



*cutting through complexity*

# Big Data @ KPMG

Sander Klous  
October, 2013  
CHEP 2013  
Beurs van Berlage  
Amsterdam

# Society changed within one generation



# Probability based logic reaches everybody

## Apples and Pears

Suppose we have two jars with apples and pears.

Our prior knowledge is that:

- Jar A contains 10 apples and 30 pears
- Jar B contains 20 of each

Fred picks a jar, without further evidence there is a 50% chance this is jar A (or B).

Fred pulls out a pear. The new probability that Fred picked bowl A is  $0.75 \times 0.5 / (0.75 \times 0.5 + 0.5 \times 0.5) = 0.6$

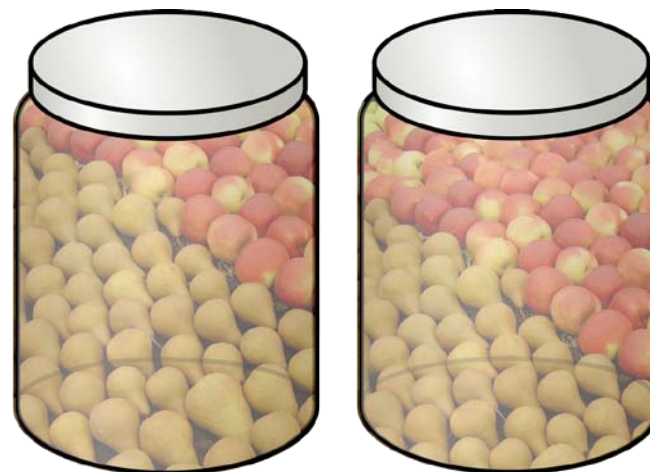
## Simpson's paradox

Although in this simple example Bayesian statistics appears to be straight forward, many subtleties arise in analysis. One of the more famous pitfalls is called Simpson's paradox:

[http://en.wikipedia.org/wiki/Simpson's\\_paradox](http://en.wikipedia.org/wiki/Simpson's_paradox)

Question:

*Do you have a higher chance of survival when falling overboard with or without a life jacket?*

