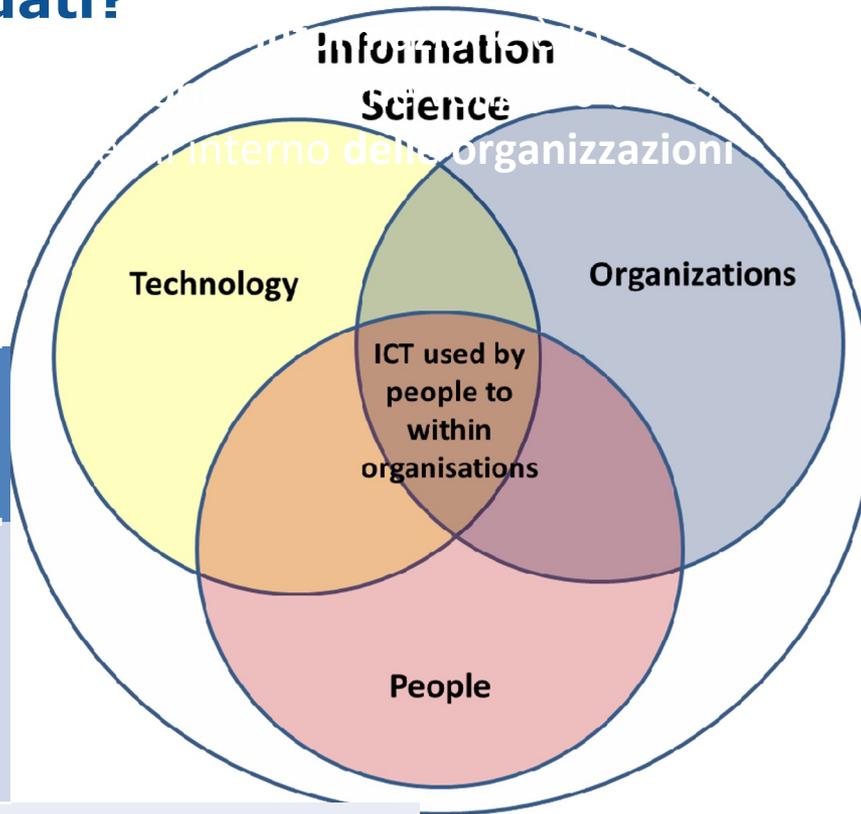
Caratteristica di una nuova disciplina:

- (a) Rappresentare un campo autonomo (*argomenti unici*)
- (b) Fornire un approccio innovativo ad argomenti filosofici sia tradizionali che nuovi (*metodologie originali*);
- (c) Stare accanto alle altre discipline, offrendo la trattazione sistematica dei propri fondamenti concettuali (*nuove teorie*).

Se una disciplina tenta di innovare in più di uno di questi ambiti contemporaneamente è prematuro, poiché si stacca troppo bruscamente dal filo normale e continuo dell'evoluzione del suo campo generale (Stent 1972).



Oltre la scienza dei dati?



