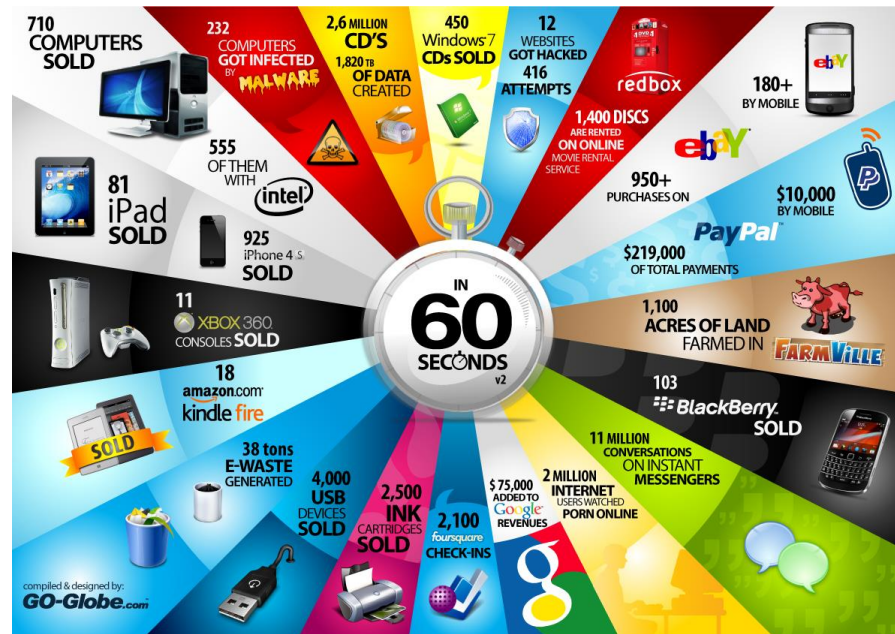


# La visualizzazione dei Big Data: opportunità e sfide

Stefano De Francisci

27 maggio 2015



# Sommario



## Complessità e rappresentazione grafica

“The world is complex, dynamic, multidimensional;  
the paper is static, flat.



How are we to represent  
the rich visual world  
of experience and  
measurement

on mere flatland?”



**E. TUFTE**, *Envisioning Information*, 1990

# Big Data

## Una definizione

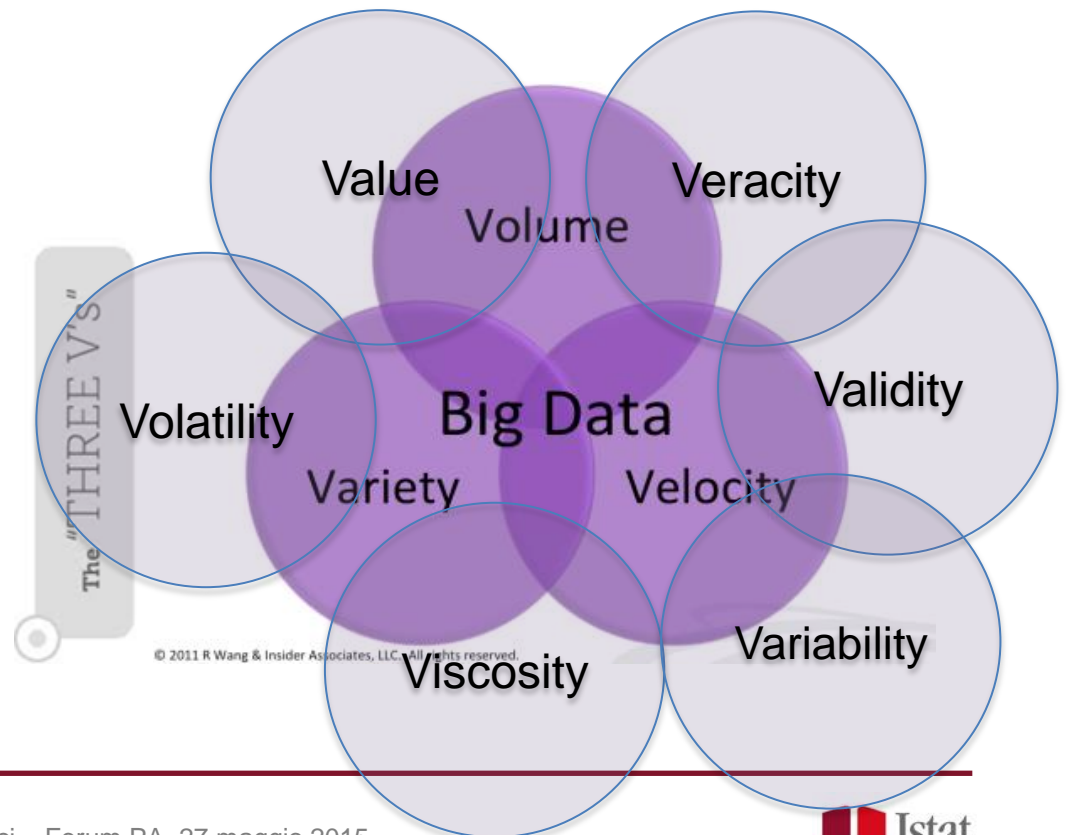
*Big data "refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze."*

[MCKINSEY GLOBAL INSTITUTE, June 2011]

**Volume:** dimensione effettiva del dataset

**Velocità:** velocità di generazione dei dati (analisi dei dati in tempo reale o quasi)

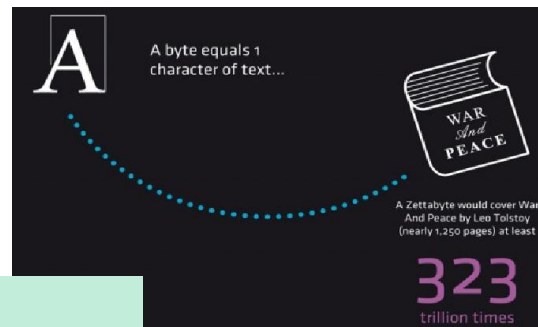
**Varietà:** varie tipologie di dati, provenienti da fonti diverse (strutturate e non)



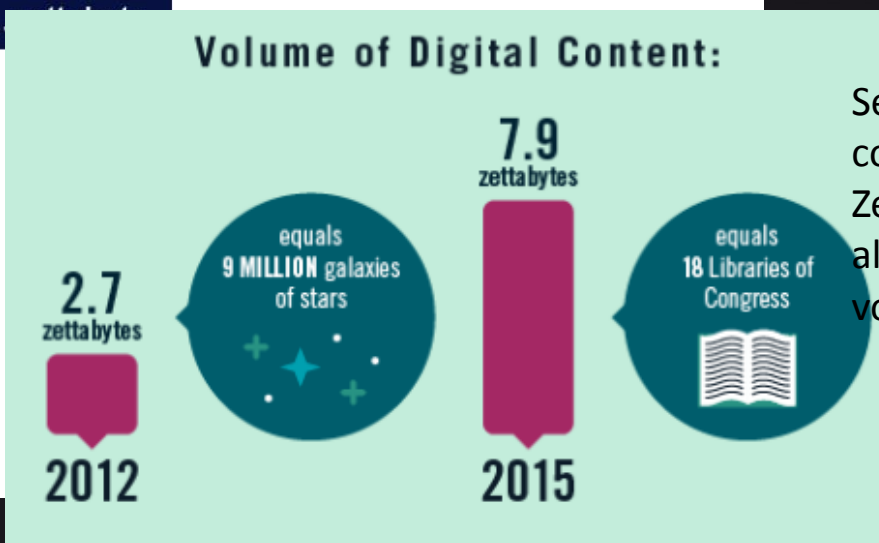
# Quanti dati nel mondo digitale?

