

# DATA ANALYTICS AND BIG DATA

A HARD ENOUGH EXPLANATION FOR NON-TECHNICAL ROLES

PART **2**

## INSIDE THE ORGANIZATION

OR HOW TO LEAD A DATA-DRIVEN REVOLUTION  
IN YOUR COMPANY BALANCING BETWEEN TECH,  
BUSINESS AND DESIGN

By [TECHBIZDESIGN.COM](http://TECHBIZDESIGN.COM)

## Inside the organization we talk...

### **About Data-driven**

The maturity levels, how data-driven are you...

### **About decisions**

Decision making process, bias in decisions...

### **About Analytical principles**

### **About Analytical value**

Value propositions, value pyramid...

### **About Data governance**

Definition, maturity and CDO Role in a Data oriented company

# The Data-driven path

A Data-Driven organization is a kind of company who uses **Data Science as a core strategy** for its business

In these types of companies, data is part of the company culture, encouraging their employees curiosity and invite them to look for new ways to exploit the information available to the company.

Data Driven is committed to the **latest data analysis technology**. They're able to optimize their resources, detect market trends and make decisions much faster than a traditional company.

These companies **don't just follow the trends, they create them**



# A Data RESISTANT company – The first level

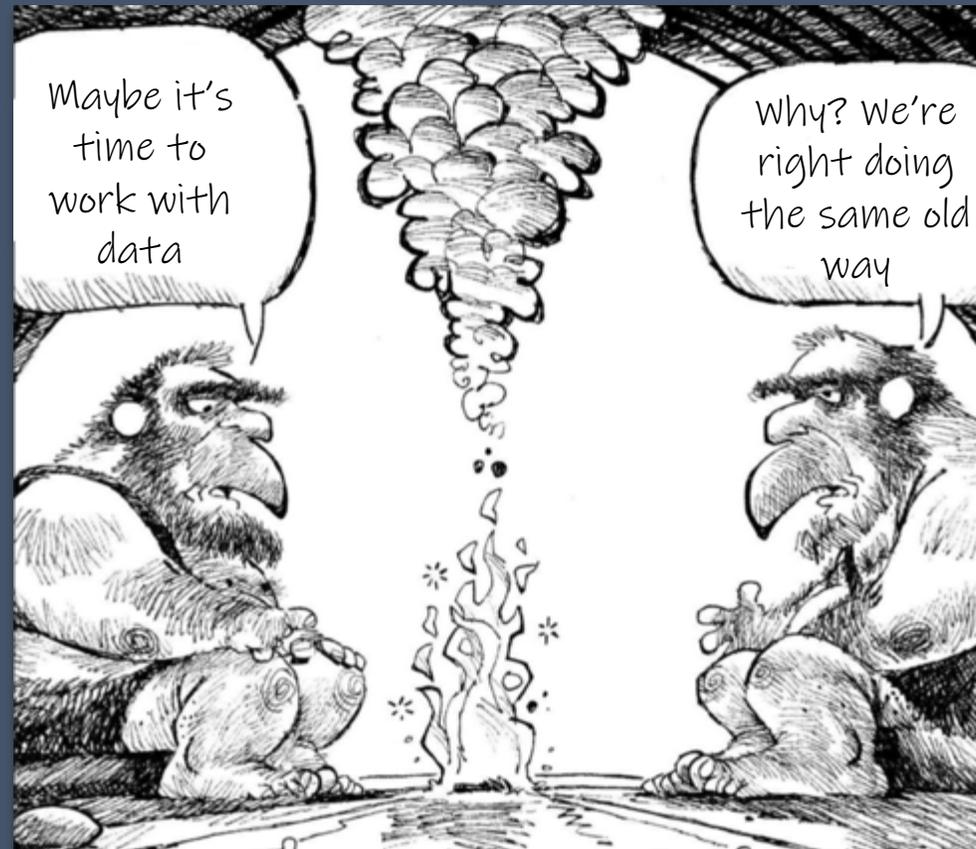
## Vision and organizational model

No specific leadership

No organizational model defined

## Investment and value generation

Their projects and initiatives are still in discussion



## Culture and talent

No analytical profiles in company

The culture is based on **intuition and experience**

## Infrastructure

**On-premise** models

Few **data internal** sources (transactional and CRM)

# A Data AWARE company – Time to think on DATA

## Vision and organizational model

Leadership is **low consolidated** and absolutely focus on reporting.

**Model not consolidated** at all. It's decentralized or controlled by a specific unit

## Investment and value generation

First **Proof of Concept** on going  
The scope is focused on a specific Business Unit



## Culture and talent

Maybe a **data scientist** in the organization. He's a **rara avis**

Basic analysis and visualization is what Data Analytics represent for the organization

## Infrastructure

**Mainly on-premise** architectures

They deal with different internal data sources

The technological stack is concentrated on **statistical and data processing software**

# A Data GUIDED company – Time to analyze the DATA

## Vision and organizational model

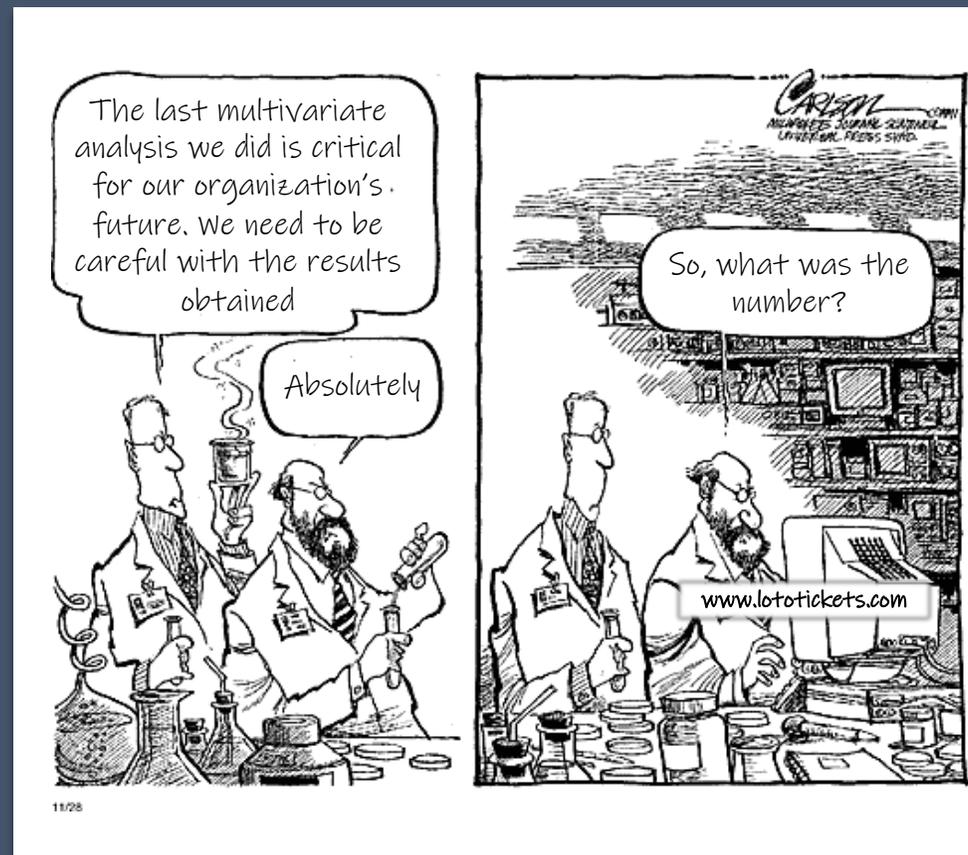
Leadership on consolidation process focus **based on data modelling**

But the model is still low consolidated. Not any defined structure

## Investment and value generation

Cross organization projects on going but in a **few Business Units**

They've got first **preliminary results**



## Culture and talent

Small **Database administration team** with data managers

They've reached basic analysis, reporting and any **multivariate model**

## Infrastructure

Mainly **hybrid architecture** with a major weight on-premise

First external data sources

Statistical and data processing software combined with **Big Data integrated packages**

# A Data SAVVY company – Looking desperately for INSIGHTS

## Vision and organizational model

They've got a **consolidated leadership** and a strong promotion of improvement and innovation initiatives

They've created **new Business lines** and the model is far consolidated

## Investment and value generation

Consolidated projects with **tangible results**



## Culture and talent

The analytical teams are involved in support to Business and decision-taking processes

They've implemented **advanced multivariate models** and some **predictive models**

## Infrastructure

Mainly **hybrid architecture** with a major weight cloud

Common use of **external data sources**

Data exploitation with script and programming languages

# A Data DRIVEN company – When STRATEGY depends on Data

## Vision and organizational model

They've got a consolidated leadership, **decisive in organization digital strategy and transformation**

Model fully consolidated

## Investment and value generation

**Multiple consolidated projects** and systems in real-time and automatize in production environment



## Culture and talent

The analytical teams are fully involved in the organization

They've reached **different business analytical stages**

## Infrastructure

Hybrid architectures **mostly cloud**

The multiple external data sources are fully integrated

The organization has a **complete usage of technology** depends on analytical domain

# So, what Data-driven level are you?

	Data RESISTANT	Data AWARE	Data GUIDED	Data SAVVY	Data DRIVEN
Vision and organizational model	<p>No leadership</p> <p>No organizational model defined</p>	<p>Low consolidated</p> <p>Model not consolidated</p>	<p>Based on data modelling</p> <p>Model low consolidated</p>	<p>Consolidated leadership</p> <p>New Business lines</p>	<p>Decisive in organization digital strategy and transformation</p> <p>Model consolidated</p>
Investment and value generation	<p>Projects in discussion</p>	<p>Proofs of Concept</p>	<p>Few Business Units pushing</p> <p>Preliminary results</p>	<p>Tangible results</p>	<p>Multiple consolidated projects in production</p>
Culture and talent	<p>Intuition and experience</p>	<p>One data scientist as a rare avis</p>	<p>DBA team with data managers</p> <p>Multivariate model</p>	<p>Advanced multivariate models</p> <p>Predictive models</p>	<p>Multiple Business analytical stages reached</p>
Infrastructure	<p>On-premise</p> <p>Internal data</p>	<p>Mainly on-premise</p> <p>Statistical and data processing software</p>	<p>Hybrid architectures</p> <p>Big Data integrated packages</p>	<p>Hybrid architecture (cloud)</p> <p>External data sources</p>	<p>Mostly cloud</p> <p>Complete usage of technology</p>

**Are there many data-driven companies? Why do they want to be one of them?**

# The Spanish case

Companies are aware of Data Analytics and Big Data potential. But they are **far from a fully adoption**.

