

**Business Science  
for governance, competition  
and innovation**

# Linking two wor(l)ds to share value

ACADEMY  MARKET

# The duty to innovate

*“That's the way it's always been done”* is often the reason to behave in a new way

Quite always solutions:

- Are not written in manual
- Are not stored in paradigms
- Cannot be found in the past

We all have to:

- Think in a new way
- Look for correlations
- Use imagination
- Focus on continue confrontation with academic research

# Mission

- Fill the gap between the big amount of meaningful data and the power of methodology's vanguard
- Fill the gap between the data and the information needed
- Use data:
  - to forecast,
  - to deeply know your own market,
  - to intercept signals of changes ...
- To provide companies the powerful tools to guarantee their future prosperity

**A revolution is going on!**

**In this context an innovative way to  
create value**

# Context

# “Tormentoni” (~catchphrases) (annoying, repetitive, storm)

- Big Data
- Analytics
- Universal solutions (DMPs, IBM Watson ...)
- Data Science
- Look alike
- Machine learning

# Not just “big” data ...

- size (of course)
- heterogeneity
- not representative samples
- lack of the original question
- new ideas and new question (a recursive process)
- processing and delivering super-speed (aggressive real time)



... it is a matter of complex data

## Why now?

New “spontaneous” stream of data:

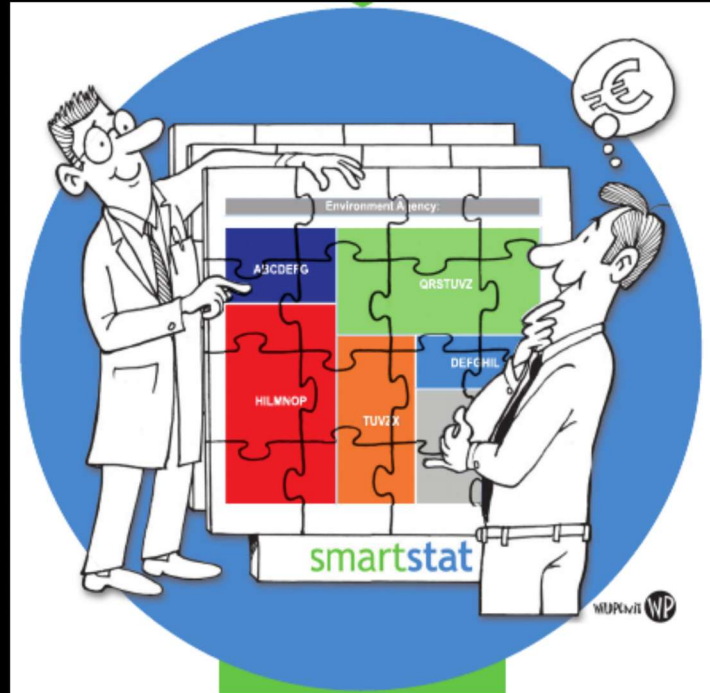
- Internet
- Mobile
- App
- Connected devices
- IoT

# Not interested in “analytics” ...

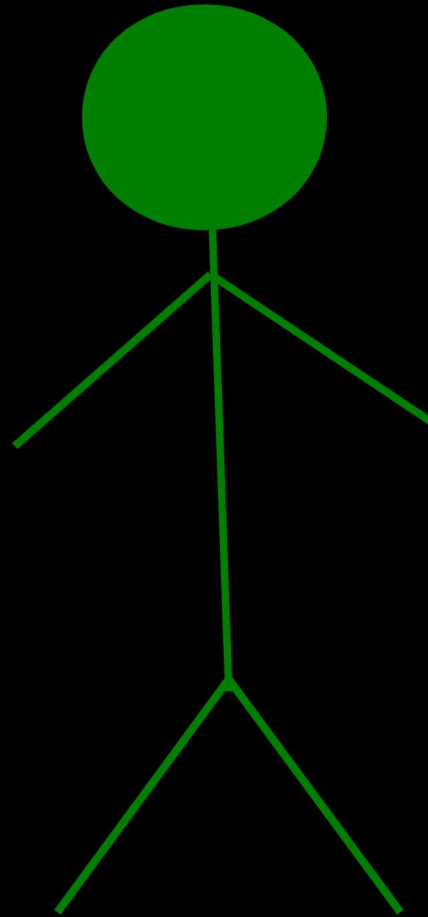


- From a lot of bit to a lot of graphs
- From a lot of bit to a lot of queries
- From a lot of bit to a lot of pivot tables, cubes ...