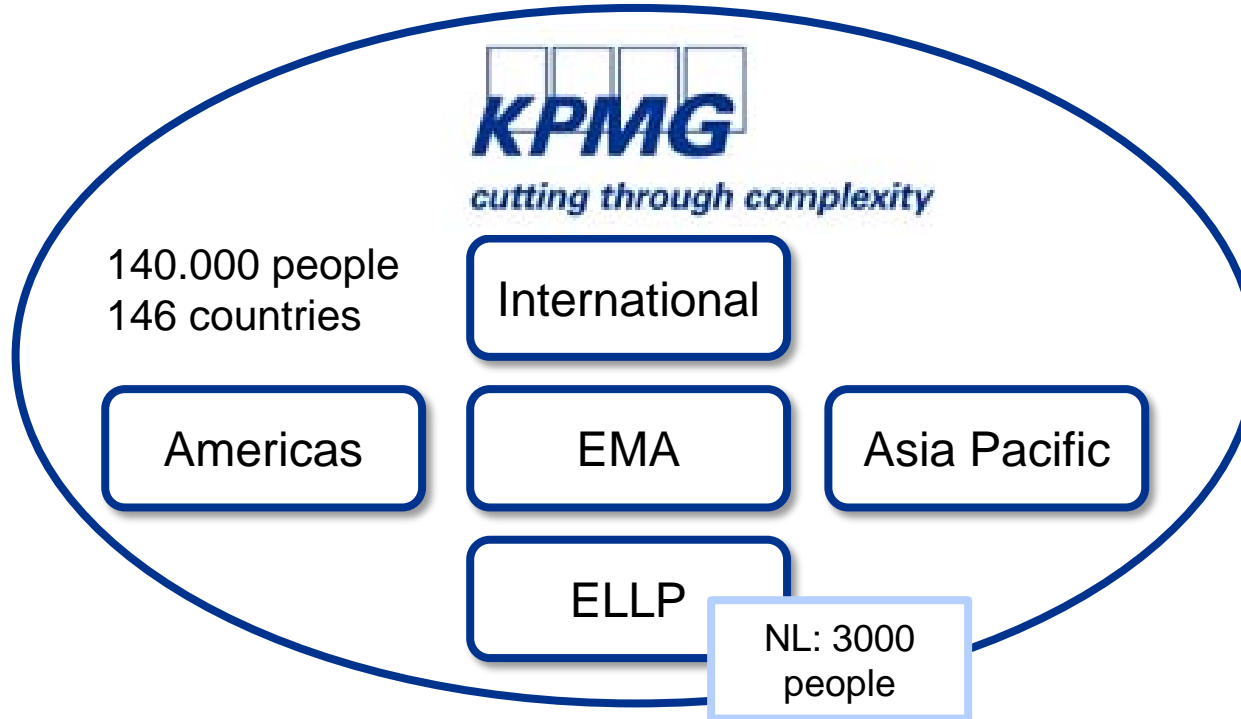


Jar A

Jar B

$$P(H_n|E) = \frac{P(E|H_n)P(H_n)}{\text{Sum}_{1}^N (P(E|H_n))}$$

# The challenge of KPMG started in 2010



- **Audit**
- **Tax**
- **Advisory**

## ***Billion dollar question: how to transform?***

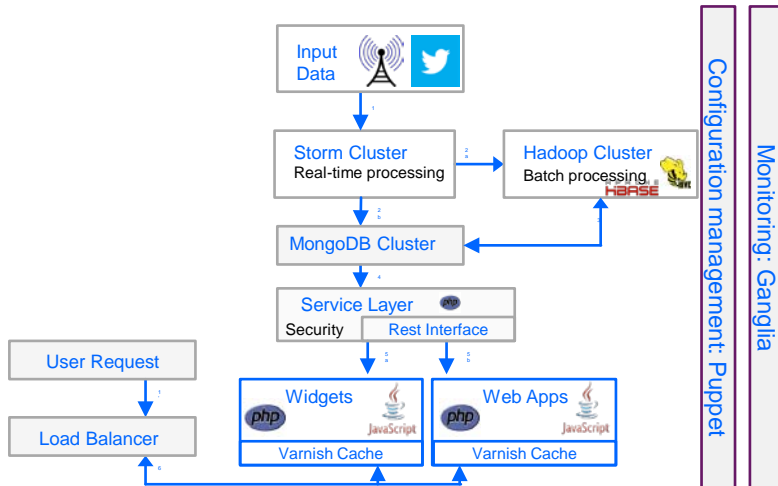
- Accountancy is about assurance
- The data driven society is about probability

## Strength and Weakness of KPMG

- Bottom up organization
- Let's contact a data scientist...