Investment is arriving to Data initiatives. But generally people are working **on Proofs of Concept**.

**Talent is rare** and hard to obtain.

Predominance of **on-premise architectures**, but with a strong **cloud presence**. At the same time, usage of programming for analysis and tools for visualization

# 40%

Don't have a specific leadership role in Big Data and Advanced Analytics or they just contracted one



Two main reasons for start Data transformation in companies

# 68%

Customer relationship

# 50%

Process automation

# 85%

Have started succeed PoCs or projects beyond other transformation projects' average rate



# 1,8M€



Yearly budget of companies with more than 200M in revenue

# 55%

Consider they don't have a data-driven decision culture



# 7,7



On average difficulties to contract an analytical profile (10 is very difficult)



Use CLOUD partial or completely to data storage

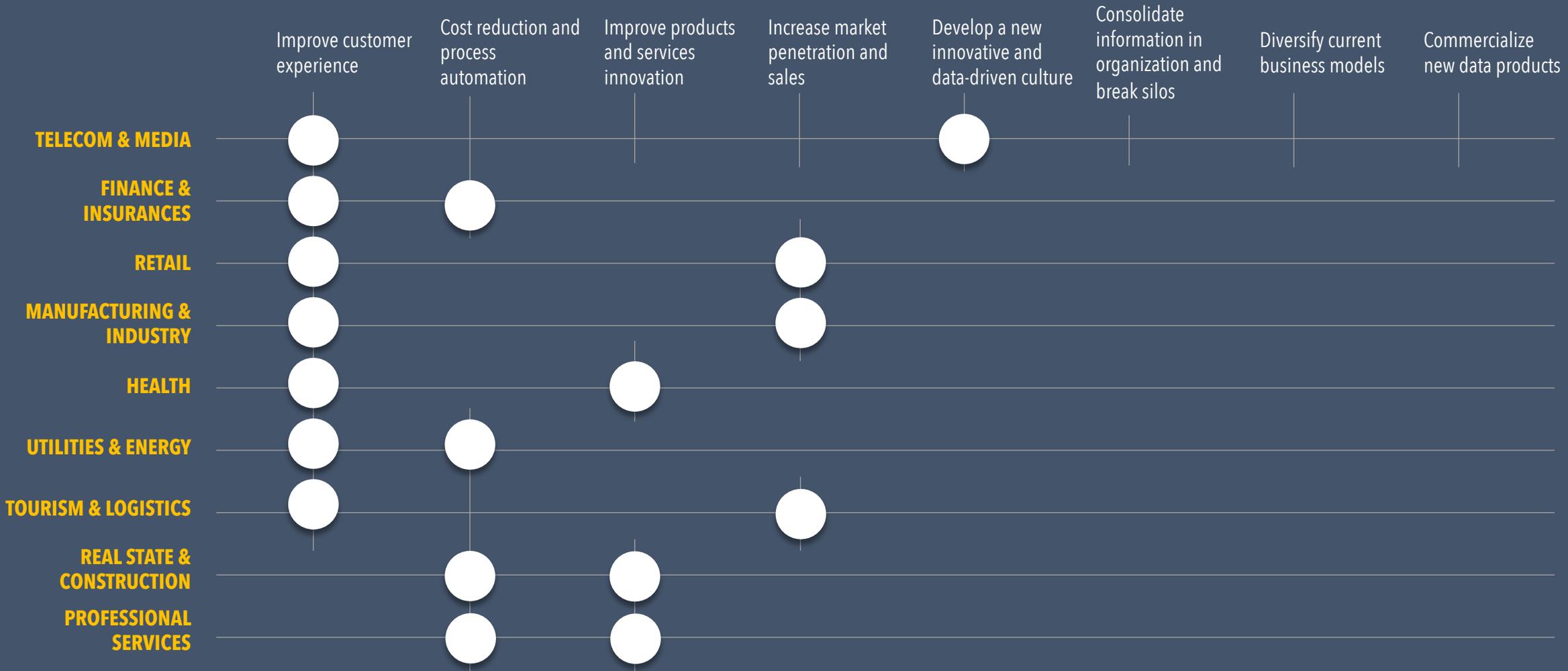
# >250%

Of the forecasted budget increase for projects and people dedicated to Big Data in next

# 3 years

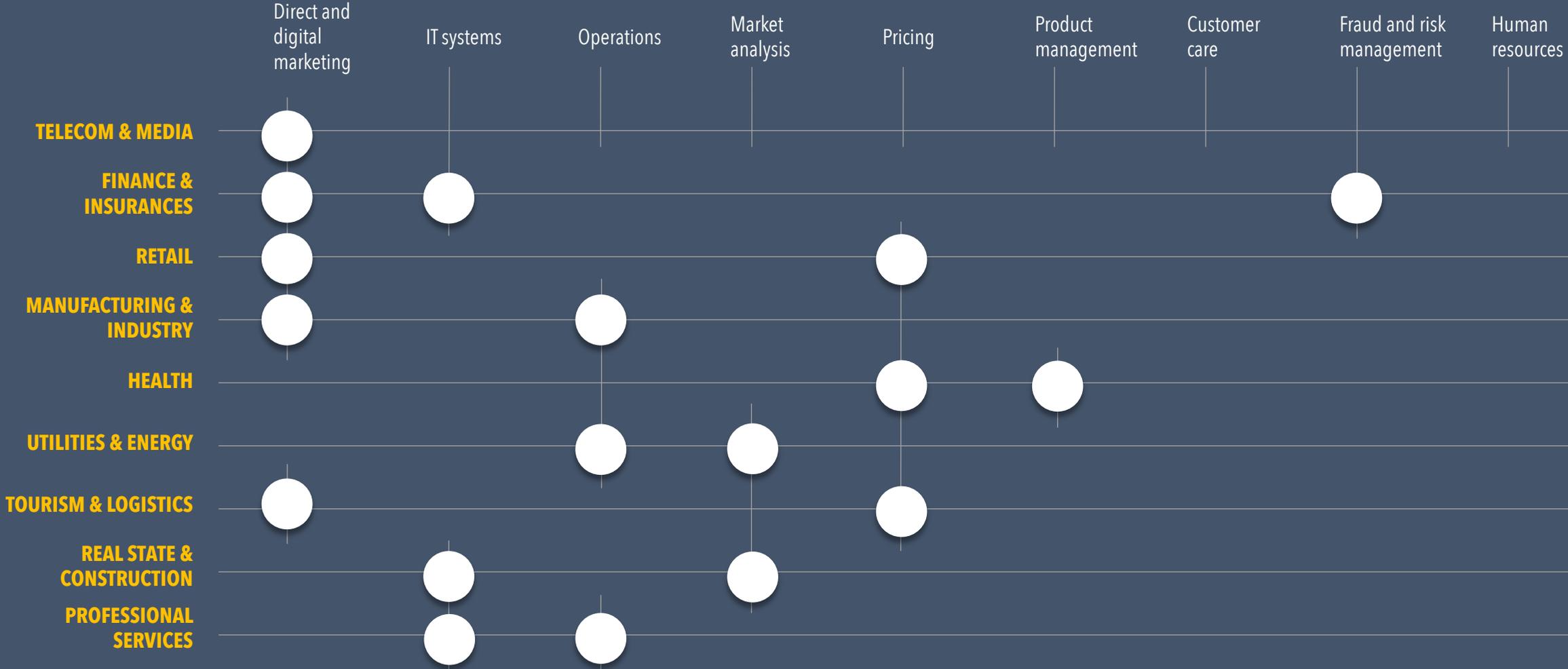


# Improve customer experience is the first motivation to be a Data-driven company



Source: Report "Adopción e impacto del Big Data y Advanced Analytics en España" – ESADE 2018

# Digital marketing is the most analytical demand from Business Units



Source: Report "Adopción e impacto del Big Data y Advanced Analytics en España" – ESADE 2018

**How can I use Data to take the right decisions?**

# Data and emotion – The situations your company has to deal with

## Simple

The relationship between cause and effect is **obvious and proven**. Best practices exist, and should be researched and applied.

For example, it's obvious that when healthcare providers in hospitals wash their hands frequently, this reduces the spread of disease



## Complicated

The relationship between cause and effect **requires analysis or some other form of investigation** and/or the application of expert knowledge.

For example, web developers have to test various website designs to see which result in fewer abandoned shopping carts



## Complex

The relationship between cause and effect is **unknown**. We can see it only in retrospect, not in advance.

For example, pharmaceutical research teams have to explore multiple compounds to determine their impact on disease conditions.



## Chaotic

The relationship between cause and effect is **unknowable**. There is no discernible relationship between cause and effect, and there may never be.

This, for example, is the situation facing medical emergency or senior leadership teams experiencing an unanticipated disruption to their supply chain.