# ... but in stylization



- readability
- information (beyond queries)
- synthesis
- understanding



- reduction of complexity
- keep of important information

# One-size-fits-all solution is impossible

- Same dataset in different context need different approach and different models
- Same needs in the same context with different data need different approach and different models
- The difference is made mostly by how data are set (it involves the highest statistical models of the research)
- Big value comes from new source of data you “discover” (got by mixing lateral thinking and deep knowledge of the up-to-date technologies)
- The correct IT solution strongly depends on the single task (it's a matter of success vs complete failure)

# Ad-hoc approach towards verticals

- A new world to discover in front of us
- R&D is vital
- Focalization on scalable projects

# Data Science is thought as:

- Data analyst's job
- Mathematicians' job
- IT's job
- Designer's job

# Our job is dialogue with:

- reality (data represents reality)
- data (data are never “wrong”)
- time (answer to “real time”<sup>\*</sup> request)
- needs (precision is not the king)

as every scientist of every discipline knows very well!

We can better define our job as:

## Business Science

<sup>\*</sup>Real time doesn't mean “instantly” but in time accordingly to the moment you need the answer. Could be 1 ns or 1 week, all the same.



# Look alike is weak

- The idea is to try to guess your behavior from someone similar to you
- But:
  - 1) you are not equal to someone similar to you
  - 2) What if no one is similar to you?

# Future is functional profiling

- You are an unique individual, and your profile is representable by many features
- Our duty is to find the correct model to guess what you will do using your features and learning how every single feature (in relation with the others) contributes to indicate a propension to act in a particular way
- Very difficult but absolutely possible! We made it many times

# Look alike approach

## TRAINING SAMPLE

user 1

feature1

50%

feature2

0%

medium  
propension

user 2

feature1

0%

feature2

100%

high  
propension

## NEW USERS

feature1

50%

feature2

50%

?

feature1

100%

feature2

100%

?

# Functional profiling approach

## TRAINING SAMPLE

user 1

feature1

50%

feature2

0%

medium  
propension

user 2

feature1

0%

feature2

100%

high  
propension

## NEW USERS

feature1

50%

feature2

50%

high  
propension

feature1

100%

feature2

100%

superhigh  
propension

# Machine learning is not the solution

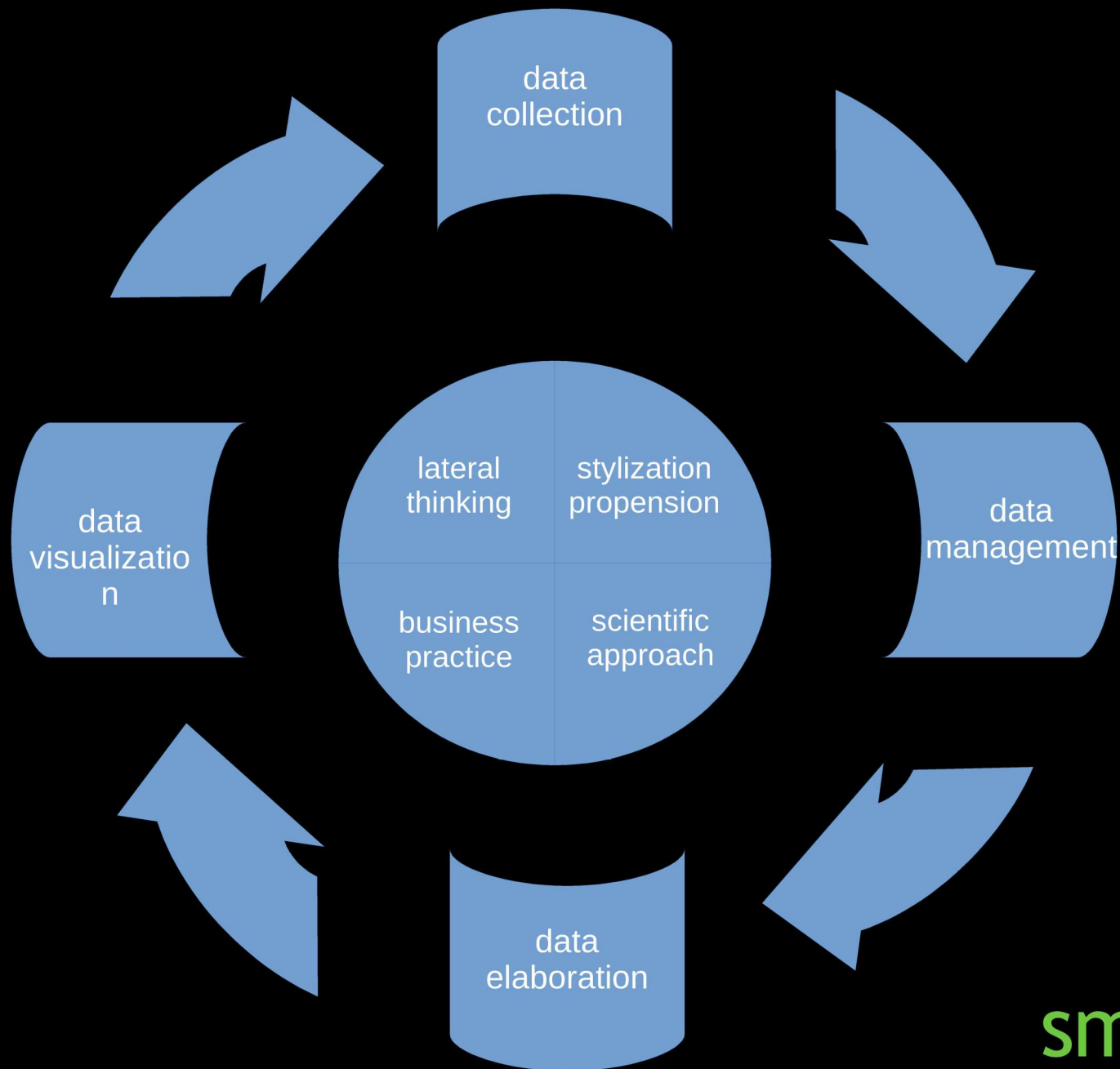
- Do what you have programmed
- To test infinite possibility you need infinite time