## Organized as a start-up within KPMG

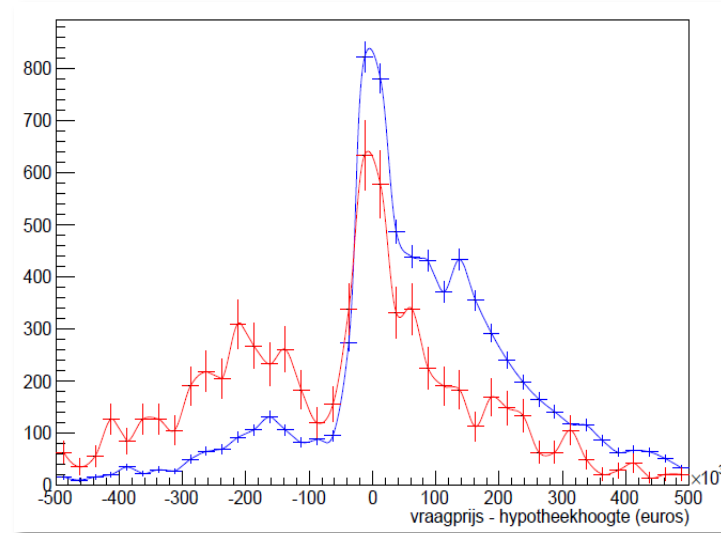
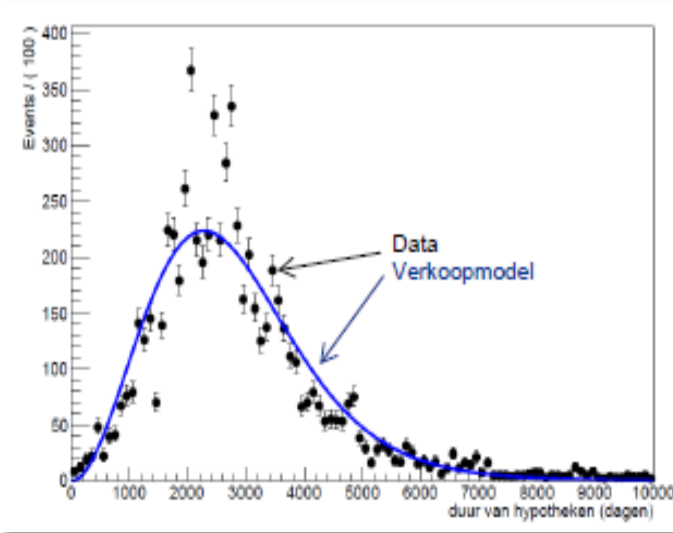
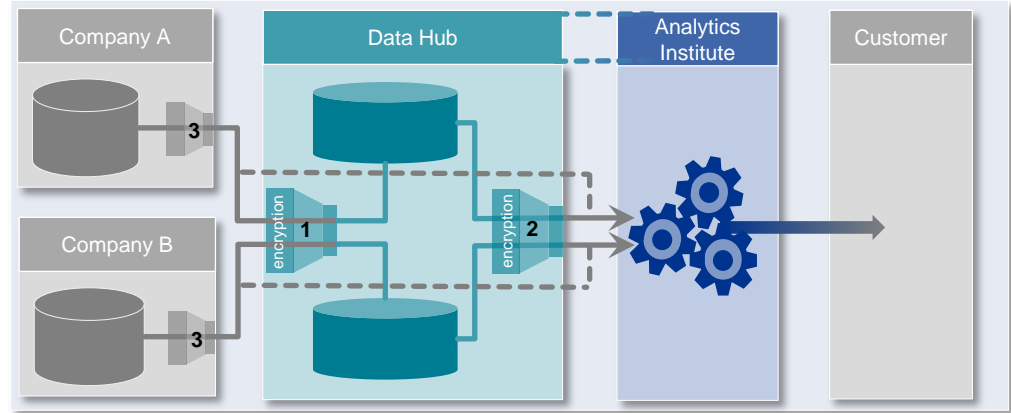
- Core team of Data Scientists
- Separated from the rest of the organization
- Our own P&L targets (*i.e.* not by the hour)
- A strong focus on improving society
- Building solutions, preferably on our own platform
- Ecosystem with partners



- **2 Vacancies**  
**(Data Scientist, Data Architect)**

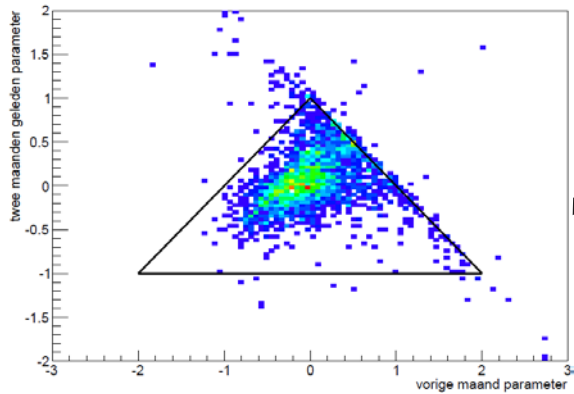
# Find statistical significant indicators