# Data and emotion – How to act in front of each situation

## Simple

**Automate?** Yes

**Frequency:** Every day

**Quantity:** Many

**Relation between cause and effect:** Known to all in advance

**Approach:** Sense, categorize and respond

**Technique:** Best practice

**Inputs & variables:** Few

**Algorithm or heuristics?** Algorithm

**Data or instinct?** Data

**Who does it best?** Machines

**Tactical, strategic or operational:**

Operational and tactical

**Examples:** Interest calculation, credit scoring

## Complicated

Yes

Frequent

Many

Requires analysis but can be known

Sense, analyze and respond

Good practice

Start with lots, end with less

Combination

Both

Machines plus human experts

Operational and tactical

Predicting employment or academic success

## Complex

Partially

Frequent

Many

Only known in retrospect

Probe, sense and respond

Emergent practice

Many

Combination

Both

Human experts

Tactical or strategic

Developing software

## Chaotic

No

Frequent

Many

Not discernible

Act, sense and respond

Novel practice, experimentation

Many and interacting

Heuristics

Emotion

Human experts with emotional intelligence plus machines

Strategic

Responding to crises

## Data and emotion – Dealing with fear and anger

Daniel Kahneman's bestseller called "**Thinking, fast and slow**"

System 1 based on intuition – Fast, automatic, frequent, emotional, stereotypic, subconscious

System 2 based on reasoning – Slow, effortful, infrequent, logical, calculating, conscious

Applied to evidence-based business scenarios, system 1 means that **any interaction that conveys an implied criticism of a person's current work can lead to immediate resistance**. This is the source of employee's defensive and unresponsive attitudes

Data may be nonjudgmental and unemotional, but when people are put in the mix, the results are far from guaranteed



**That's none of my business syndrome**

# Data and emotion – Cognitive bias and logical fallacies

## Cognitive bias

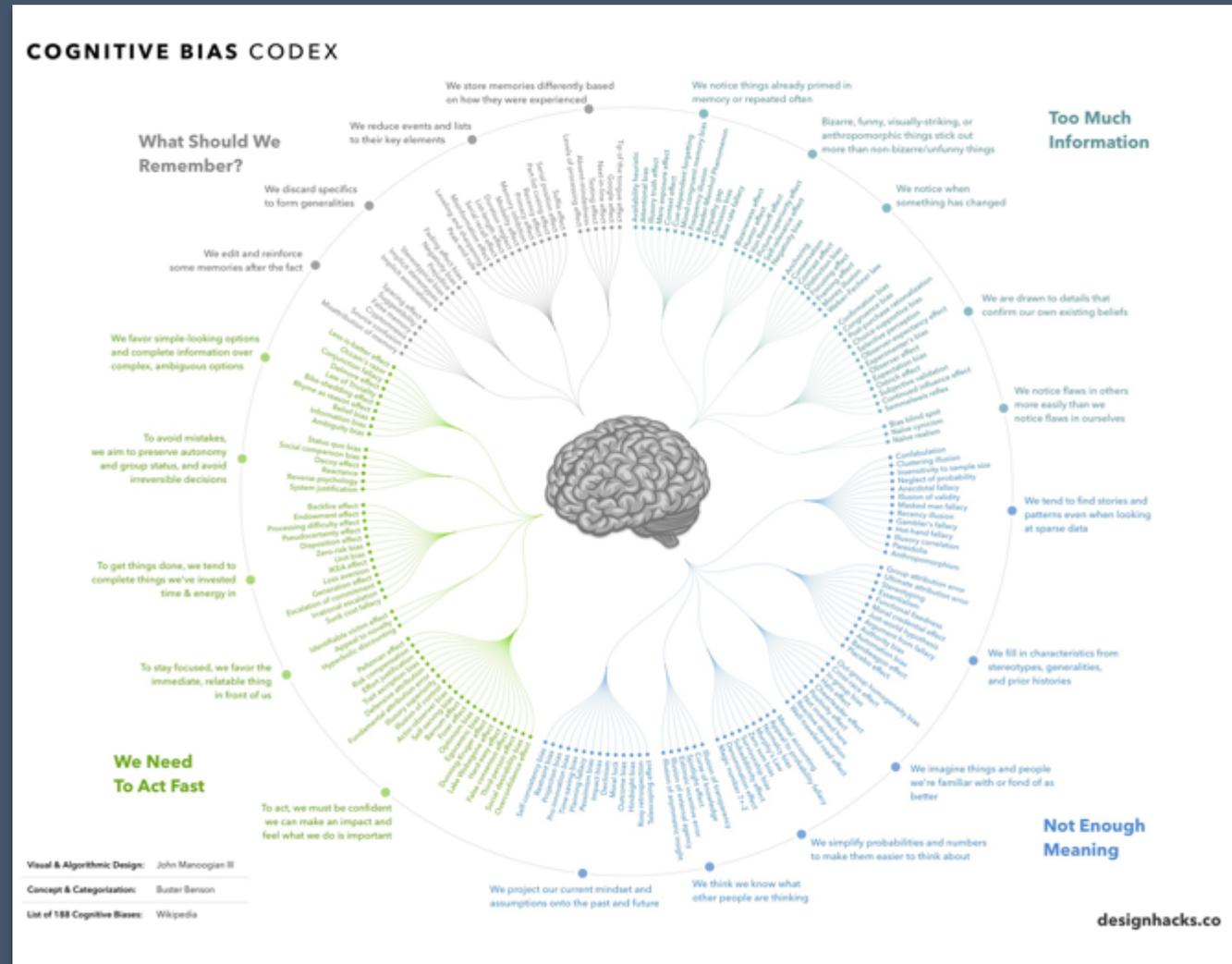
Cognitive biases are our default inclinations towards certain kind of thinking when we interpret and process information from the world around us. Everyone has biases and they affect greatly our behavior, decisions and judgements.

It's about **INTERPRETATION**

## Logical fallacies

Logic fallacies are mistakes in arguments - done deliberately or by accident. These mistakes include misuse of evidence, misuse of language, erroneous line of reasoning and distortion of an issue. Logical fallacies are "nonsequiturs", meaning that the conclusion doesn't follow from what preceded it.

It's about **COMMUNICATION**



# Data and emotion – Some famous and dangerous Bias

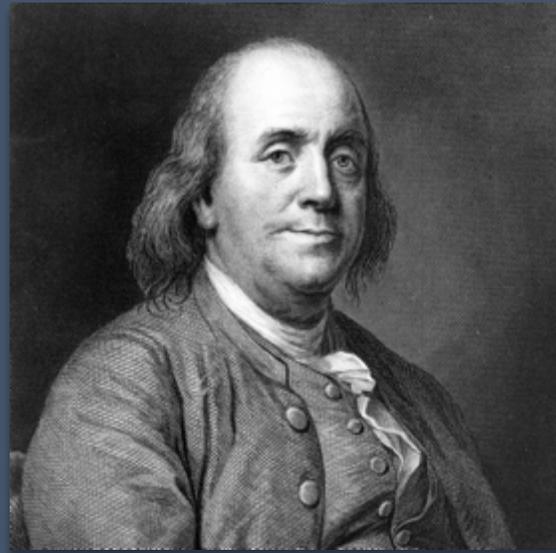
## Overconfidence

Also called the **Dunning-Kruger effect**. The more ill-informed and incompetent someone is, the more confident they are of their position. In contrast, the more informed a person becomes, the less confident they tend to become.



## Benjamin Franklin effect

People reason that they help others because they like them, even if they do not, because their minds struggle to **maintain logical consistency between their actions and perceptions**.



## Validation by frequency effect

The more times you listen to a fact, even if it is a lie, the more predisposed it is to validate it. The basis of the current trend of **FAKE NEWS**



## Peltzman effect

Better to lower a small risk to zero than larger profit assuming risk or the the tendency to take **greater risks** when perceived **safety increases**.



# Data and emotion – Other cognitive bias or the things you probably do

## 1. Anchoring Bias

People are **over-reliant on the first piece of information** they hear.

In a salary negotiation, whoever makes the first offer establishes a range of reasonable possibilities in each person's mind

## 2. Availability heuristic

People **overestimate the importance of information is available** to them.

A person might argue that smoking is not unhealthy because they know someone who lived 100 and smoked three packs a day

## 3. Bandwagon effect

The probability of one person **adopting a belief increases** based on the number of people who hold that belief.

This is a powerful form of groupthink and is reason why meetings are often unproductive

## 4. Blind-spot

Failing to **recognize your own cognitive biases** is a bias itself.

People notice cognitive and motivational biases much more in others than themselves

## 5. Choice-supportive

When you choose something, you tend to **feel positive about** it, even if that choice flaws.

Like how you think your dog is awesome, even if it bites people every once in a while

## 6. Clustering illusion

This is the tendency to **see patterns in random events**. It's key to various gambling fallacies.

Like the idea that red is more likely to turn up on a roulette table after a string of reds

## 7. Confirmation

We tend to **listen only to information that confirm our preconceptions**.

One of the many reasons it's so hard to have an intelligent conversation about climate change

## 8. Conservation

Where people **favor prior evidence over new evidence** or information that has emerged.

People were slow to accept that the Earth was round because they maintained their early understanding

## 9. Information

The tendency to **seek information when it doesn't affect action**. More information is not always better.

With less information, people can often make more accurate predictions

## 10. Ostrich effect

The decision to ignore dangerous or negative information by burying one's head in the sand, like an ostrich.

Research suggests that investors check the value of their holdings significantly less often during bad markets

# Data and emotion – Other cognitive bias or the things you probably do

## 11. Outcome

Judging a decision **based on the outcome** rather than how exactly the decision was made in the moment.

Just because you won a lot of money in Vegas doesn't mean gambling your money was a good decision

## 12. Zero – risk

Sociologists have found that **we love certainty** even it is counterproductive.

Eliminating risk entirely means there is no chance of harm being caused.

## 13. Placebo effect

When **simply believing** that something will have a certain effect on you causes it to have that effect.

In medicine, people given fake pills often experienced the same physiological effects as people given the real pills.

## 14. Pro-innovation

When a proponent of an innovation tends to **overvalue its usefulness** and undervalue its limitations.

Sounds familiar Silicon Valley?

## 15. Recency

The tendency to weight the **latest information more heavily** than older data.

Investors think the market will always look the way it looks today and make unwise decisions.