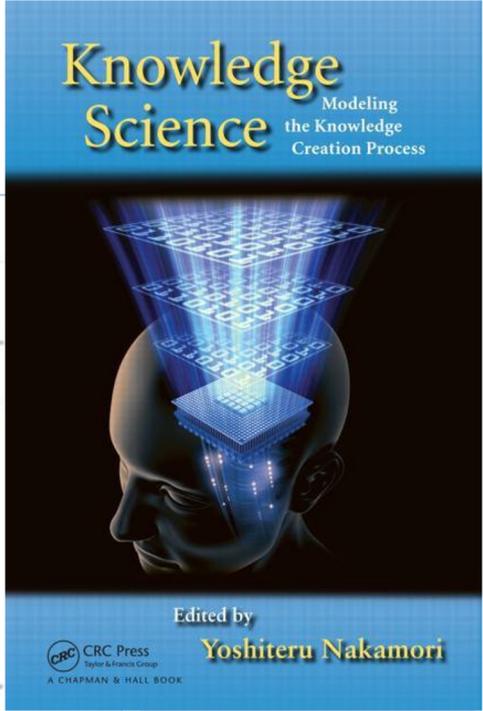
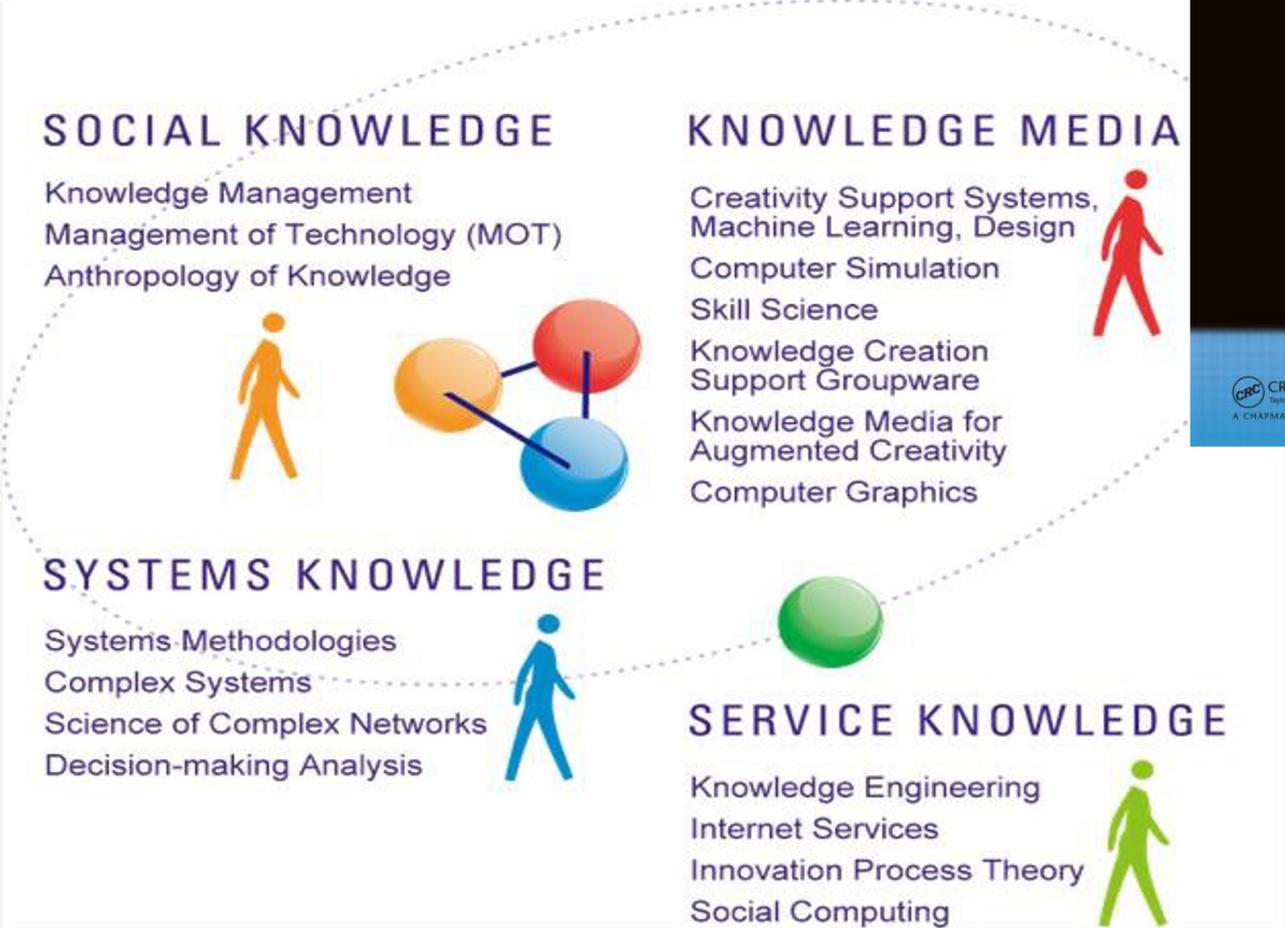
La scienza dell'informazione si trova all'intersezione tra tecnologia, persone e organizzazioni .

È una disciplina distinta e si concentra sulle tecnologie dell'informazione e della comunicazione (TIC) utilizzate dalle persone per gestire le informazioni

Informazione	Conoscenza
Scienza dell'informazione	

Oltre la scienza dei dati?

The School of Knowledge Science consists of four major content areas.