## Connecting external and internal data.

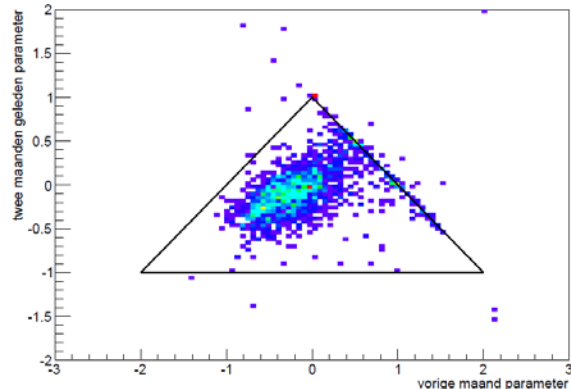


# Modeling with linear differential equations to describe behavior

## Pattern recognition, financial health prediction

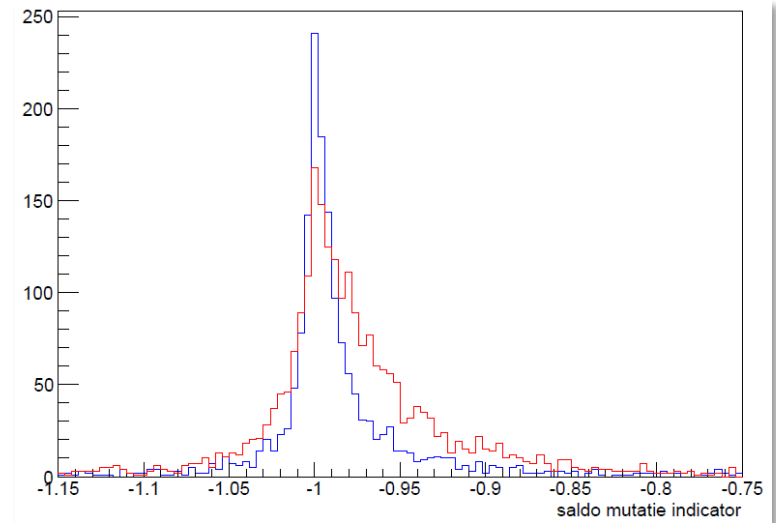


Parameters a & b towards Financial Health Dept.



Parameters a & b towards Non-Financial Health Dept.

Expenses within the range for sustainable behavior



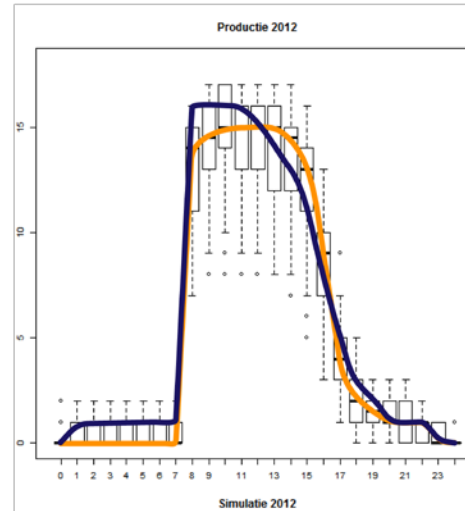
*Customers who were never considered by the Financial Health Dept. balanced their spending with their income.*

# Monte Carlo simulation to predict operational efficiency

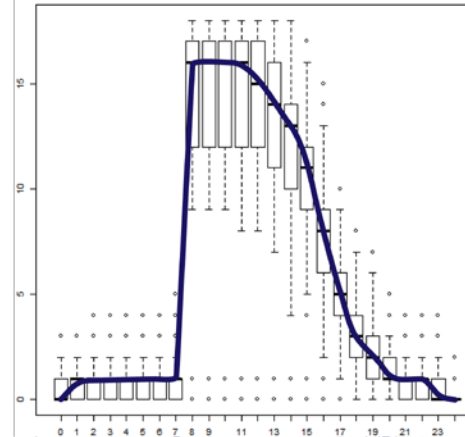
Monte Carlo simulation to analyze operational efficiency of Operation Room planning at a hospital.