## WHAT IS A ZETTABYTE?

1,000,000,000,000 .....gigabyte  
1,000,000,000,000 .....terabyte  
1,000,000,000,000 .....petabyte  
1,000,000,000,000 .....exabyte  
1,000,000,000,000 .....zettabyte

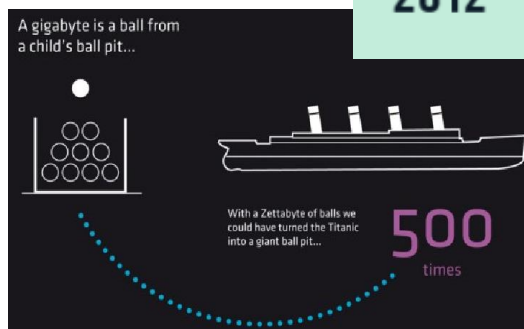


## Volume of Digital Content:

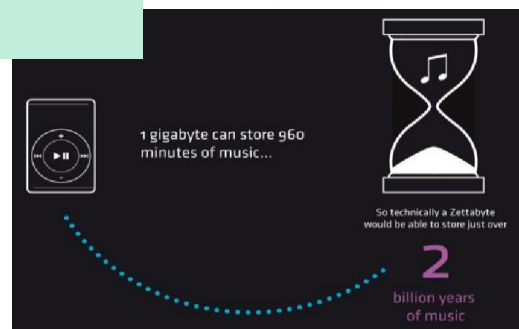


Se una lettera dell'alfabeto corrisponde a un Byte... uno Zettabyte corrisponde ad almeno 323 mila miliardi di volumi di Guerra e Pace...

...una pallina a 500 transatlantici pieni di palline...



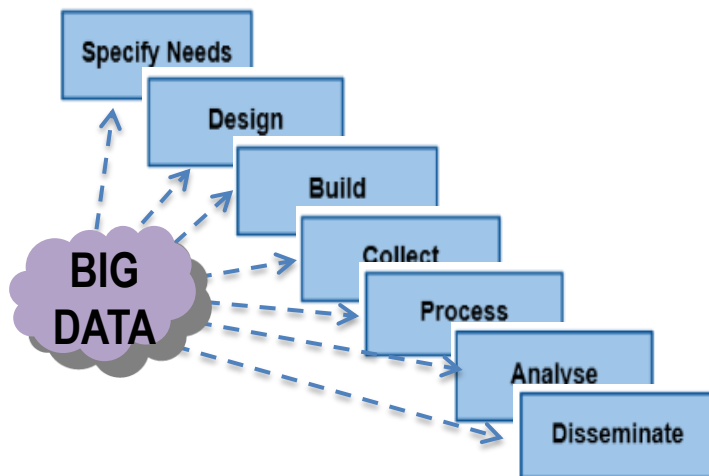
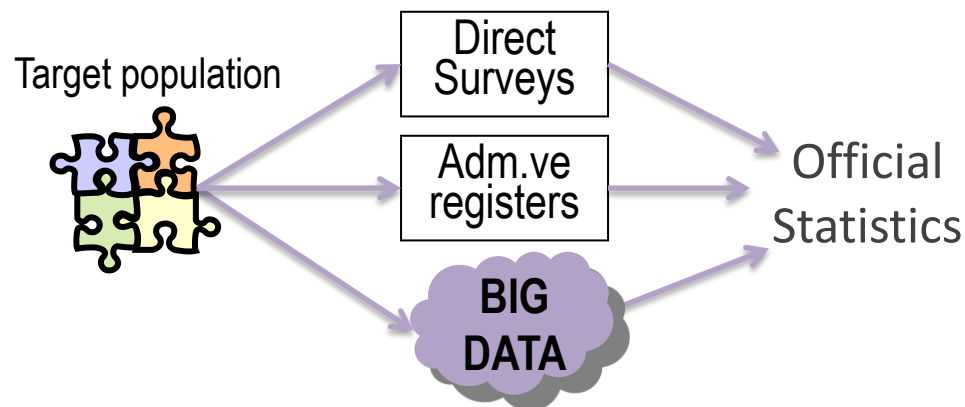
... e un gigabyte di musica (960 minuti)... a 2 miliardi di anni di musica



# Big Data e statistica ufficiale

## Nuova opportunità

Big Data come fonte *addizionale* alle fonti tradizionali della statistica ufficiale