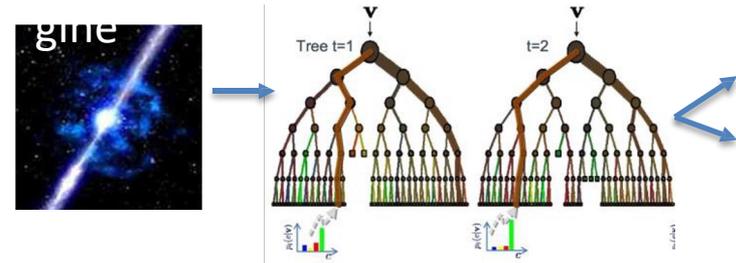
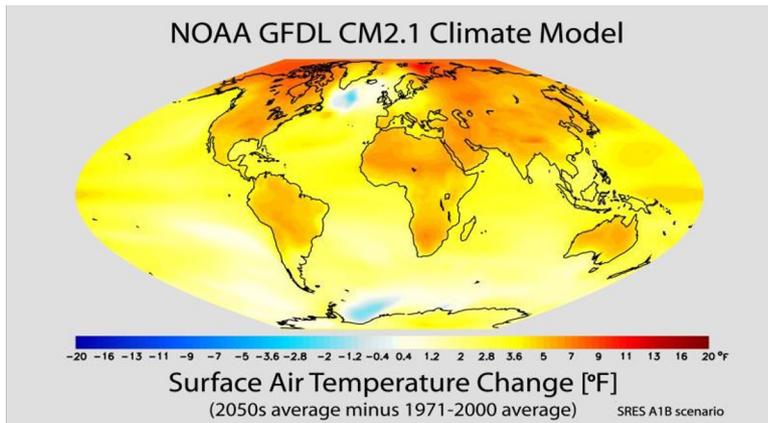


Scienza
della
conoscenza

Scienza dei dati, analisi e consegna del software

Componente	Analisi tradizionale	Consegna del software tradizionale	Scienza dei dati
Utensili	SAS, R, Excel, SQL, strumenti interni	Java, controllo del codice sorgente, Linux, integrazione continua, test unitari, segnalazioni di bug e gestione dei progetti	R, Java, librerie scientifiche Python, Excel, SQL, Hadoop, Hive, Pig, Mahout e altre librerie di machine learning, github per il controllo del codice sorgente e la gestione dei problemi
Metodi analitici	Regressioni, classificazioni, misurazione dell'accuratezza della previsione e della copertura/errore, campionamento	N / A	Classificazione, clustering, rilevamento di similarità, raccomandazioni, apprendimento supervisionato e non supervisionato, calcoli su piccola e larga scala, misurazione dell'accuratezza della previsione e della copertura/errore
Squadra Struttura	Statistici, matematici, scienziati	Sviluppatori, Project Manager, Ingegneri di sistema	Matematici, statistici, scienziati, sviluppatori, ingegneri di sistema
Lasso di tempo	<p>O:</p> <ul style="list-style-type: none"> • Solitamente ricerca e scoperta continua all'interno di un team dell'organizzazione <p>O:</p> <ul style="list-style-type: none"> • Progetto specifico per determinare le risposte 	Ciclo regolare di rilascio del software, consegna continua, ecc.	<p>O:</p> <ul style="list-style-type: none"> • Fase di scoperta/apprendimento che porta allo sviluppo del prodotto <p>O:</p> <ul style="list-style-type: none"> • Ricerca e prodotto continui invenzione/miglioramento