... but a tool to help **human** learning

- Help to check the correctness of human intuitions
- If well directed shorts time-to-solution

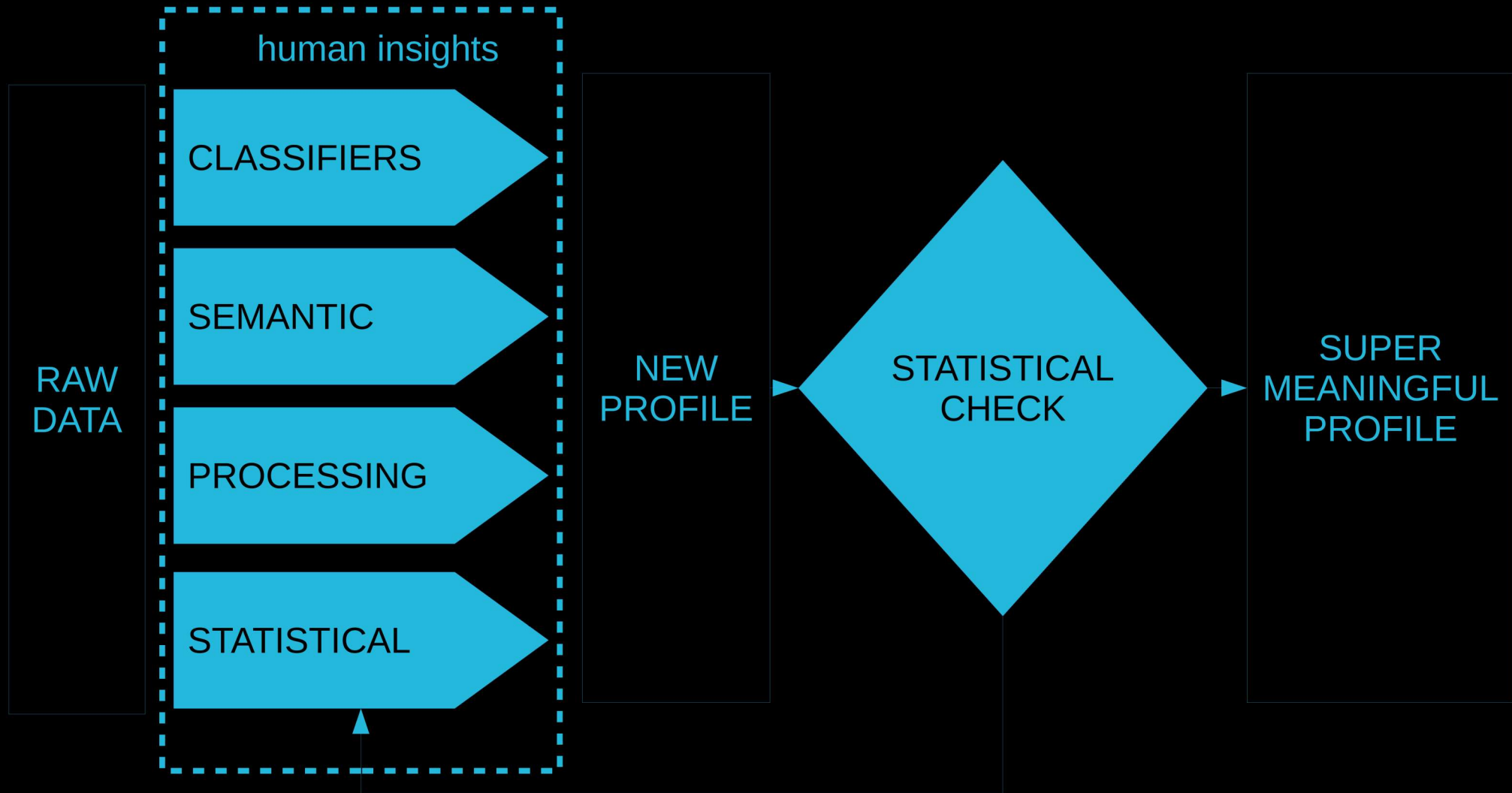
**Our contribute**

# Recursive approach





# Features quest



# What we use

- **Machine learning:** all processes are developed in solution able to continuously learn from the data flow to:
  - improve performance
  - identify seasonality
  - adapt to the continues changes
- **Artificial Intelligence:** our systems are a complex structure of software solutions capable to simulate human ability to learn from reality and to adapt to it
- **Embedding of human intelligence** know-how: use of Bayesian models to allow human to teach the engine how to better do its job
- **Neural Network:** Simulating the approach of human brain in some situation system are able to achieve impressive results in term of performance and delivery time
- **Deep learning:** with the heavy use of neural network we can easily develop powerful deep learning solution
- **Vanguard sophisticated models:** we usually develop solutions using all the bleeding edge model of the international statistics research reading the source code and modifying them according to our needs (for instance we have used genetics biostatistical models adapted to analyze cookies)

# Why Smartstat?

# Innovative approaches

- Beyond one-size-fits-all solutions -> **ad hoc solution**

Agnostic and continuously updated approach towards:

- technologies
- statistical models

- Beyond just big data -> **complex data**

Open mind approach using not "superclean" data and looking for innovative ideas to improve their informational power.