## 16. Salience

Our tendency to focus on the **most easily recognizable features** of a person or concept.

When you think about dying, you might worry about being mauled by a lion, as opposed to what is statistically more likely, like dying in a car accident.

## 17. Selective perception

Allowing our expectations to influence **how we perceive the world**.

An experiment involving a football game between students from two universities showed that one team saw the opposing team commit more infractions.

## 18. Stereotyping

Expecting a group or a person to have **certain qualities without having real information** about the person.

It allows us to quickly identify strangers as friends or enemies, but people tend to overuse and abuse it.

## 19. Survivorship

A error that comes from **focusing only on surviving examples**, causing us to misjudge a situation.

For instance, we might think that being an entrepreneur is easy because we haven't heard of all those who failed.

## 20. Barnum effect

People accept general and vague information when they are led to believe that the **information has been specifically tailored** to be about them.

# Data and emotion – Some relevant logical fallacies

## 1. Argument from incredulity

This is basing **an argument on personal disbelief**.

This provokes that a decision cannot go forward because it contradicts the point of view of the misinformed individual.

## 2. Correlating proven causation

This is a faulty assumption that a correlation between two variables implies that **one causes the other**.

So, the action is taken based on a mistaken premise.

## 3. Base rate fallacy

This is making a probability judgment **based on conditional relationships**, without taking into account the effect of prior status.

The starting conditions are misjudged, leading to unreasonable or invalid expectations.

## 4. False authority

This is **using an expert of dubious credentials**, or using only one opinion to sell a product or idea.

Wrong solutions are adopted for the wrong problems because they were recommended by an "authority."

## 5. Anecdotal fallacy

People **use a personal experience** or an isolated example, instead of reasoning or compelling evidence. An anecdote is not data.

This extrapolates from a small set of anecdotes

## 6. False dichotomy

**Two alternative statements** are held to be the only possible options, when in reality, there may be other options.

Leaders force a decision without looking at other alternatives..



**Fallacies have a more negative connotation than bias**

**Which are the rules or principles to go in the right direction? How can I get the value from Data?**

# Mantras for a Data-driven organization



## Data driven

Adopt an **experimental mindset**

1. Challenge data-decision making
2. Anticipate unintended consequences of metrics
3. Experiment with new data and analytical models



## Privacy

Be the guide an **guardian** of your customer's data

1. Privacy, as we known, it no longer exist
2. Leverage transparency to create trust
3. Treat customers' data as they want it to be treated



## Artificial Intelligence

**Automate** the manual | Promote the **creative**

1. Algorithmic decision making is inevitable
2. Focus on augmentation, not replacement
3. Play to the respective strengths of AI and humans

# Connecting Data and Analytics to Business value

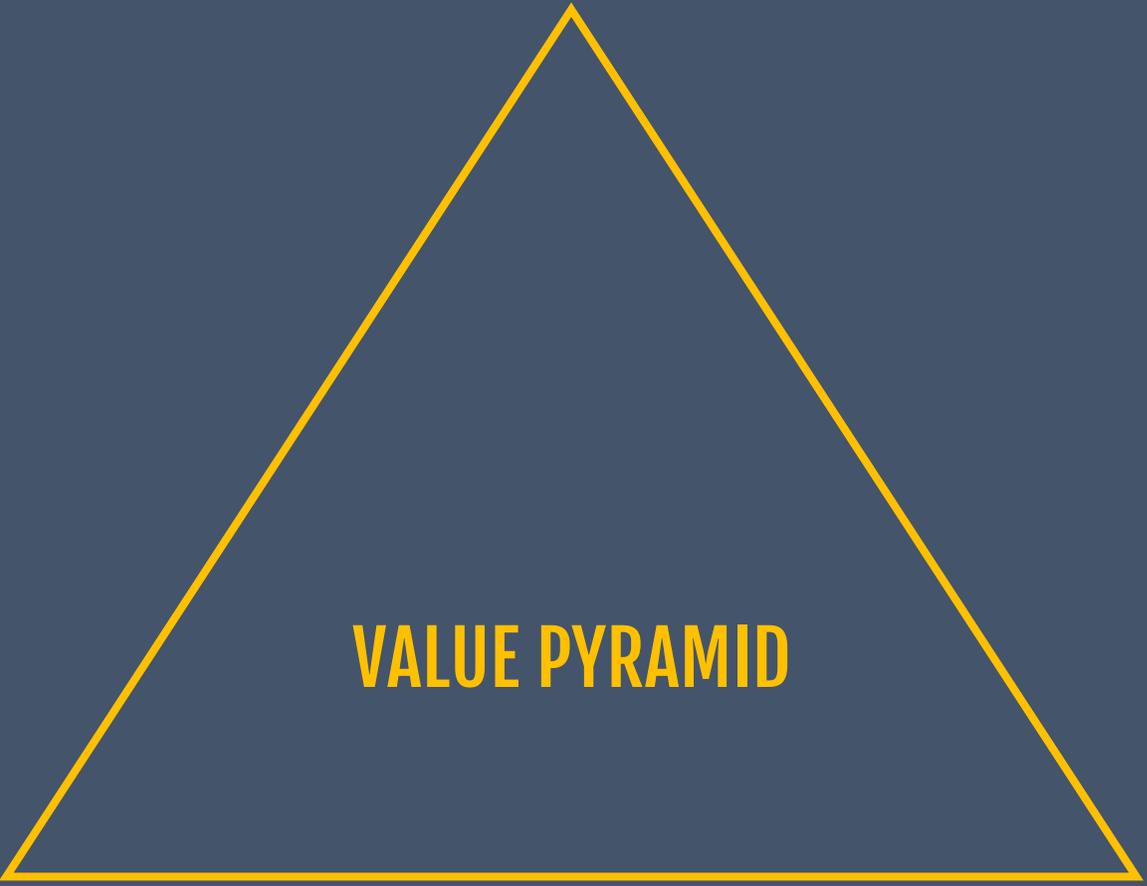
Bringing business value from Data Analytics is not an easy thing. Hundreds of small problems can occur and provoke a misalignment between technology and business.

We need to draw a consistent path then.

First, we should start **identifying the needs and problems** are impacting most in our business co-creating with our stakeholders (Regulators, investors, suppliers/partners, community, consumer, employees...)

Second, we need to recognize the key elements of any use case identified. We need to recognize **value** (What is the question, business problem or target outcome? How is value realized?). We need to recognize **information** (What data or data sources e involved?). And we need to do the same with **analytics** (What analytical or data science methods are applied to the data)

Once we've done that, it's time to move along the **value pyramid**...



VALUE PYRAMID

# Connecting Data and Analytics to Business value – Top down approach

Identify a stakeholder and select an outcome

Examples:

- . Reduction in production-line stop occurrences by 50% and increased production volume by 12%.
- . Evolved workforce mentality from "fit it and restart" to "observe and prevent"
- . Determine financial results

**STEP 1**



# Connecting Data and Analytics to Business value – Top down approach

Identify a stakeholder and select an outcome  
Examples:  
. Reduction in production-line stop occurrences by 50% and increased production volume by 12%.  
. Evolved workforce mentality from "fit it and restart" to "observe and prevent"  
. Determine financial results

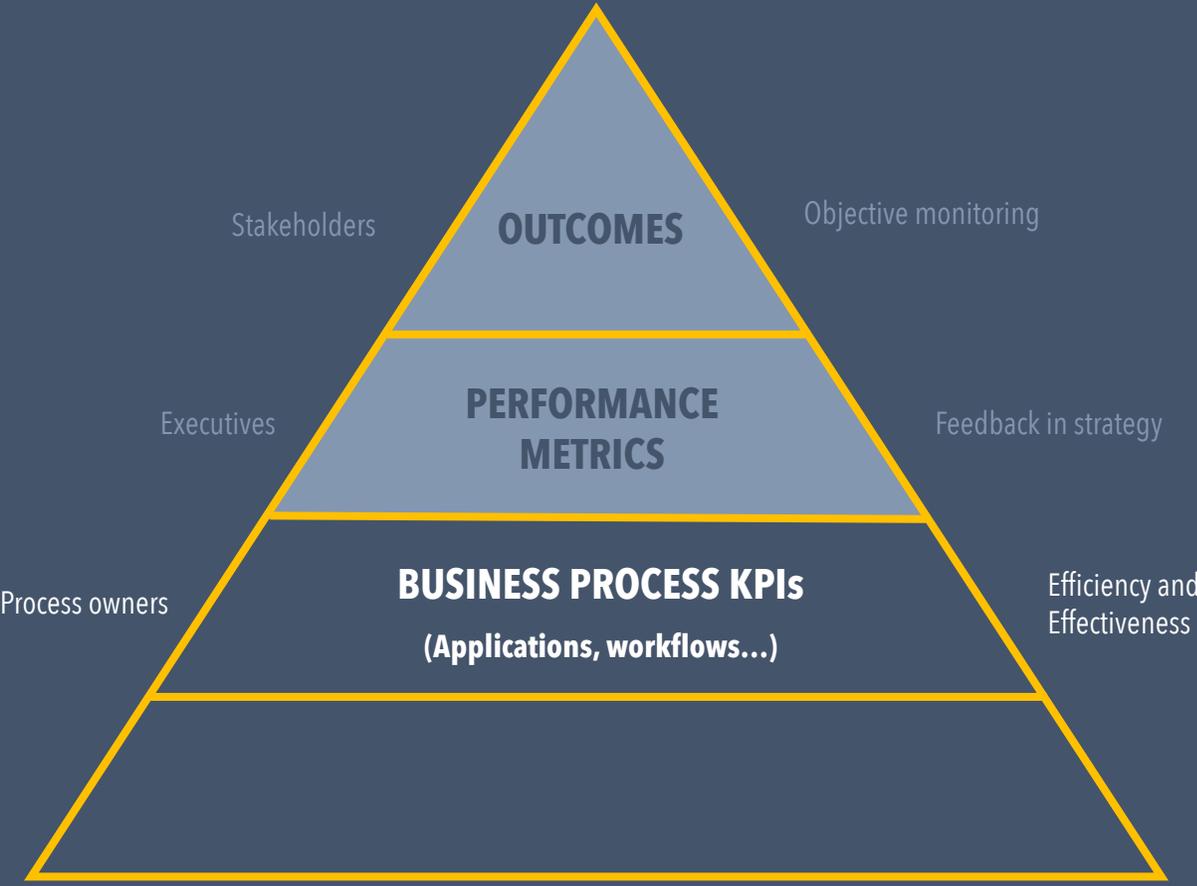
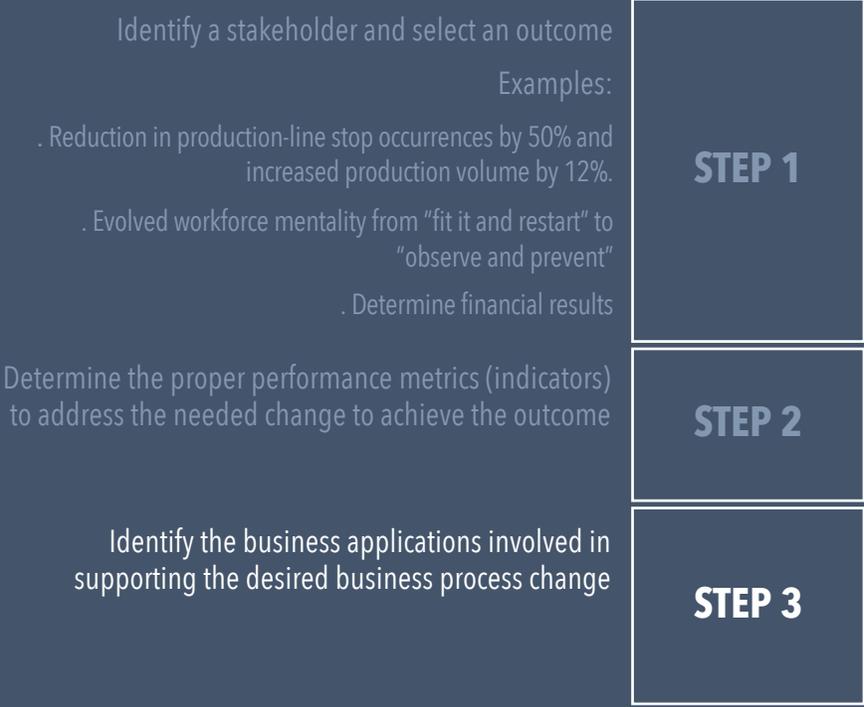
**STEP 1**