Contrasto: informatica scientifica



Modellazione scientifica

Modelli basati sulla fisica

Strutturato per problemi

Per lo più deterministico, preciso

Esegui su supercomputer o cluster di elaborazione di fascia alta

Approccio basato sui dati

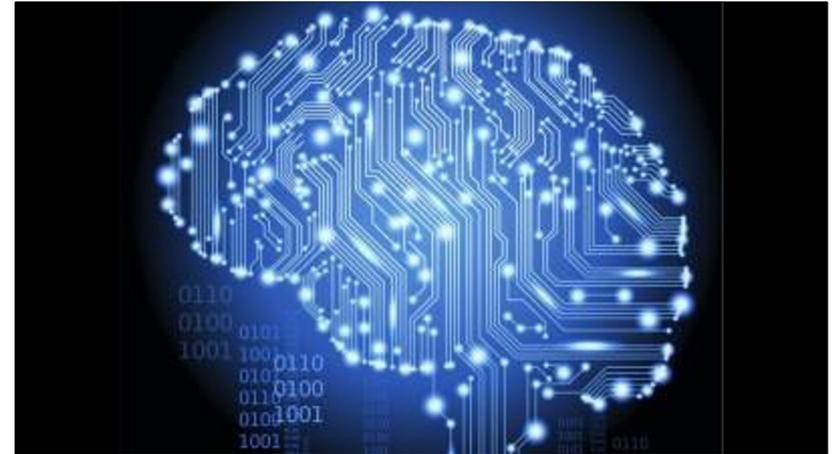
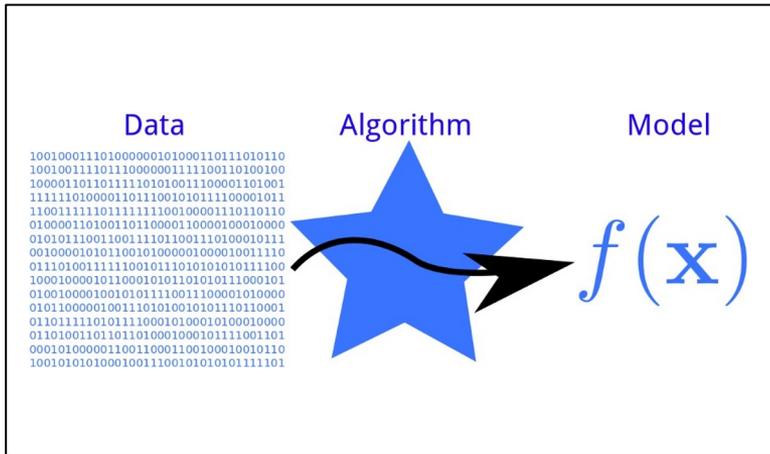
Il motore di inferenza generale sostituisce il modello

Struttura non correlata al problema

I modelli statistici gestiscono la vera casualità,
e **complessità non modellata** .

Funziona su cluster di computer più economici (EC2)

Contrasto: apprendimento automatico



Apprendimento automatico

Sviluppare nuovi modelli (individuali).

Dimostrare le proprietà matematiche di Modelli

Migliorare/convalidare alcuni, relativamente set di dati piccoli e puliti

Pubblica un articolo 😊

Scienza dei dati

Esplora molti modelli, costruisci e ottimizza gli ibridi

Comprendere le proprietà empiriche dei modelli

Sviluppare/utilizzare strumenti in grado di gestire enormi set di dati

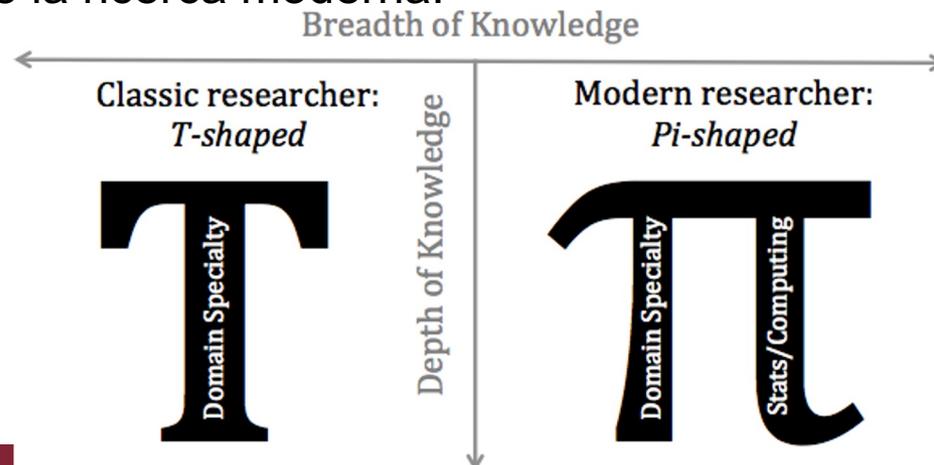
Agire!