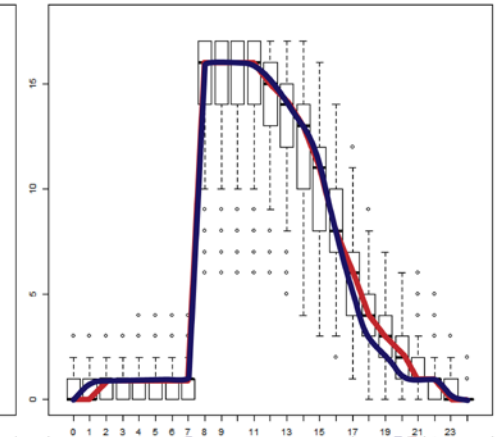
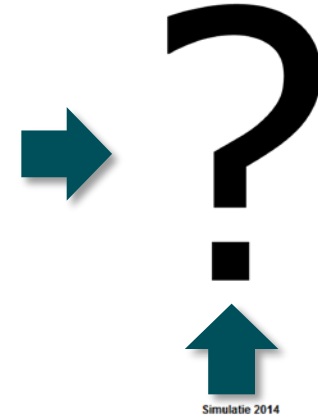
Reality



Simulation



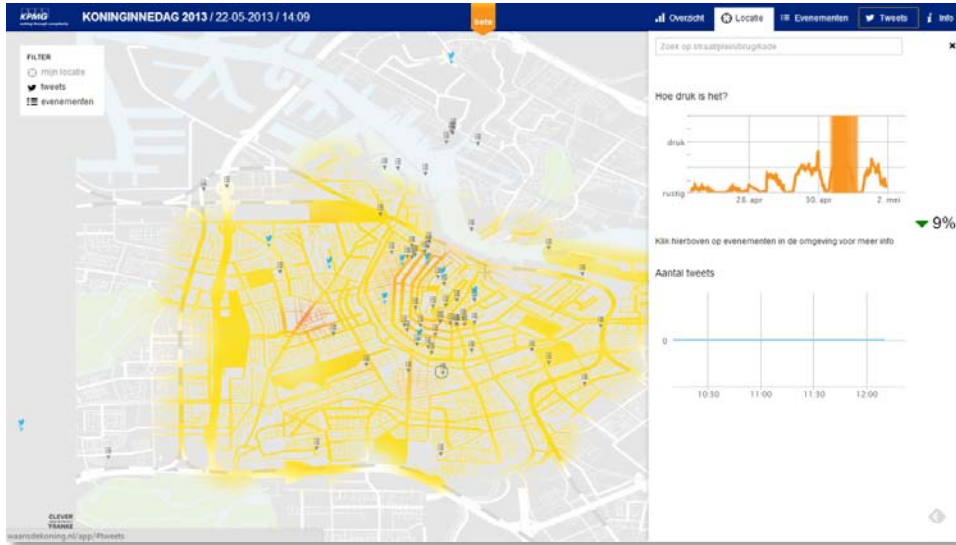
2012



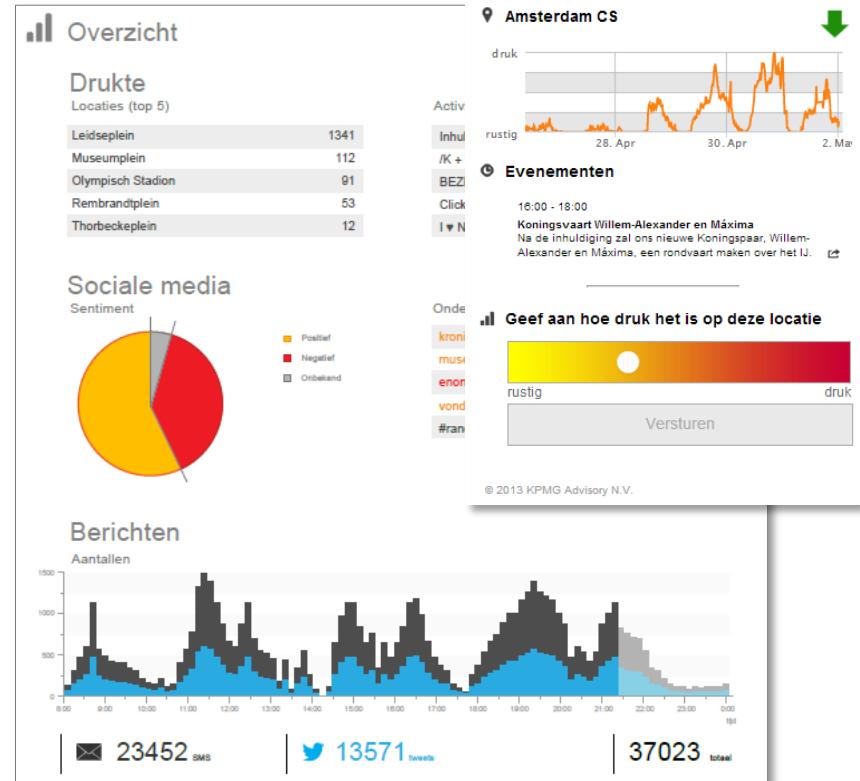
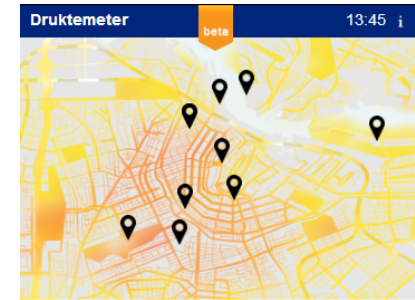
2014