Crucial step is data assessment to feed models with meaningful data – to do this most innovative and sophisticated models and approach are needed (if not we get “garbage in - garbage out”)

- Beyond “data science” -> **Business Science**

Approach situation as a real scientists (both of the founders are graduated in Physics)

- Beyond “analytics” -> **stylization**

Always towards new way to see the data according to the sensibility, the time and the needs of management

- Beyond “look alike” -> **functional profiling**

We started from the beginning approaching our project with functional profiling

# Ad hoc: technology cornered

- old paradigms are outmoded
  - super computer to compensate software's inefficiencies
  - super reliable machines
  - one task for one machine
  - the existence of the perfect architecture able to solve all problems
- new paradigms
  - scalability
  - many low costs server
  - use of abstraction layer
  - ad hoc infrastructure according to the need of the project

# Ad hoc: new data approach

- Old principle of relational SQL database
  - normalization
  - no redundancy
  - primary key to identify tables
  - separated tables for each correlated data set
  - consistency
  - everything in a monolithic DB
- new approach
  - heterogeneous data -> tables with different fields in different records
  - real time needs -> super redundancy, not so important consistency
  - enormous amount of data -> better light than structured DB
- noSQL, many possibilities
  - key/value (Redis, MemcacheDB ...)
  - Column (Cassandra, Hbase, ...)
  - Document (mongoDB, Couchbase, ...)
  - Graph (OrientDB, Neo4j ...)