Determine the proper performance metrics (indicators) to address the needed change to achieve the outcome

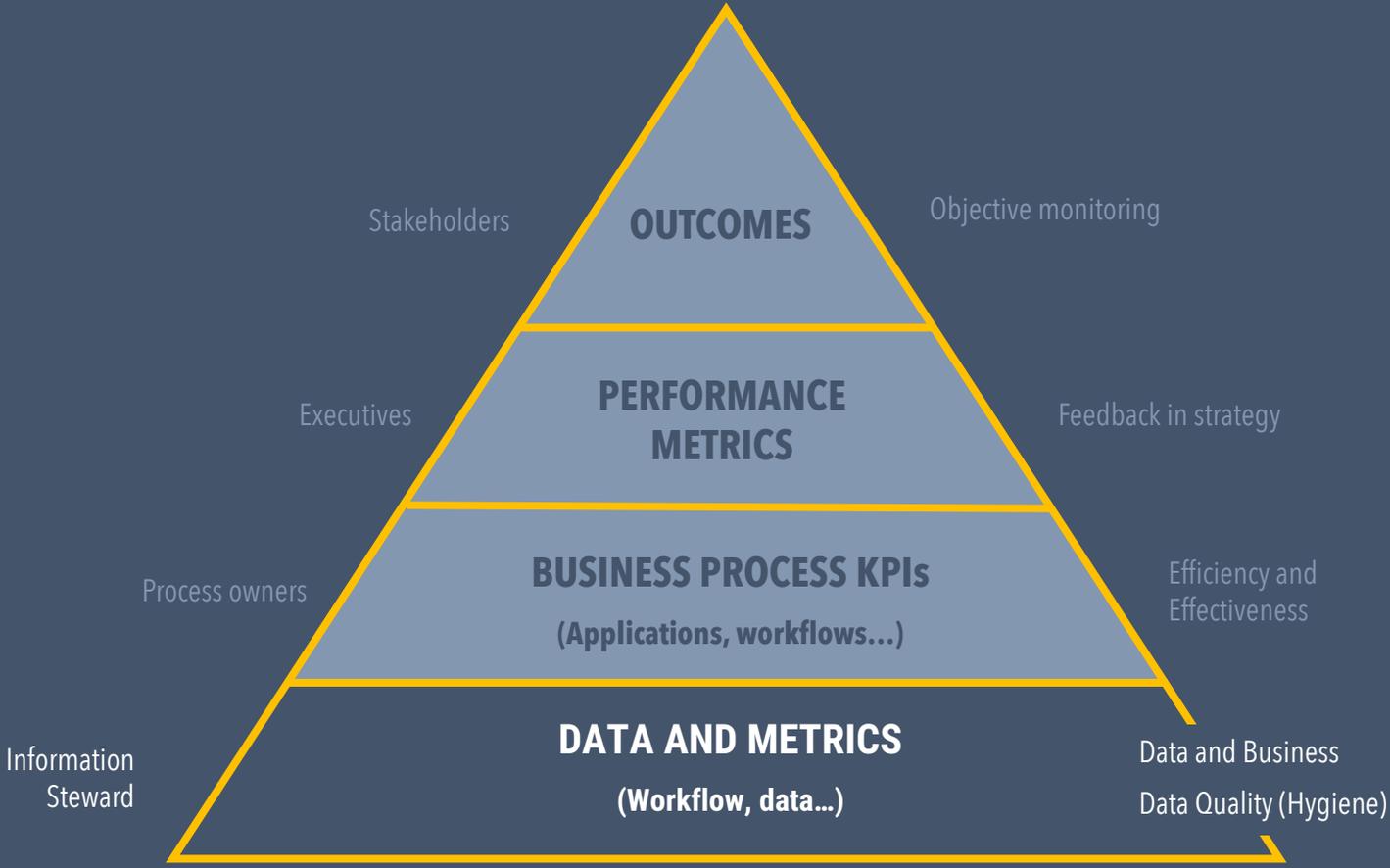
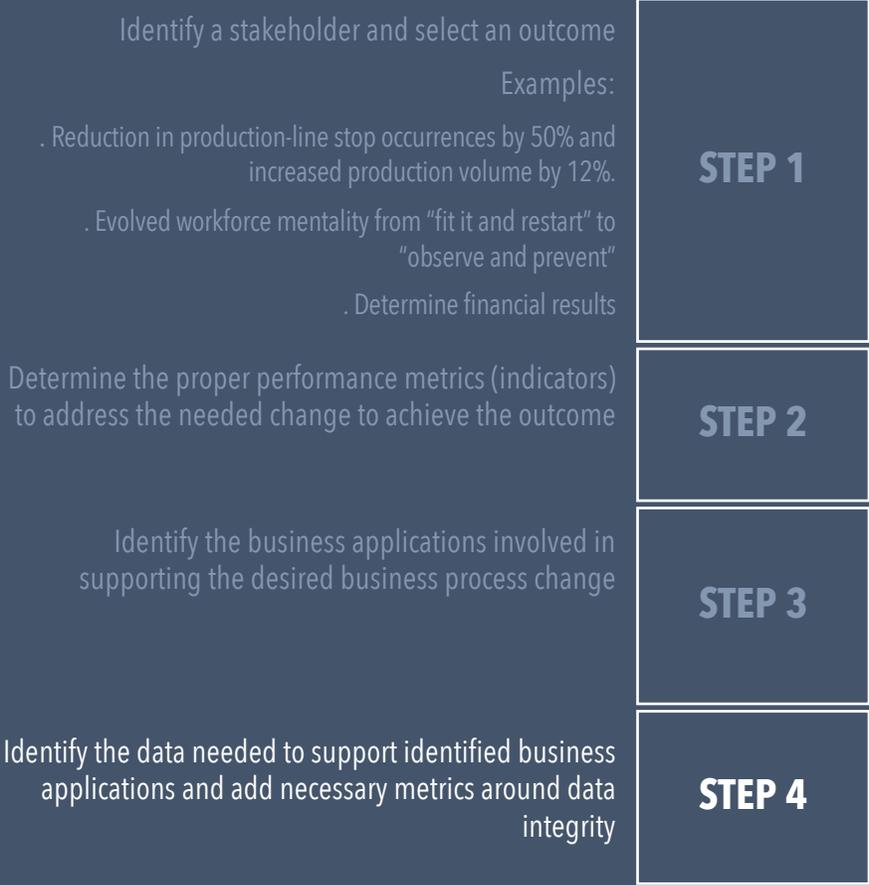
**STEP 2**