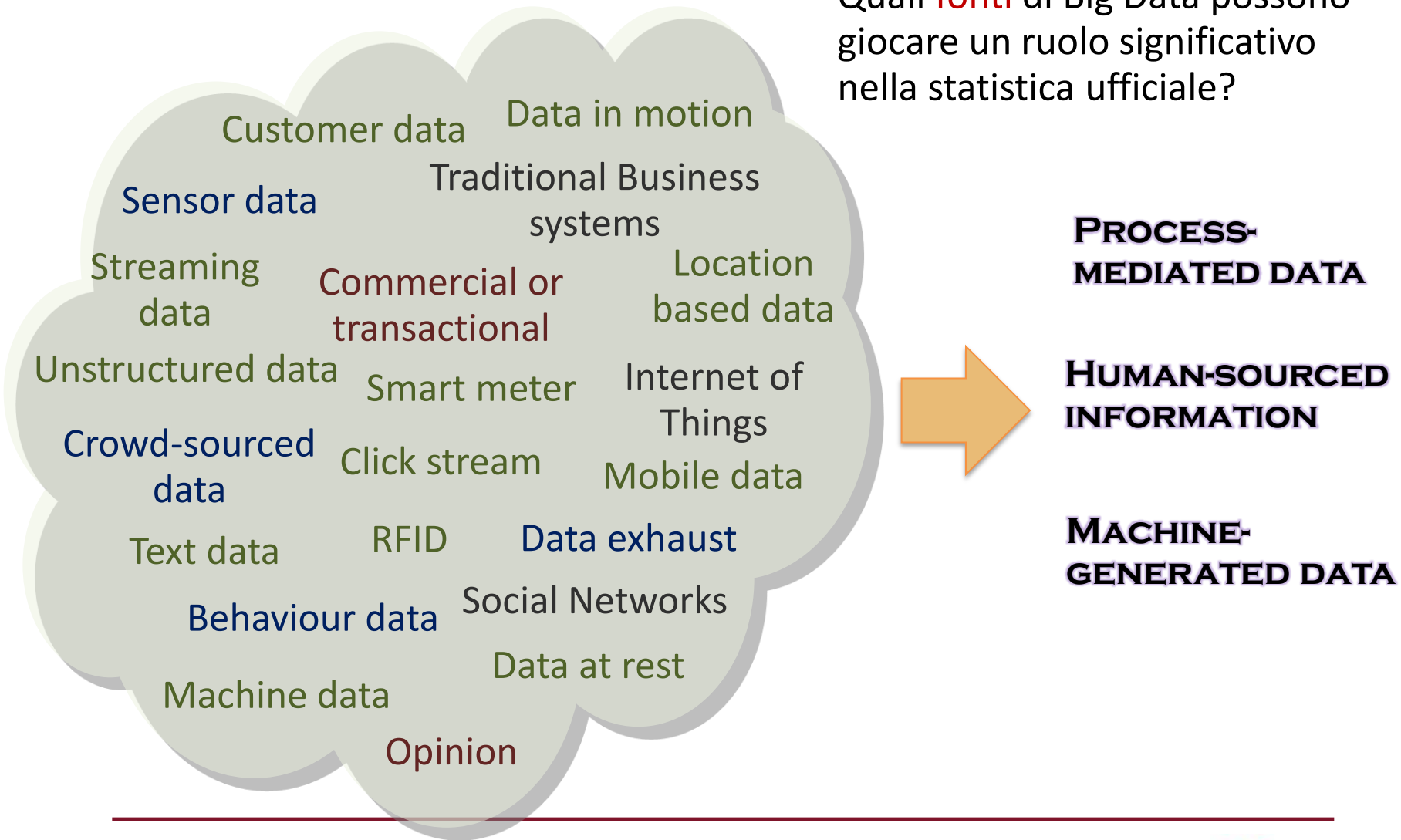


## Nuova sfida

Andare oltre lo stadio sperimentale ed entrare nell'età della maturità

# Un accenno alle fonti

Quali **fonti** di Big Data possono giocare un ruolo significativo nella statistica ufficiale?



## Social Networks

Dati prodotti tramite l'interazione umana con dispositivi digitali (anche mobili):

- Blog posts
- Twitter messages
- User-generated maps



## Traditional Business systems



Dati raccolti da sistemi tradizionali in modalità passiva:

- Medical records
- Commercial transactions
- Banking/stock records

**HUMAN-SOURCED INFORMATION**

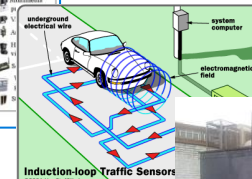
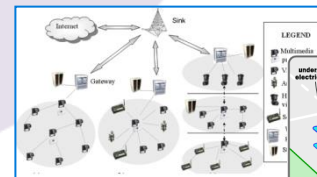
**PROCESS-MEDIATED DATA**

**MACHINE-GENERATED DATA**

## Internet of Things

Sensori e macchine utilizzati per misurare e registrare eventi e situazioni nel mondo fisico

- Satellite imaging
- Climate & environmental sensors
- Road & traffic sensors
- Logs & Web Logs

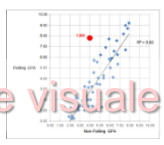
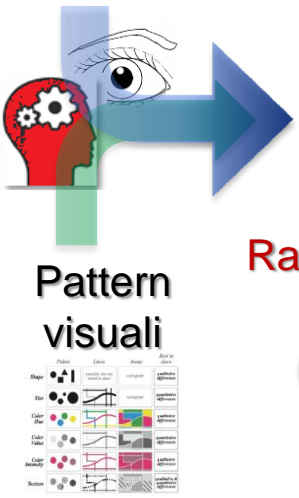




# Il processo cognitivo visivo

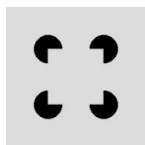
**Dati grezzi**

	Settembre	Ottobre	Novembre	Dicembre	Gen	Feb	Mar	Apr	Mai	Giugno	Luglio	Ago	Sette	Ottobre	Nov	Dicem	Totale
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464
Albergo	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	28	464