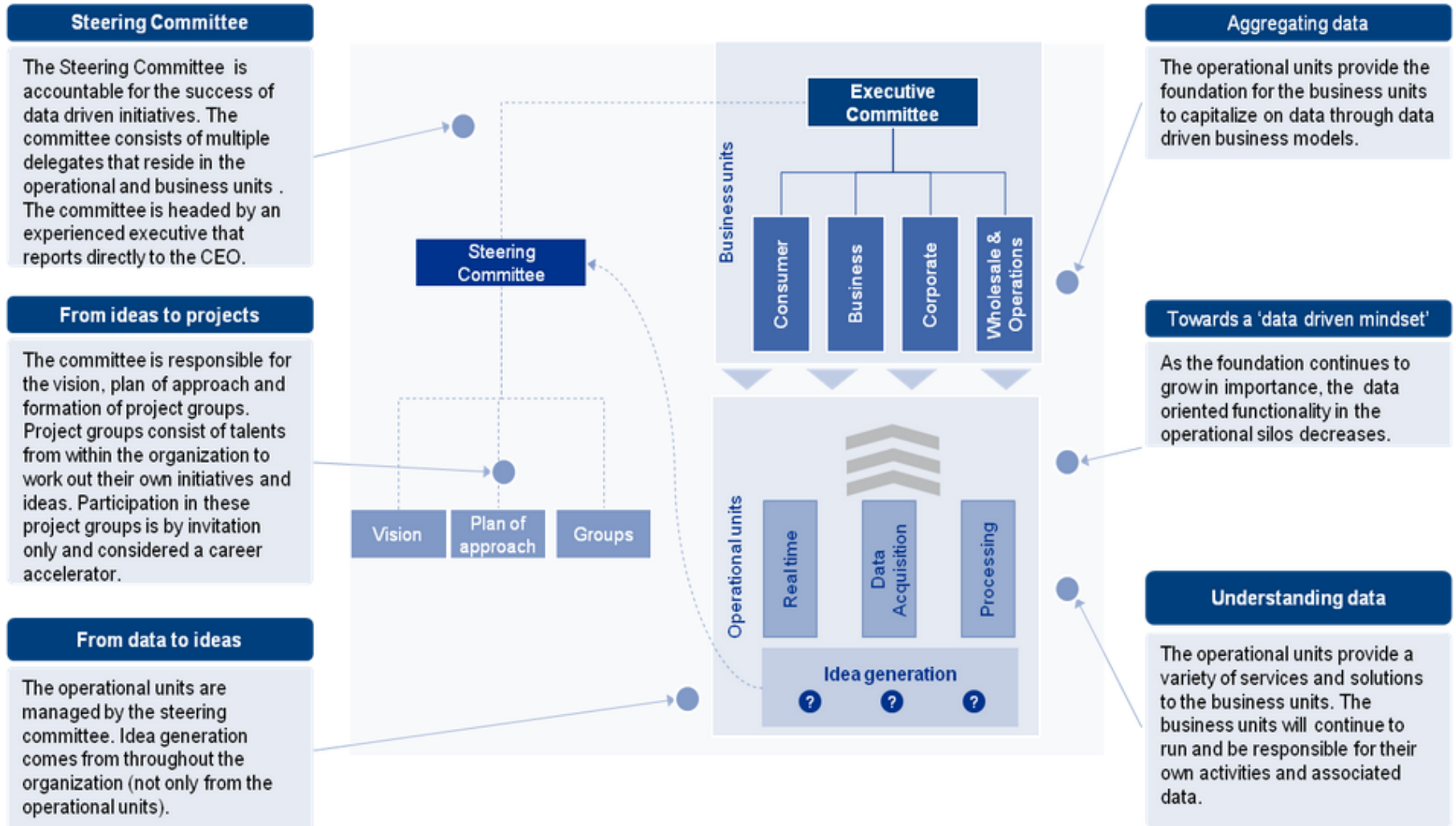
# Processing “large” amounts of data in real-time to identify threats