# Connecting Data and Analytics to Business value – Top down approach

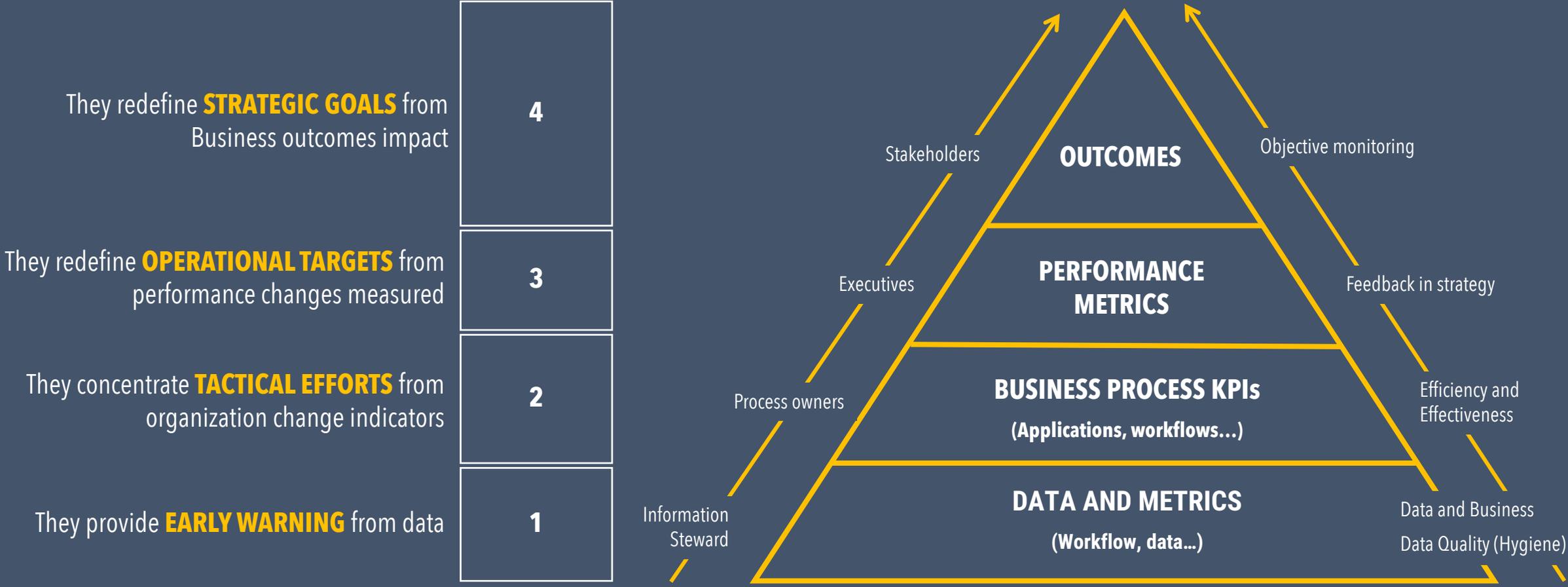


# Connecting Data and Analytics to Business value – Top down approach



Source: Gartner public info and Internet references

# Connecting Data and Analytics to Business value – Bottom Up feedback



Source: Gartner public info and Internet references

**All this stuff seems quite complicated to manage. Is it  
easy to drive and control?**

# Sounds familiar?

Hours and hours spent preparing for the meeting

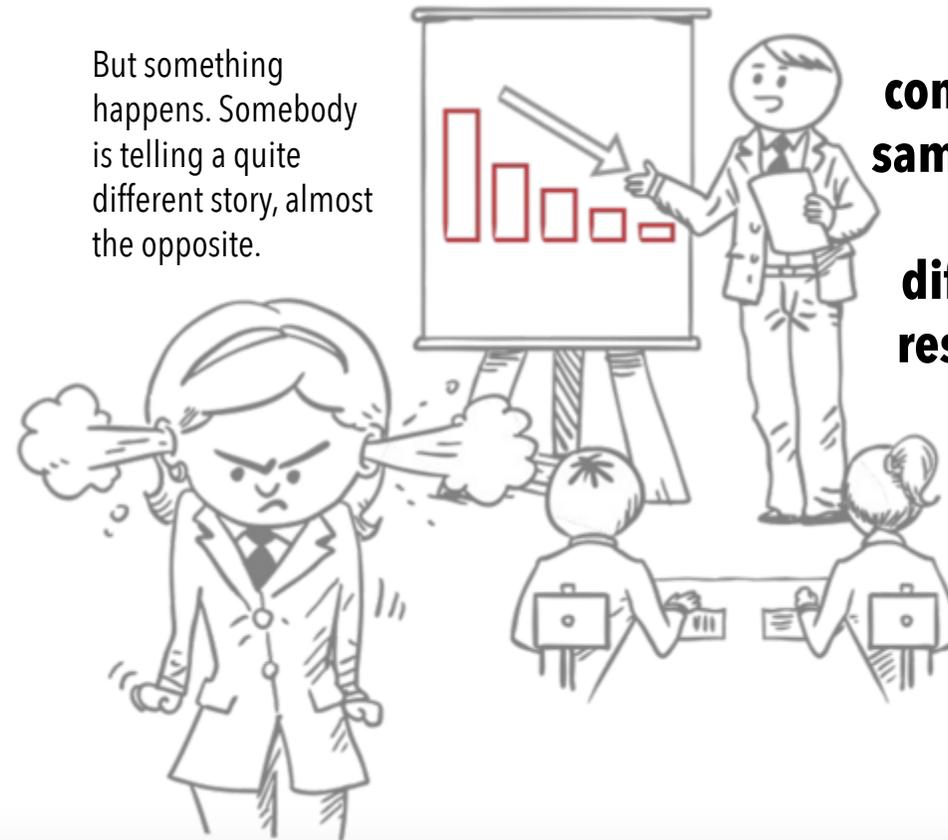
So much time collecting data from finance, IT and the data lake

You did your homework. You made the proper analysis. You prepared relevant insights and wrote some recommendations

**You're ready to go**



But something happens. Somebody is telling a quite different story, almost the opposite.



**Same company, same data but different results...**

# Other consequences of a bad Data Governance

Your meetings dissolve into **data brawls**

Your decisions are primarily made by associated staff who owns the IT systems.

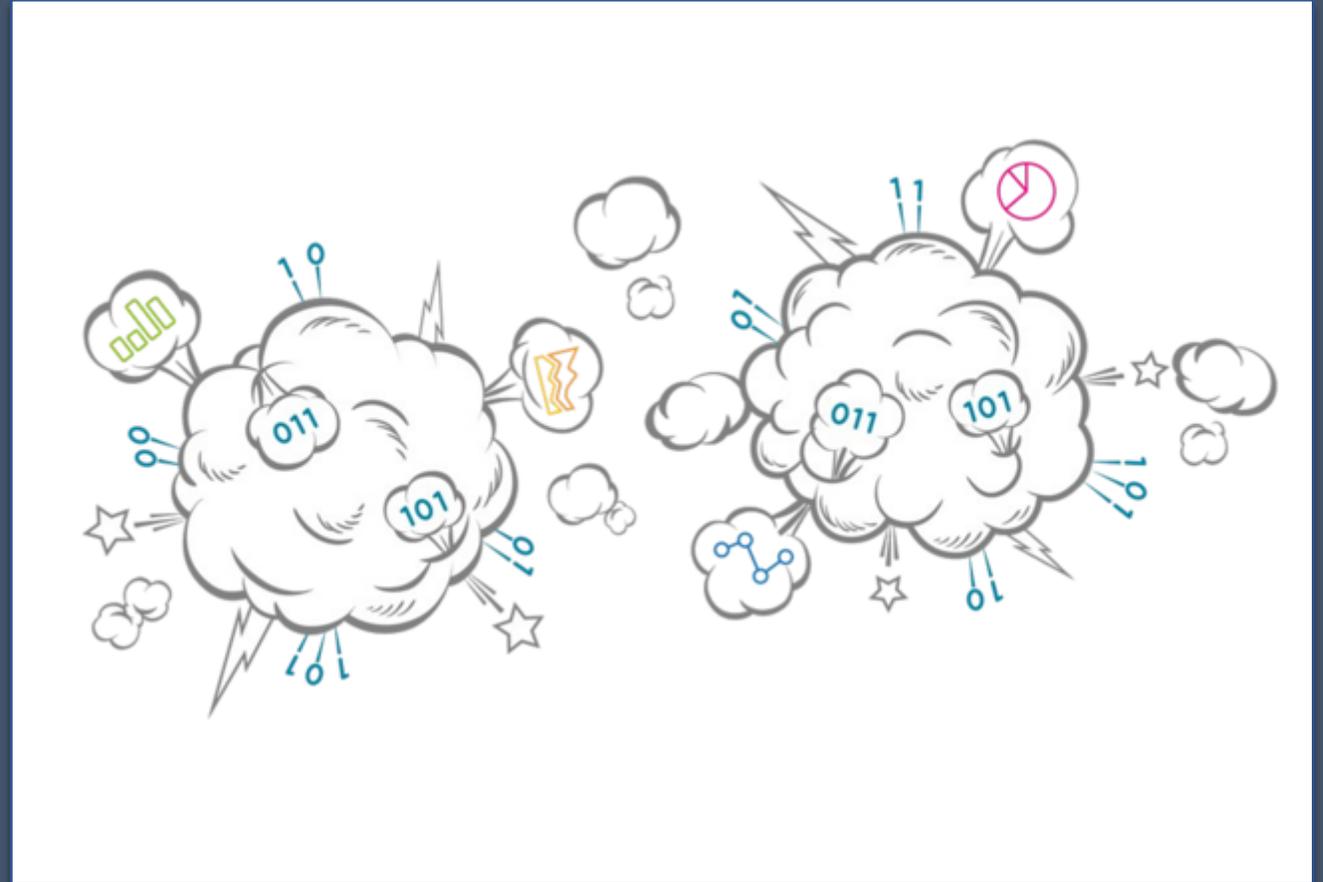
Your data availability and collection are **inconsistent**. It depends on third parties' good deeds.

Your data usability and quality are **not guaranteed**. You never know that's the most accurate or current.

Your data integrity is **inconsistent**. Maybe you're using a local duplicate and old dataset because there's no unique data policy.

And so forth...