**Analisi visuale**

**Sintesi visuale**



**Interazione visuale**

**Esplorazione**

**Dashboard**

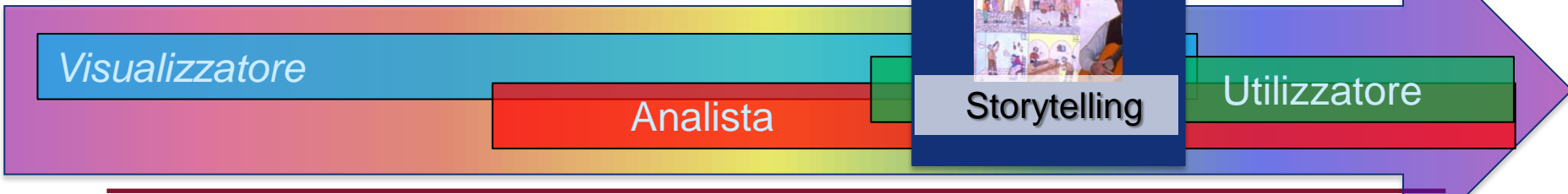
**Esplorazione visuale**

**Presentazione grafica**

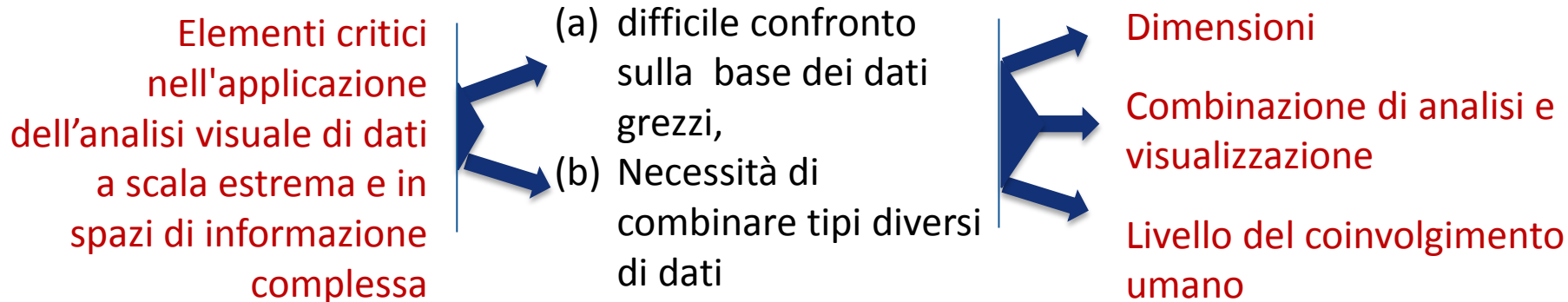
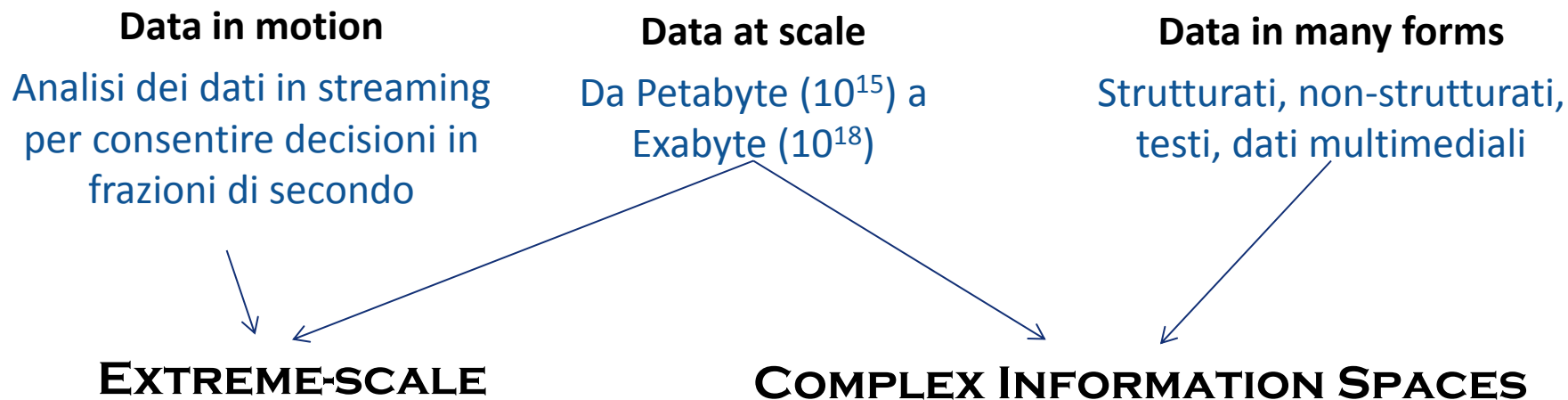
**Infografica**

**Storytelling**

- Problem-solving**
- Sense-Making**
- Making decisions**



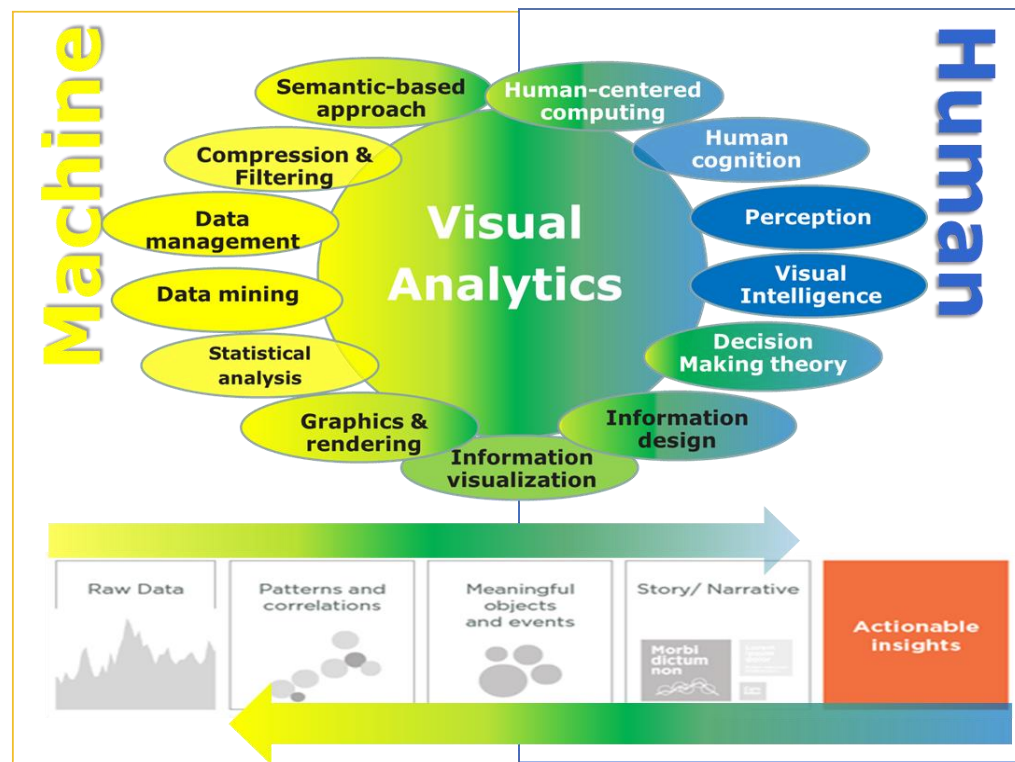
# Quando i dati diventano... *big*



# Analisi automatica e interazione visuale

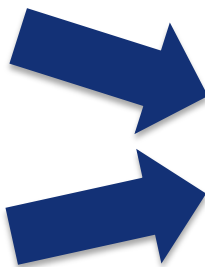
Per sostenere l'intero ciclo di vita dei Big Data, un buon sistema di analisi visiva deve combinare i vantaggi dell'**analisi automatica** con **tecniche interattive** per esplorare i dati.

Dietro questa caratteristica tecnica c'è lo scopo più profondo di integrare la **capacità di analisi di un computer** con la **capacità dell'analisi umana**.