# Abstraction layers in continue evolution

- File manager
- Data manager
- Elaboration manager
- Most famous:
  - Hadoop (inspired to Google MapReduce e Google file system)
  - Storm (from Twitter)



# Ad hoc: data elaboration

- Different tools (each with different pros):
  - R (free availability of all vanguard models -> R&D)
  - Julia (future, very high performance and easy programming)
  - Python (powerful, ductile and easy integrable)
  - Java (super scalable)
- Many models from the vanguard of the international research (a lot of R&D)
  - from the simpler to humanly understand the result and improve the approach
  - to the bleeding edge model to get best performance



# The importance of Semantic

# A winning couple

- All success cases are about human behaviour studies
- The language is a representation of thinking, comprehend language means to deeply know how a person is and what he/she thinks. Use semantic data carries very rich data.
- Using semantic and sentiment analysis engines we extract profile with a very big amount of data
- So huge amount of richful data used with our vanguard approach are a very strong weapon to get success
- Facts continuously prove the rightness of my starting intuition

# Customers



la Repubblica

GRUPPO  
MONDADORI



Associazione Italiana Sommelier

FINECO  
BANK

GEDÌ  
GRUPPO EDITORIALE



QUATTORRUOTE



retelit

JCDecaux



nielsen

FANUC  
ROBOTICS

ITALTEL



CANALI



SORIN GROUP  
AT THE HEART OF MEDICAL TECHNOLOGY

AdmantX  
semantic advertising data

YOOX



PROVINCIA  
MONZA  
BRIANZA

BPM  
BANCA POPOLARE  
DI MILANO

MerloniProgetti

Tisettanta



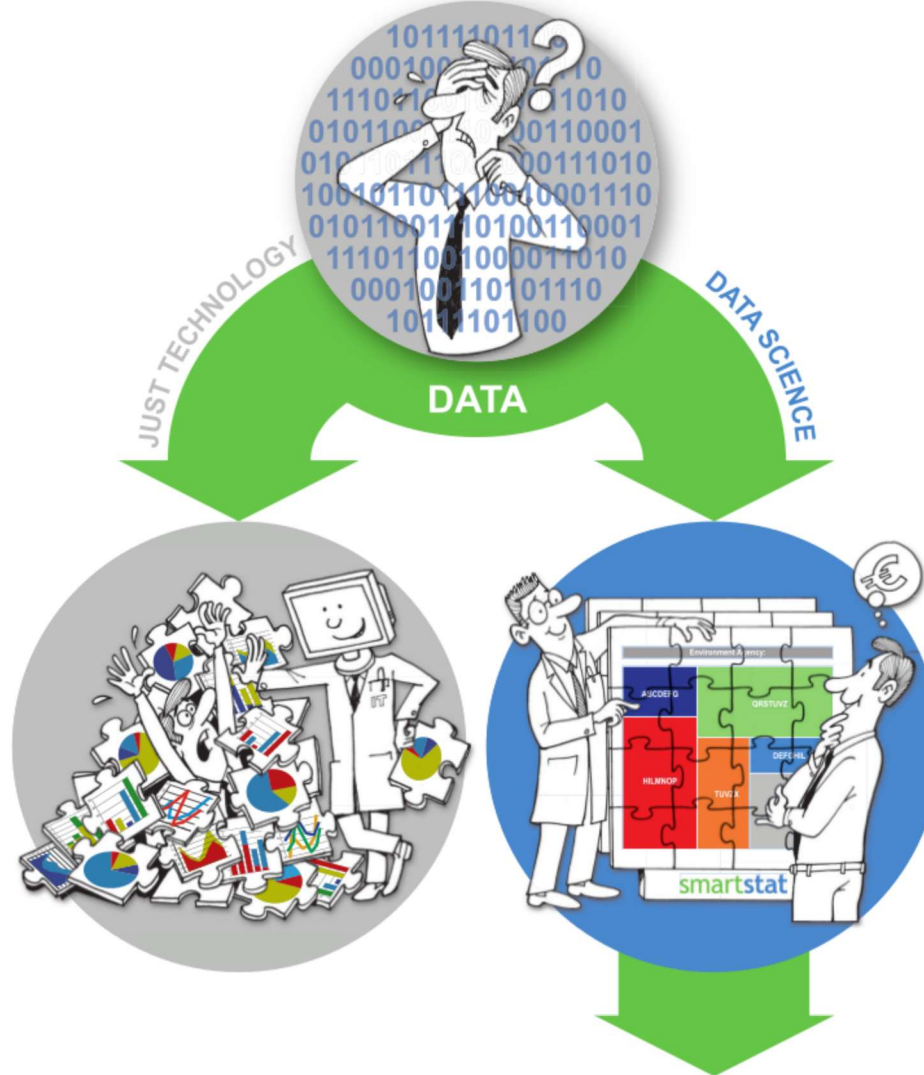
GEOMAG

sky

ENGINEERING

smartstat

# BIG DATA, BIG CONFUSION?



**smartstat**  
Data science for your business