So, a trustworthy data, a common understanding, a complete traceability, and a transparent data ownership is needed



# Data governance inside the organization is...



# Data governance for users is...

A system of record for Data assets that should be the authoritative source of information for any given data asset used by/valuable for the organization

**FIND | Know where the data comes from**

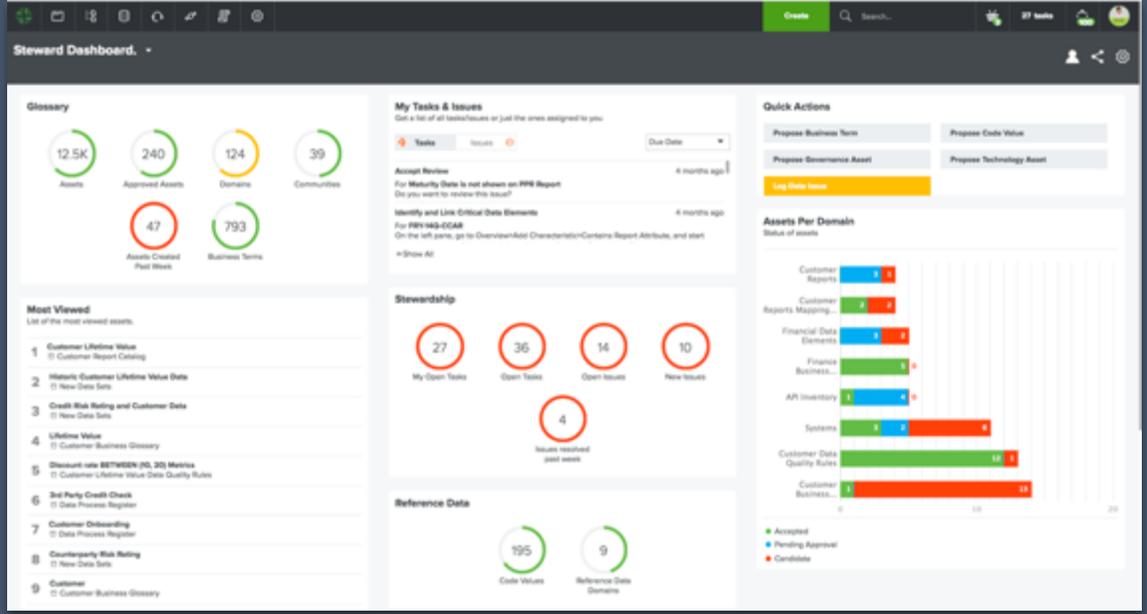
- 1. Catalog
- 2. Data dictionary

**UNDERSTAND | Know what the data means**

- 1. Business glossary
- 2. Reference data

**TRUST | Know that the data is right**

- 1. Policy manager
- 2. Data helpdesk
- 3. Stewardship

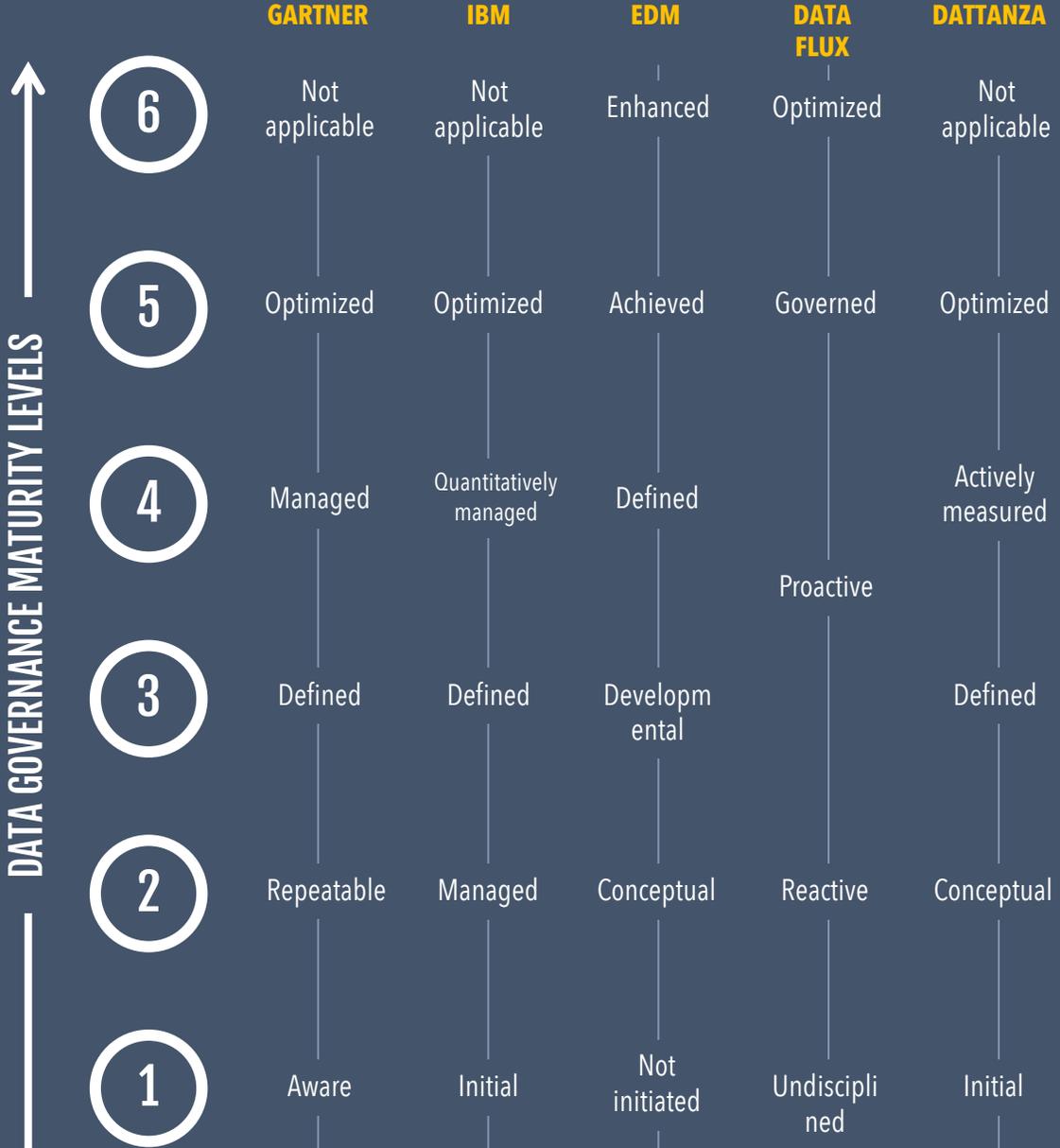


**Collibra customers achieve 510% ROI with their data governance and catalog initiatives**

# The Data governance path

There are several approaches to face a Data Governance maturity model. Gartner, IBM, CMM, EDM... but most of them put the emphasis on same things.

- . Process and the nature of the outcomes with each achieved process state
- . Adoption by People, Process and Technology
- . Implementation of capabilities along with Risk and related Benefits
- . Perceived data value from non-monetization to monetization data
- . Data being traversed from being a transactional asset to an enterprise asset
- . Implementation of Org structure, Policy, Lineage, Metadata, Funding, and Culture Change
- . Effectiveness of Accountability, Formalization, Principles of managing and governing data



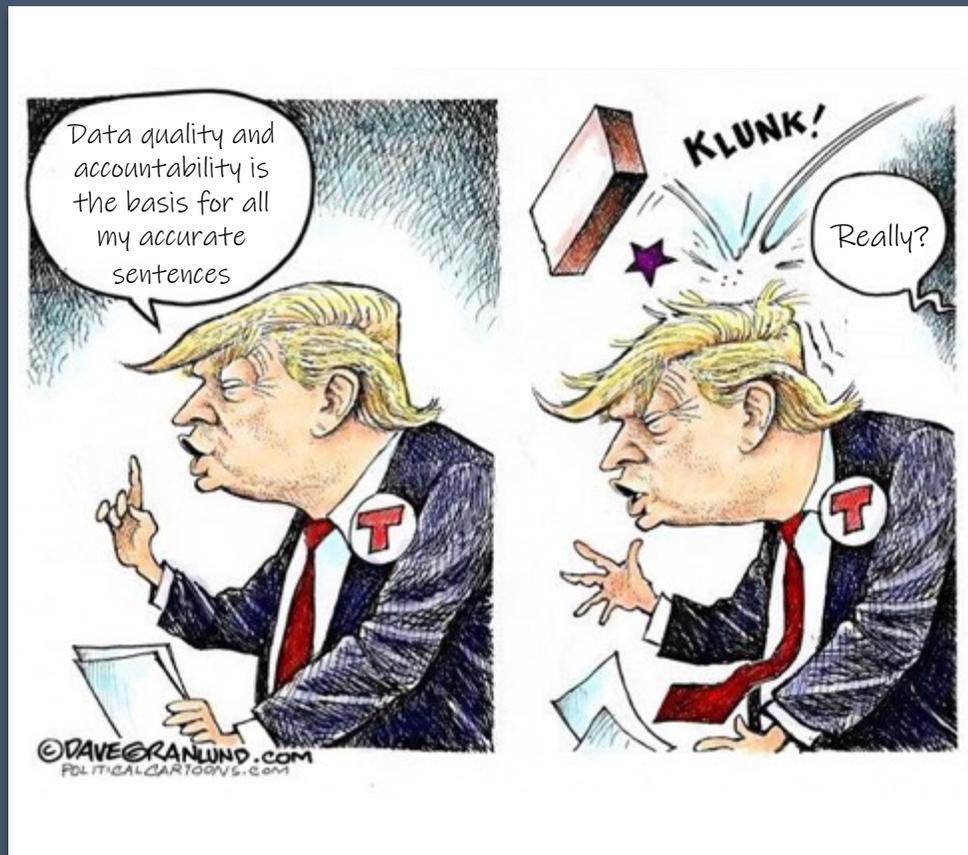
# Data Governance first level – AWARE

## Accountability

Executive level accountability is not defined. Scope of System Owners (stewards / owners) are not defined or named.

## Quality

Profiling or data quality measures of data elements are not in place or data quality standards are not published.



## Metadata

System characteristics are not defined or documented. Business, technical and operational metadata is not effectively cataloged.