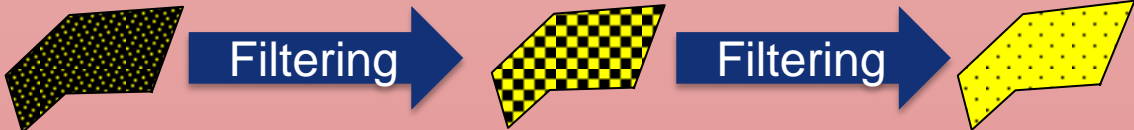
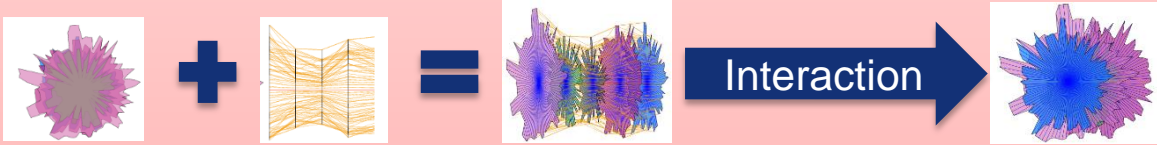
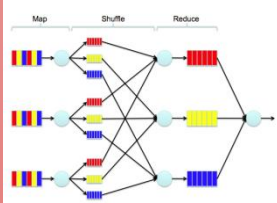
Volume, velocity, variety

Mappare dati complessi in più  
semplici forme visuali di  
conoscenza



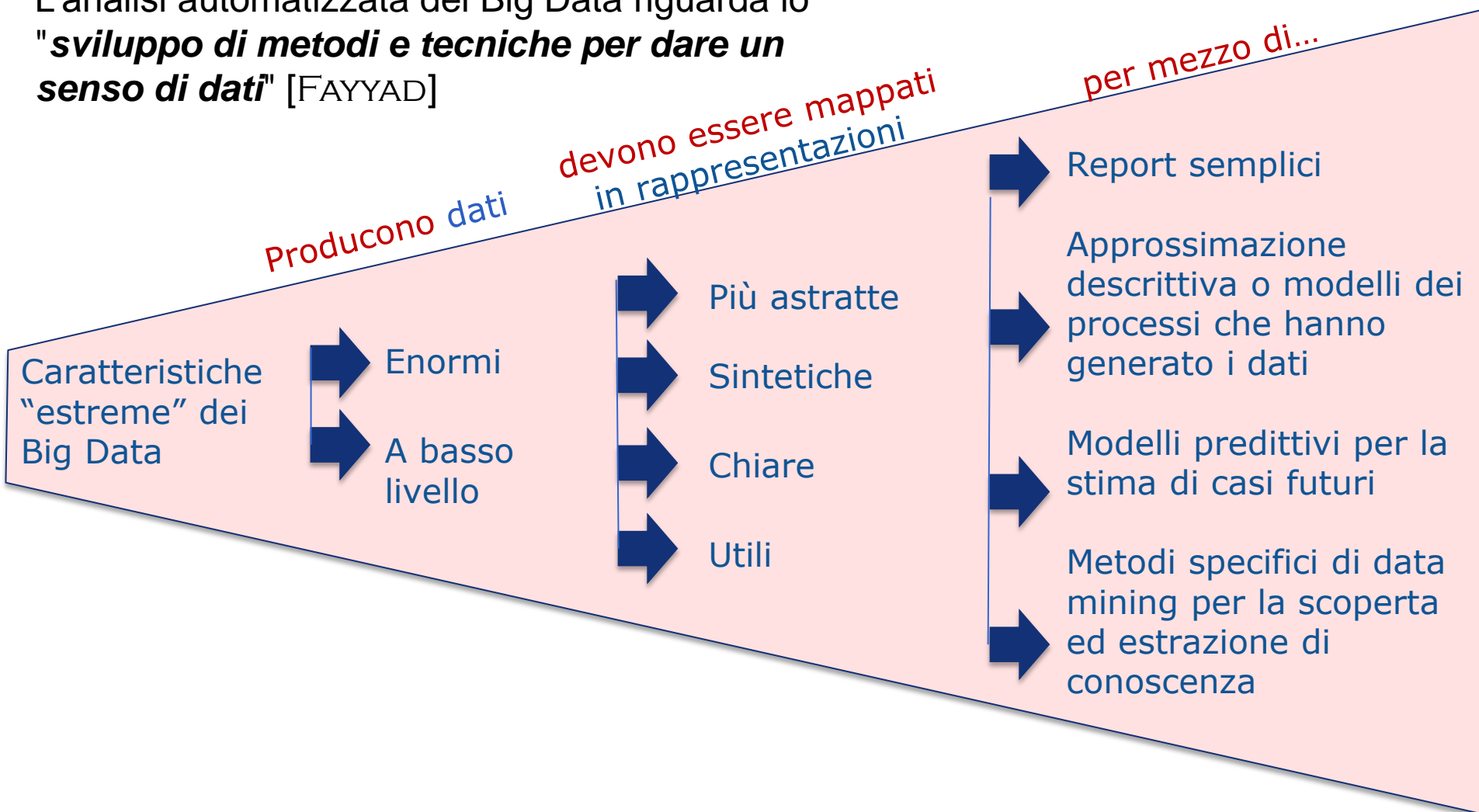
- Definizione appropriata in fase di progettazione
- Raggiungimento di un corretto peso e bilanciamento delle due componenti

# Tre scuole di pensiero

Enfasi su...	Metodo	Riferimento
<b>Data reduction</b>	<p><b>Big Data → Medium Data → Small Data+ R</b></p> 	Wickham
<b>Visual interaction</b>	<p><b>Nuovi pattern di rappresentazione grafica + Interazione</b>            StarGlyphs + Parallel coordinates</p> 	Carpendale
<b>HCP</b>	<p><b>Divide et impera + Calcolo parallelo</b></p> 	Bowei Xi

# Analisi automatica

L'analisi automatizzata dei Big Data riguarda lo **"sviluppo di metodi e tecniche per dare un senso di dati"** [FAYYAD]



## Visualizzazione interattiva

Nel contesto di Big Data possono essere adottate alcune categorie essenziali dell'interazione come ***base di ragionamento analitico*** [YI – ET AL.]

- **Select** (contrassegna qualcosa come interessante)
- **Explore** (mostrami qualcos'altro)
- **Reconfigure** (mostrami una differente disposizione)
- **Encode** (mostrami una differente rappresentazione)
- **Abstract/elaborate** (mostrami più o meno dettagli)
- **Filter** (mostrami qualcosa *sotto certe condizioni*)
- **Connect** (mostrami I concetti collegati)