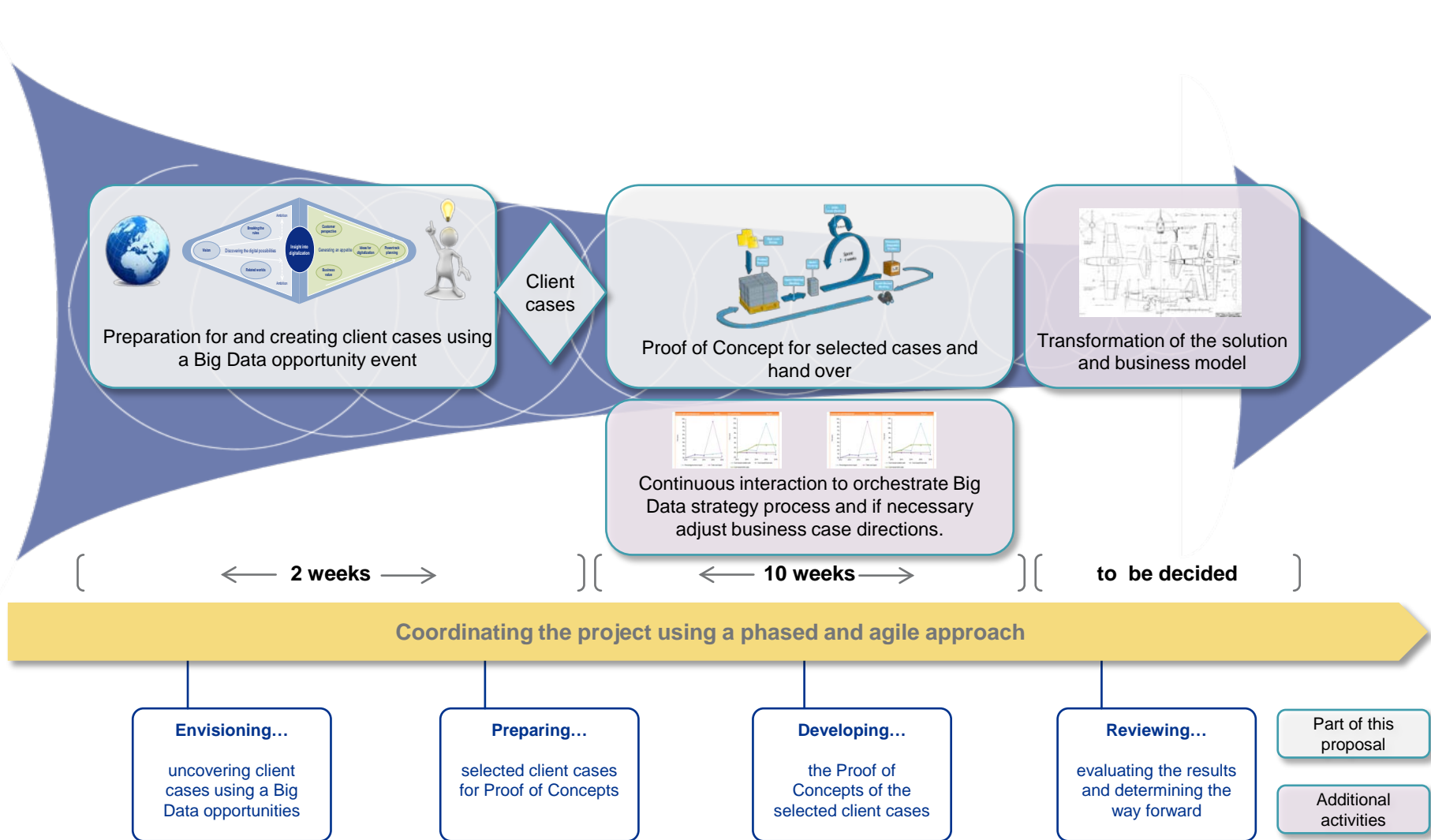
## Threat index



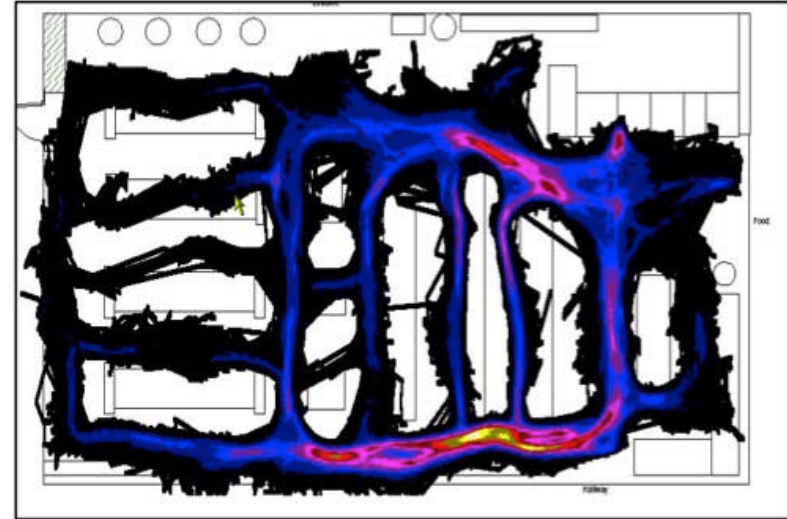
# Create data driven organizations



# Avoid an overkill of structure and hierarchy



# Our newest toy: WiFi tracking for various applications



## KPMG CHEP Challenge

- Find the iPad
- Take it home

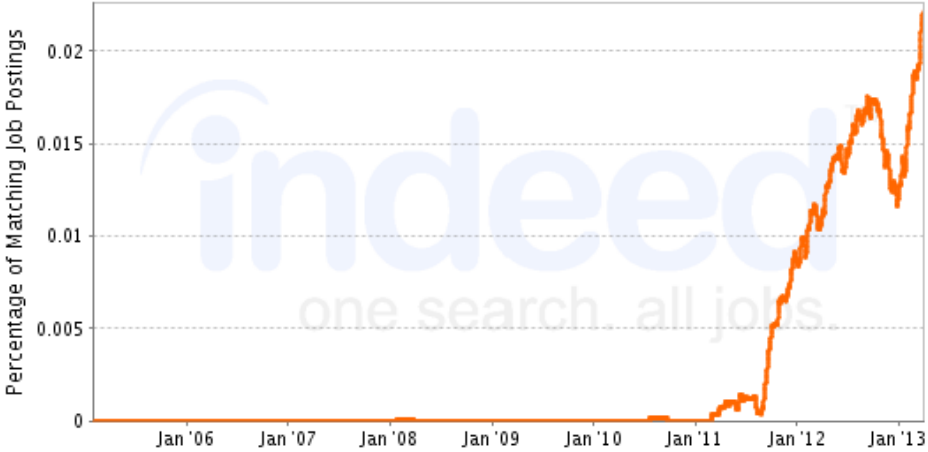


### Profile



Job Trends from Indeed.com

— "data scientist"



# Thank you



Sander Klous  
Data & Analytics lead  
+31 6 5420 8399  
[klous.sander@kpmg.nl](mailto:klous.sander@kpmg.nl)  
[s.klous@uva.nl](mailto:s.klous@uva.nl)  
[sander@nikhef.nl](mailto:sander@nikhef.nl)  
[@sanderklous](https://www.instagram.com/sanderklous)