## Provisioning

Data is distributed on a reactive basis with little to no enterprise planning, control or governance.

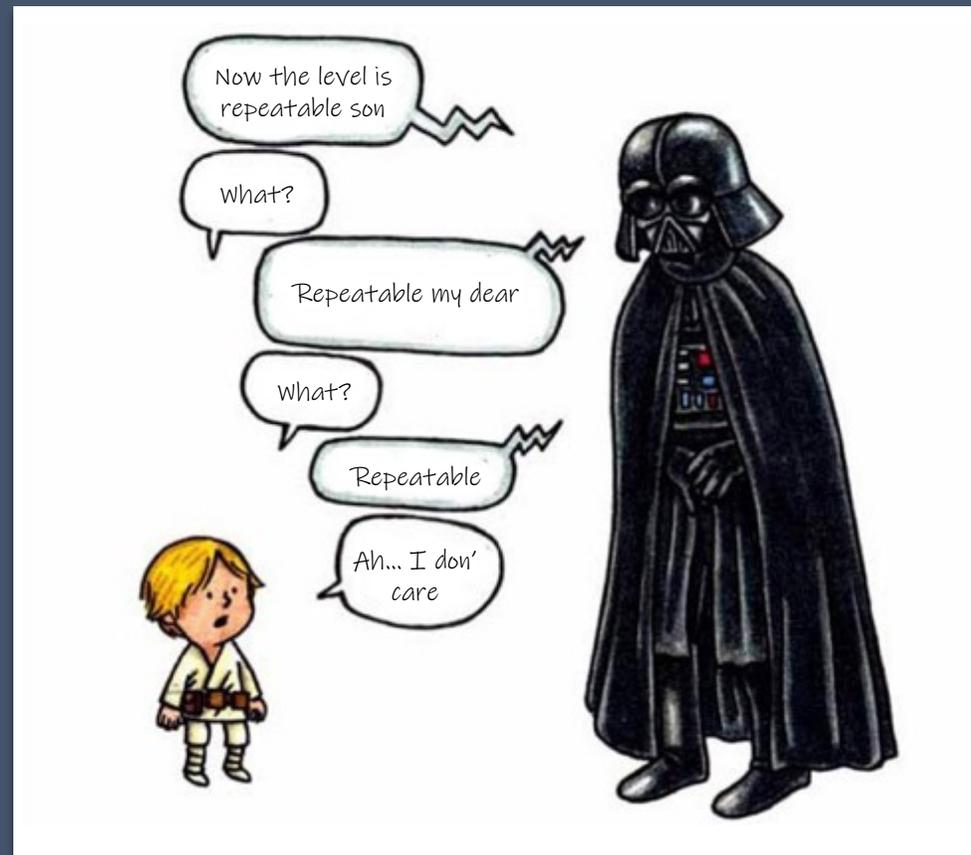
# Data Governance second level – REPETEABLE

## Accountability

Scope of **System Owners is defined**, and stewards / owners are named.

## Quality

**Key business elements** are identified and **data quality standards** are created for them based on analysis of **profiling** or data **quality measurement** results.



## Metadata

Baseline documentation for data stores, data flows, and data element definitions are available.

## Provisioning

**Registration** as a data provider is complete. All of the system's business domains and provisioning roles are documented. The **data provider assessment information** is complete.

# Data Governance third level – DEFINED

## Accountability

Executive level accountability is defined by **Enterprise Governance**.

## Quality

**Key business elements** are measured against **data quality standards** and data quality **reports** are published on an established **measurement schedule**.



## Metadata

Documentation for data lineage, data stores and data element definitions are maintained and available to the enterprise.

## Provisioning

**SLAs** with data consumers are in place; all **data flows** into and out of system are documented. Projects comply with enterprise **Data Provisioning Strategy Framework and Guidelines**.

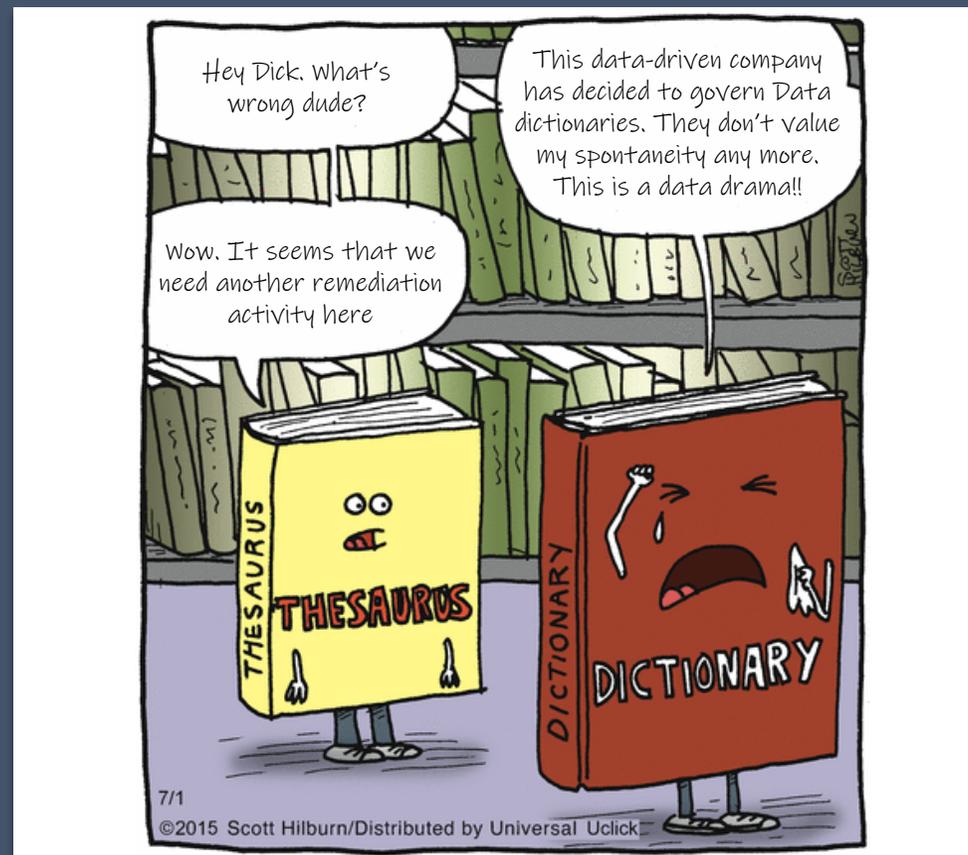
# Data Governance fourth level – MANAGED

## Accountability

**Local Data Governance** bodies are chartered and routines are in place.

## Quality

**Data quality goals** are in place and drive **remediation** activities.



## Metadata

**Data Dictionaries** are governed. A **Business Domain Language** is governed and mapped to data elements.

## Provisioning

**Data lineage** is documented for all prioritized data flows involving the system. **Target provisioning points** are designated and a **domain provisioning roadmap** is in place.

# Data Governance fifth level – OPTIMIZED

## Accountability

Enterprise and Local Data Governance sustainability routines are in place.

## Quality

A data **quality response plan** with defined **thresholds** for key business elements and data **quality self-monitoring activities** are defined and followed.



## Metadata

Metadata is easily accessible across domains. The BDL is mapped to the CBL.

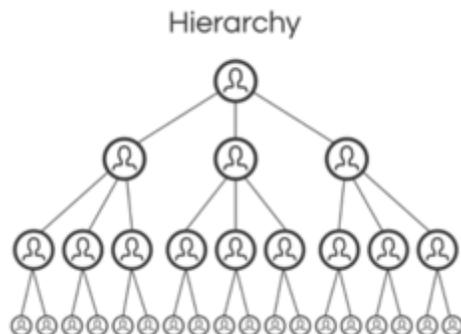
## Provisioning

Data provisioning and consumption occurs according to the **domain roadmap**. **Provisioning governance** is in place and drives ongoing funding and execution decisions.

# A shift in Data Governance is happening – From hierarchies to networks

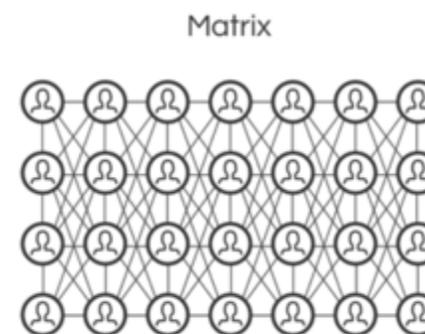
## HIERARCHICAL DATA GOVERNANCE (SYSTEM OF **RECORD**)

- . CDO as a Coordinator: Inward-oriented / Traditional Data / Service
- . Defensive: Risk-driven
- . Scarcity: Few consumers, few producers
- . Compromises on old obsolete cost assumptions of digital power
- . Use of digital optimizes to some extent
- . Not scalable for big data by larger 'data scientist' populations



## NETWORKED DATA GOVERNANCE (SYSTEMS OF **ENGAGEMENT**)

- . CDO as an Experimenter: Outward / Big Data / Strategy
- . Offensive: Value-driven
- . Abundance
- . Many Producers - Data Democratization | Eliminate Breadlines | Consumerization of BI and cheap digital power | Many serve many | Supports customer
- . Many Consumers - Data Amazonification | Access, SLA, Trust, Secure Cloud...



## Who's the CDO?

It's the abbreviation of **Chief Data Officer**

He/she takes care of Analytics excellence, Information management excellence and leads change for impact

He/she defines strategy, applies best methods, embraces cutting-edge technologies and moves organization forward in data-driven maturity levels

Her responsibilities are:

- . Enterprise Data Strategy
- . Data Governance and Management
- . Data Analytics
- . Data Architecture
- . Data Culture

"The executive suite needs someone who can oversee the strategic business application of its information assets enterprise-wide"

**Michelline Casey**  
- **Federal Reserve Board CDO**



*That's all Folks!*

Coming soon

**DATA ANALYTICS AND BIG DATA – INSIDE THE ORGANIZATION**