Contrasto: ingegneria dei dati

	Scienza dei dati	Ingegneria dei dati
Approccio	Scientifico (Esplorazione)	Ingegneria (sviluppo)
I problemi	Illimitato	Delimitato
Percorso verso la soluzione	Iterativo, esplorativo, non lineare	Per lo più lineare
Formazione scolastica	Di più è meglio (comune del dottorato)	BS e/o autodidatta
Abilità di presentazione	Importante	Non così importante
Esperienza di ricerca	Importante	Non così importante
Programmazione Competenze	Non così importante	Importante
Competenze sui dati	Importante	Importante

Scienza dei dati e mondo accademico

- Nelle parole di Alex Szalay, questo tipo di ricercatori devono essere “a forma di Pi” in contrapposizione al più tradizionale ricercatore “a forma di T”. Secondo Szalay, un classico programma di dottorato genera ricercatori a forma di T: scienziati con una conoscenza generale ampia ma superficiale, ma con profonde abilità ed esperienza in un'area particolare. La nuova generazione di ricercatori scientifici, i data scientist, deve essere a forma di Pi: cioè, mantengono la stessa ampia ampiezza, ma si spingono più in profondità sia nella propria area tematica che nei metodi statistici o computazionali che aiutano a guidare la ricerca moderna:



Scienza dei dati e mondo accademico

- In un post di Jake Vanderplas nel 2014 relativo alla discussione di SciFoo su: ***Academia and Data Science*** , sono state discusse le seguenti domande.
- Ti incoraggio a sviluppare i tuoi pensieri su di essi e ad elaborare la tua valutazione
 - Dove si inserisce Data Science all'interno dell'attuale struttura del università e istituti di ricerca?
 - Che cosa vogliono i data scientist accademici dalla loro carriera? Come può il mondo accademico offrirlo?
 - Quali fattori potrebbero spingere il mondo accademico verso il riconoscimento e la ricompensa scienziati dei dati nei campi dei domini?
 - Riconoscendo che i laureati continueranno a lavorare sia nel mondo accademico che nell'industria, come possiamo prepararli al meglio per il successo in entrambi i mondi?

Applicazioni di scienza dei dati

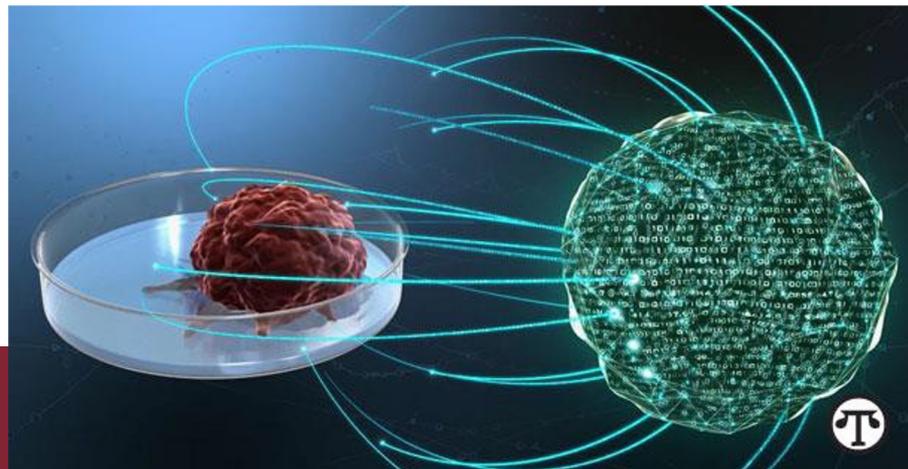
	Attività commerciale	Assistenza sanitaria	Partenza urbana
Riepilogo	Dalla progettazione automobilistica alle assicurazioni fino alla consegna della pizza, le aziende utilizzano la scienza dei dati per ottimizzare le proprie operazioni e soddisfare meglio le aspettative dei propri clienti.	L'assistenza sanitaria di domani potrebbe sembrare più efficiente grazie a cose come le cartelle cliniche elettroniche. Potrebbe anche sembrare molto più efficace. All'orizzonte si profilano una riduzione delle riammissioni, migliori cure e una diagnosi precoce.	Per la prima volta nella storia umana, più persone vivono nelle città che nelle aree suburbane o rurali. Un campo emergente chiamato "informatica urbana" combina la scienza dei dati con le sfide uniche che devono affrontare le città in crescita del mondo
Che cosa sta accadendo?	Strada a doppio senso per l'auto elettrica Ford Focus	Ridurre le riammissioni ospedaliere	Affrontare il traffico di megalopoli
	Un migliore rilevamento delle frodi aumenta la soddisfazione dei clienti	Migliori decisioni al punto di cura	Combattere il crimine con la "polizia predittiva" dei dati
	Approfondimenti sull'e-commerce: la salsa segreta di Domino		
Cosa è possibile	Utilizzo dei dati social per selezionare punti vendita di successo	Esami medici tramite specchi da bagno	Strumentare le città