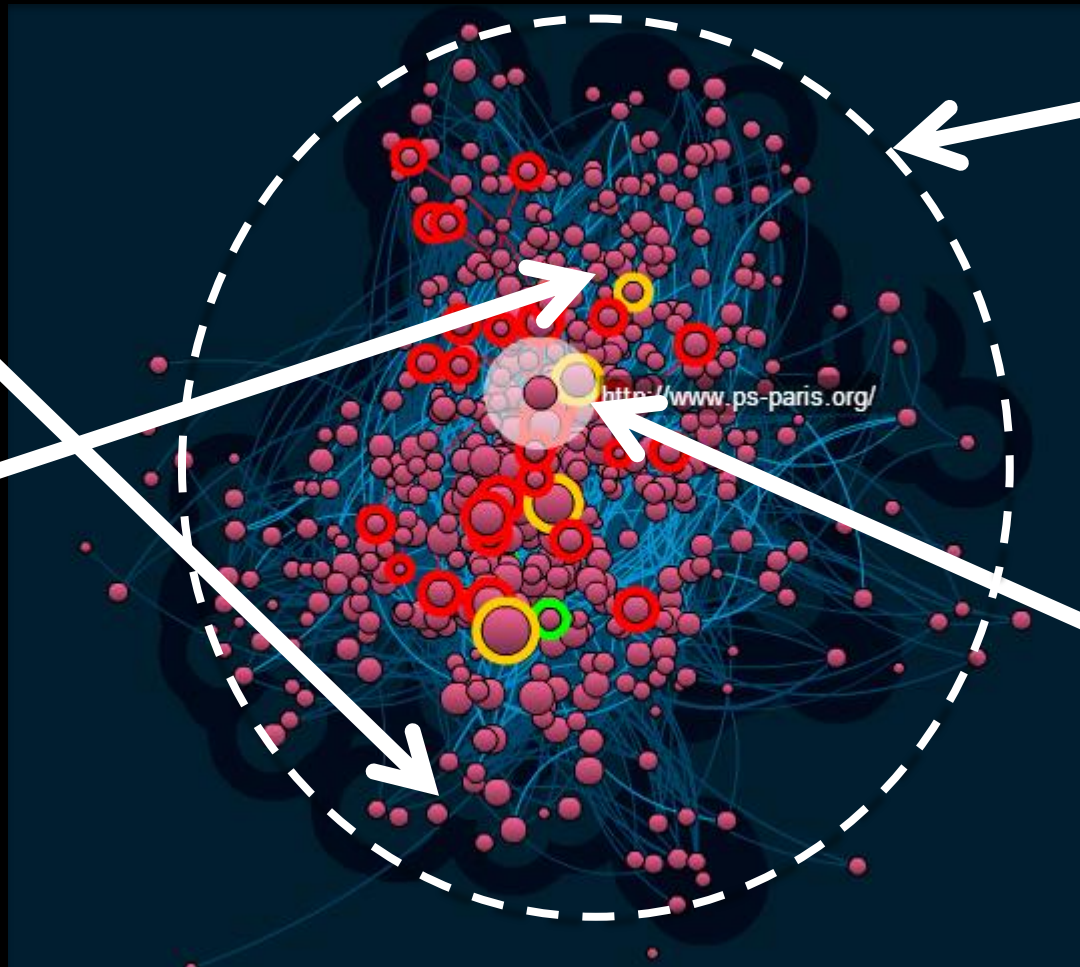
<http://www.cs.tufts.edu/comp/250VA/papers/yi2007toward.pdf>

Select (mark something as interesting)

Abstract/  
elaborate  
(show me more or less detail)

Explore  
(show me something else)

Filter (show me something conditionally)



# La classificazione Big Data di UNECE

## 1. Social Networks (human-sourced information)

**Social Networks**

Blogs and comments

Personal documents

Pictures: Instagram, Flickr, Picasa

Videos: Youtube etc.

Internet searches

Mobile data content: text messages

User-generated maps

E-Mail

## 2. Traditional Business systems (process-mediated data)

**Data produced by Public Agencies**

Medical records

**Data produced by businesses**

Commercial transactions

Banking/stock records

E-commerce

Credit cards

## 3. Internet of Things (machine-generated data)

**Data from sensors**

*Fixed sensors*

Home automation

Weather/pollution sensors

Traffic sensors/webcam

Scientific sensors

Security videos/images

*Mobile sensors (tracking)*

Mobile phone location

Cars

Satellite images

**Data from computer systems**

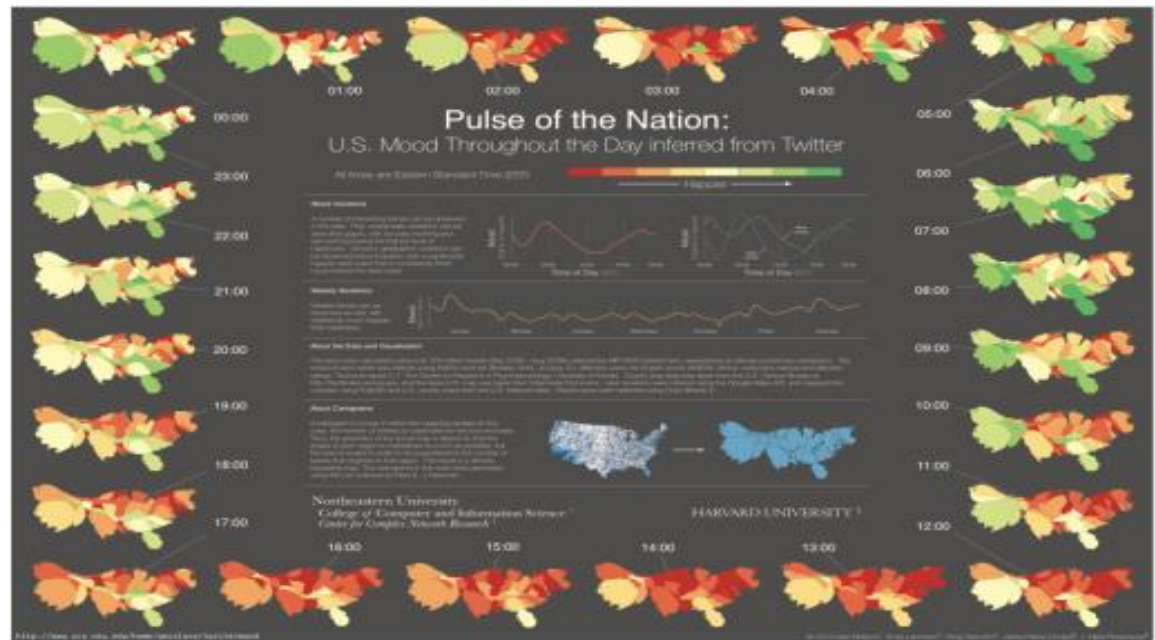
Logs

Web logs



# Human emotion

«This video shows the mood in the U.S., as inferred using over 300 million tweets, over the course of the day. The maps are represented using density-preserving cartograms»



<https://www.youtube.com/watch?v=ujcrJZRSgkg>

# Tweetcatcha

«TweetCatcha seeks to uncover the organic nature of news as it travels through Twitter over time, by examining the movement of NY Times articles through Twitter»

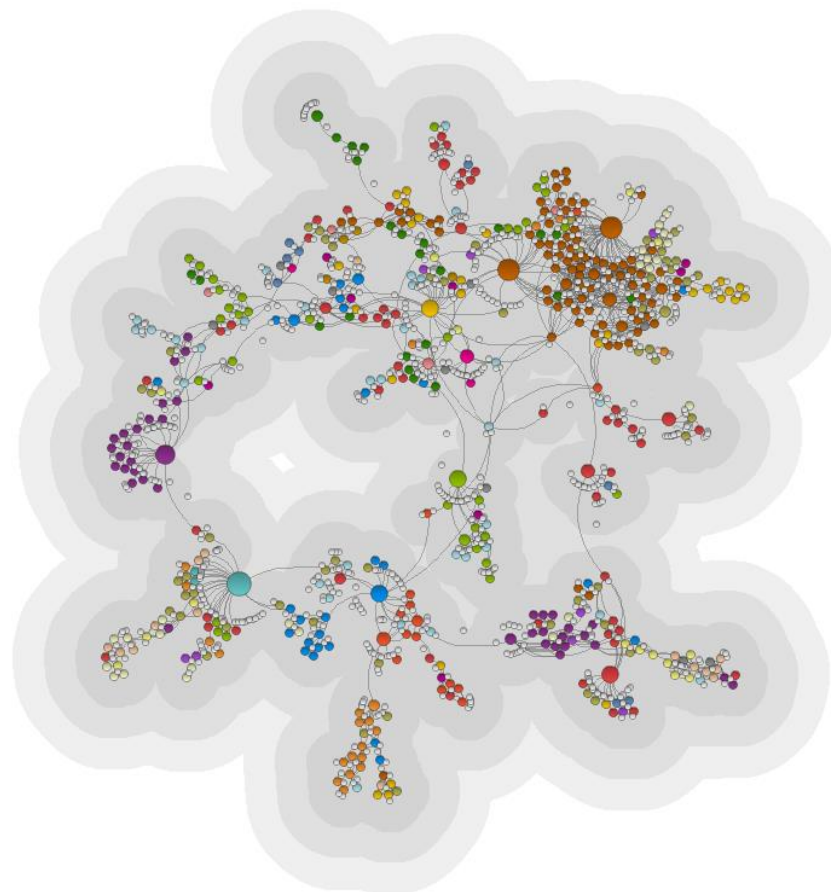


<http://a.parsons.edu/~drumb588/tweetcatcha/>

Medical records

# Human disease network

«The diseasome website is a disease/disorder relationships explorer and a sample of an innovative map-oriented scientific work. Built by a team of researchers and engineers, it uses the Human Disease Network dataset and allows intuitive knowledge discovery by mapping its complexity»



<http://diseasome.eu/map.html>

Mobile phone location

# Urban Mobs

«Cette visualisation représente la quantité de SMS envoyés le soir de la fête de la musique (21 juin 2008). On peut découvrir à partir de 17h une forte activité aux alentours du Parc des Princes que nous pouvons mettre en parallèle avec le concert de Tokio Hotel ce soir là. On remarque ensuite un autre foyer d'activité à l'hippodrome d'Auteuil correspondant au concert organisé par France 2»

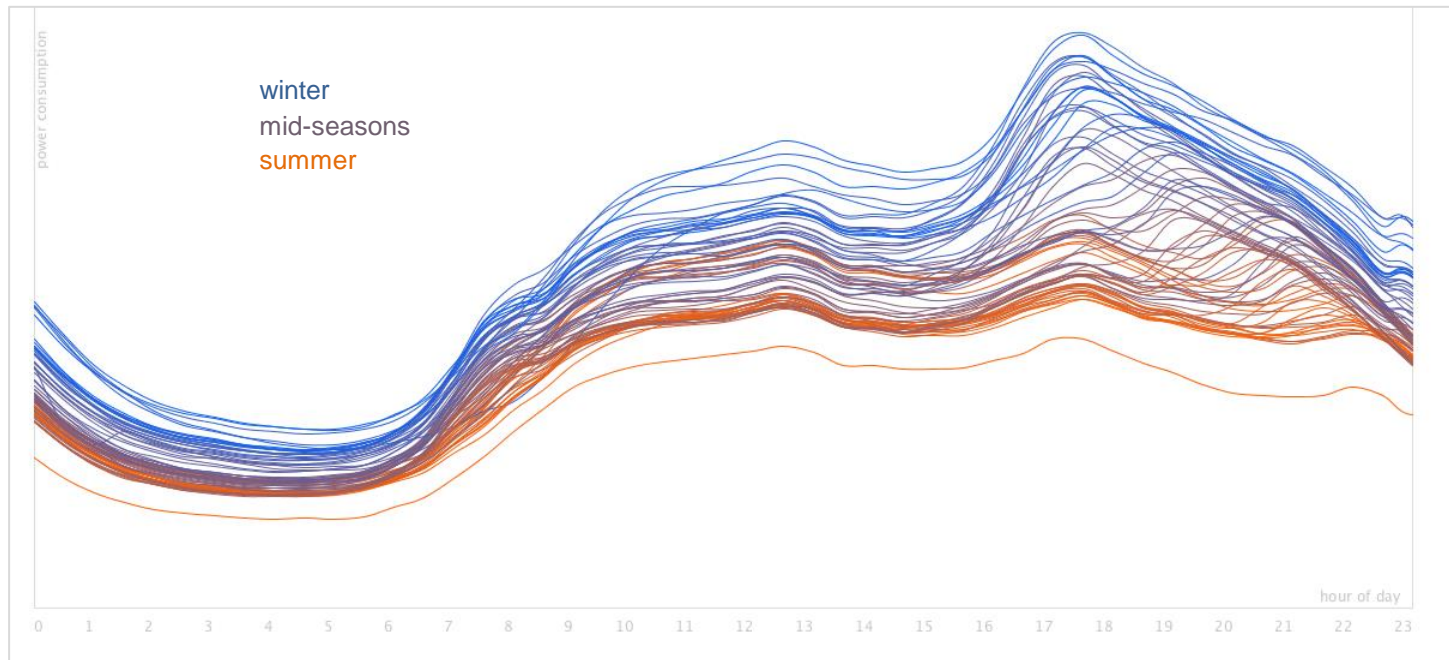


<http://www.urbanmobs.fr/fr/france/>

Fixed sensors

# Visualization of Irish Data

Weekly consumption per hour of day over a year (IE)