Contrasto: scienze computazionali



Scienza dei dati: caso di studio sulla ricerca sul cancro

- Il cancro è una malattia incredibilmente complessa; un singolo tumore può avere più di **100 miliardi di cellule** e ciascuna cellula può acquisire mutazioni individualmente. La malattia è in continua evoluzione, evoluzione e adattamento.
- Sfrutta la potenza dell'analisi dei Big Data e dell'elaborazione ad alte prestazioni.
- Sfrutta modelli sofisticati e algoritmi di apprendimento automatico per identificare modelli potenzialmente collegati al cancro
- Enorme quantità di elaborazione e riconoscimento dei dati



Scienza dei dati: caso di studio sull'assistenza sanitaria

Stanford Medicine, Google team up to harness power of data science for health care

Stanford Medicine will use the power, security and scale of Google Cloud Platform to support precision health and more efficient patient care.

**AUG 8
2016** Stanford Medicine and Google are working together to transform patient care and medical research through data science.

The new collaboration combines Stanford Medicine's excellence in health-care research and clinical work with Google's expertise in cloud technology and data science. Stanford's forthcoming Clinical Genomics Service, which puts genomic sequencing into the hands of clinicians to help diagnose disease, will be built using Google Genomics, a service that applies the same technologies that power Google Search and Maps to securely store, process, explore and share genomic data sets.



Lloyd Minor, dean of the School of Medicine, says the collaboration with Google marks a "milestone for the future of patient care and research."
Glenn Matsumura

Stanford Medicine includes the Stanford School of Medicine, Stanford Health Care and Stanford Children's Health. Together, Stanford Medicine and Google will build cloud-based applications for exploring massive health-care data sets, a move that could transform patient care and medical research.

"Stanford Medicine and Google are committing to major investments in preventing and curing diseases that afflict ordinary people worldwide. We're proud to be setting this milestone for the future of patient care and research," said Lloyd Minor, MD, dean of the School of Medicine.

The agreement — considered key to Stanford Health Care's development of the Clinical Genomics Service — makes Google Inc. a formal business associate of Stanford Medicine. As such, Google and Stanford will both comply with the Health Insurance Portability and Accountability Act, a federal law that regulates the privacy and security of medical information. HIPAA requires that Stanford Medicine patient data stored on Google Cloud Platform servers stay private. Patient information will be encrypted, both in transit and on servers, and kept on servers in the United States.

LEADING IN PRECISION HEALTH



Stanford Medicine is leading the biomedical revolution in precision health, defining and developing the next generation of care that is proactive, predictive and precise.

[Learn more →](#)

A LEGACY OF INNOVATION



Stanford Medicine's unrivaled atmosphere of breakthrough thinking and interdisciplinary collaboration has fueled a long history of achievements.

[View timeline →](#)

Scienza dei dati: elezioni di casi di studio

- Le campagne di Obama nel 2008 e nel 2012 sono accreditate per il loro uso riuscito dei social media e del data mining.
- Micro-targeting nel 2012
 - <http://www.theatlantic.com/politics/archive/2012/04/the-creepiness-factor-how-obama-and-romney-are-getting-to-know-you/255499/>
 - <http://www.mediabizbloggers.com/group-m/How-Data-and-Micro-Targeting-Won-the-2012-Election-for-Obama---Antony-Young-Mindshare-North-America.html>
- Microprofili costruiti da più fonti accessibili tramite AP, dati aggiornati in tempo reale basati su visite porta a porta, acquisti media mirati, e-mail e messaggi Facebook altamente mirati.
- 1 milione di persone hanno installato l'app Facebook di Obama che dava accesso alle informazioni sugli "amici".

Scienza dei dati: caso di studio Internet delle cose (IoT)