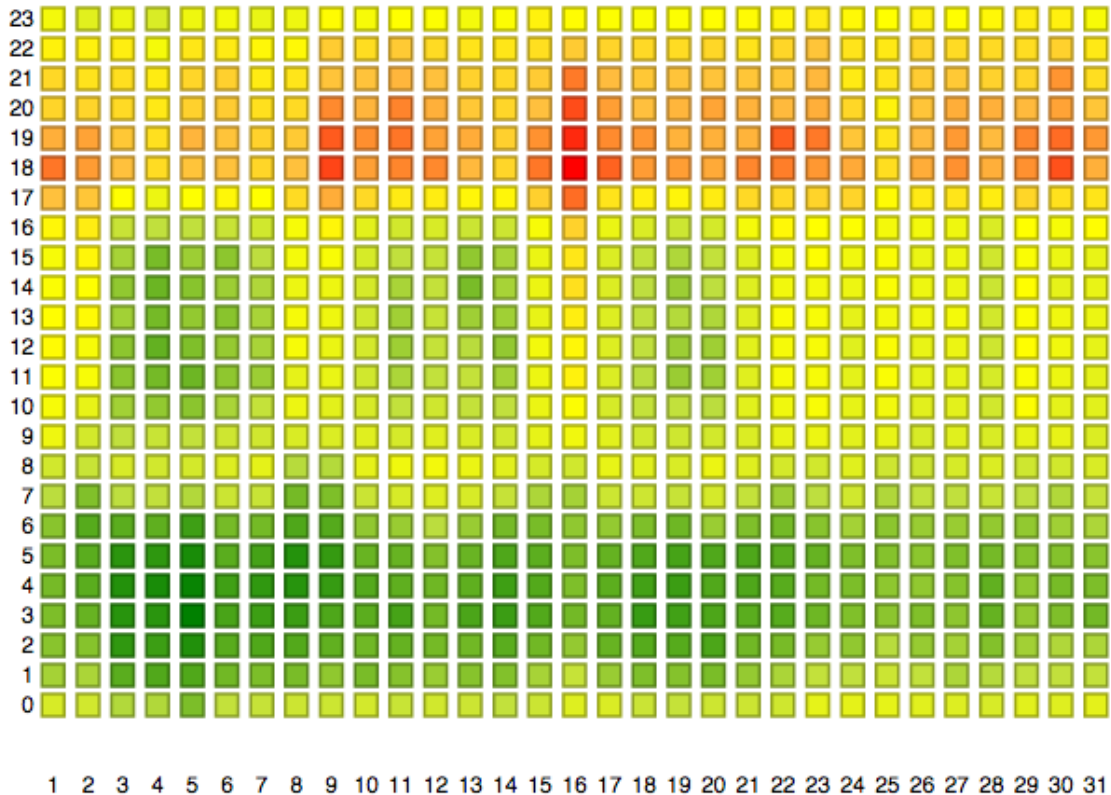
Tool:  
Processing

Source: [International collaboration project on The Role of Big Data in the Modernisation of Statistical Production](#) (overseen by the [High-Level Group for the Modernisation of Statistical Production and Services](#))

Fixed sensors

# Visualization of Canadian Data

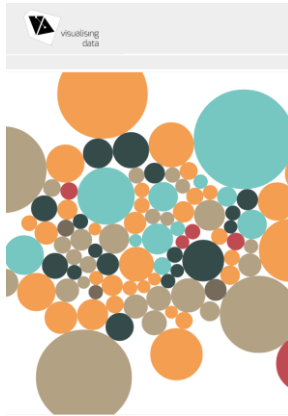
Hourly consumption per day on a single month



Tool:  
Pentaho

Source: [International collaboration project on The Role of Big Data in the Modernisation of Statistical Production](#) (overseen by the [High-Level Group for the Modernisation of Statistical Production and Services](#))

# Risorse in rete. Esempi



## 39 Data Visualization Tools for Big Data

On 02/15/14, In Cloud Computing, by Andre Lusic

Big Data is more valuable when visualized and analyzed

Data visualizations are everywhere today. From creating a visual representation of data points to impress potential investors, report on progress, or even visualize concepts for customer segments, data visualizations are a valuable tool in a variety of settings. When it comes to big data, weak tools with basic features sort out it. The following 39 tools (listed in no particular order) are some of the best, most comprehensive, sophisticated-yet-flexible visualization tools available—and all are capable of handling big data.

Many of these tools are Open-Source, free applications that can be used in conjunction with one another or with your existing design applications, using JavaScript, JSON, SVG, Python, HTML5 or drag-and-drop functionality with no programming required at all. Others are comprehensive business intelligence platforms capable of sophisticated data analysis and reporting, complete with a multitude of ways to visualize your data. Whether you need to analyze data and determine the best ways to present it to clients or partners, or you have a visual layout in mind and need a tool to bring your concept to life—there's a tool on this list to serve your needs.

► [Pomaps](#) v0.5.1.0  
► [Pomaps](#) v0.5.1.0

Pomaps is a free JavaScript library for making dynamic, interactive maps in modern web browsers

Pomaps is a project from [Impresso](#) and [Stamen](#)

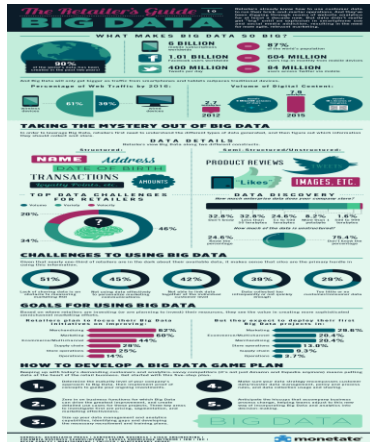
<http://www.visualisingdata.com/>

<http://blog.profitbricks.com/39-data-visualization-tools-for-big-data/>

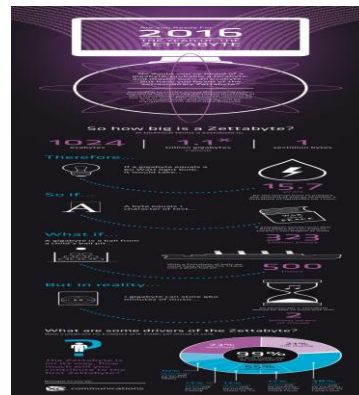
DailyInfographic

The image shows the DailyInfographic website interface. It features a grid of data visualizations and articles. The top left article is titled 'The Top 10 Most Dangerous Jobs' and is dated April 15, 2015. The top right article is titled 'Rental Home Management' and is dated April 13, 2015. Below these are two more articles: 'Overcoming The Fear Of Starting A New Business' dated March 31, 2015, and 'Service Trends Around The World' dated March 27, 2015. Each article includes a small thumbnail image and a brief description.

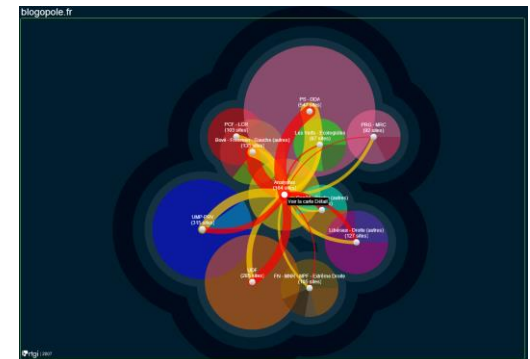
<http://www.dailyinfographic.com/>



<http://content.monetate.com/h/i/12311844-the-retailer-s-guide-to-big-data>



<http://www.dailyinfographic.com/2016-the-year-of-the-zettabyte-infographic>



<http://blogopole.observatoire-presidentielle.fr/>