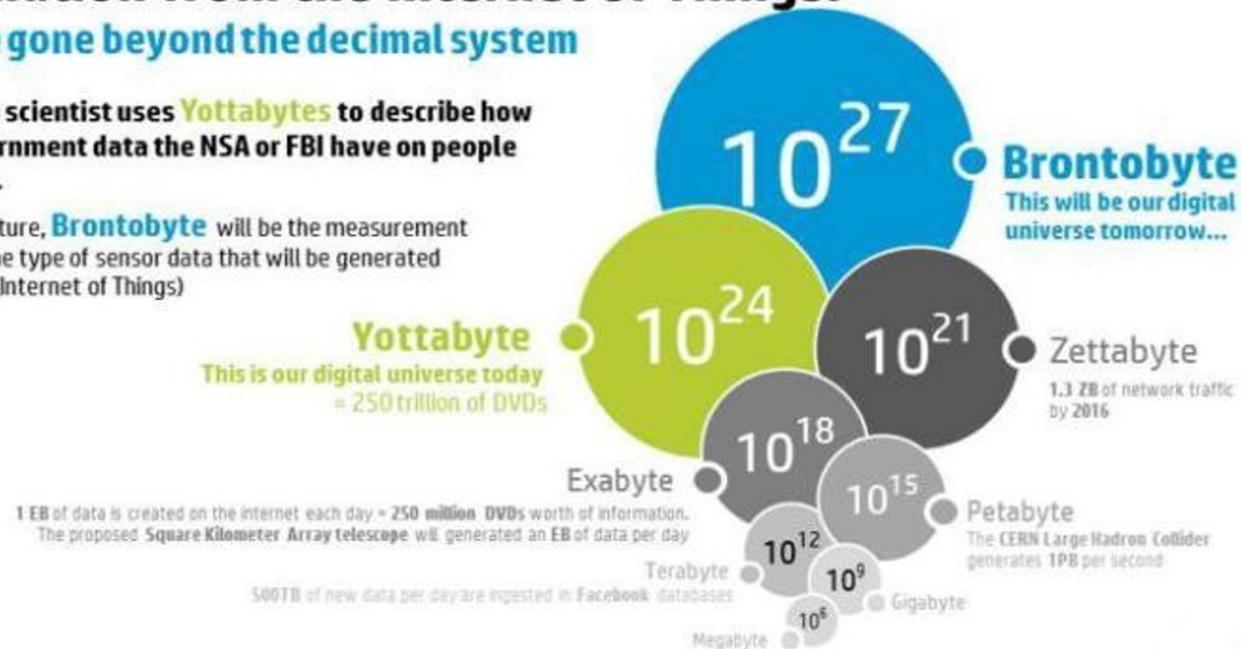
- L'Internet delle cose è in rapida crescita. Si prevede che entro il 2020 saranno connessi più di 25 miliardi di dispositivi.

Information from the Internet of Things:

We have gone beyond the decimal system

Today data scientist uses **Yottabytes** to describe how much government data the NSA or FBI have on people altogether.

In the near future, **Brontobyte** will be the measurement to describe the type of sensor data that will be generated from the IoT (Internet of Things)



- L'Internet delle cose (IOT) produrrà presto un volume enorme e una varietà di dati a una velocità senza precedenti. Se "Big Data" è il prodotto dell'IOT, "Data Science" ne è l'anima.

Scienza dei dati: caso di studio sull'analisi dei clienti



**Marketing &
Advertising**

Leveraging customer data to move ever closer to the elusive goal of truly personalized marketing: the right offer, at the right time, in the right location and context, to the right person.



**Customer
Service**

By capturing and analyzing the data from customer touch points within an organization, companies can identify customer pain points and issues proactively and update their customer service FAQs or other communications with existing customers.



**Retention &
Loyalty**

Using customer data and analytics, these companies deploy and refine predictive models that help them retain customers with proactive approaches. Investments, in terms of offers and upgrades, can be made at the right time to increase the likelihood of retaining desirable customers.



**Customer
Experience**

The experience that customers have with companies matters a great deal. Other recent research has highlighted the critical connection between experience and company financial performance.

Punti essenziali

- **I Big Data hanno dato origine alla Data Science**
- **La scienza dei dati affonda le sue radici in solide basi di matematica e statistica, informatica e conoscenza del dominio**
- **Professione sexy – Data Scientist 😊**
- **Non tutto ciò che riguarda i dati o la scienza è Data Science!**
- **I casi d'uso per Data Science sono convincenti**

Conclusione

- In questa sezione hai imparato
 - Quali sono le sfide dei Big Data
 - Cos'è esattamente la Data Science e cosa significano i dati
 - Gli scienziati lo fanno
 - La scienza dei dati si contrapponeva ad altre discipline
 - Casi di studio e casi d'uso

Domande?



A close-up, chest-up portrait of Darth Vader from Star Wars. He is wearing his iconic black helmet and breathing apparatus. The lighting is dramatic, highlighting the metallic textures of his mask and the folds of his chest plate. The background is solid black.

Grazie!