# Business intelligence

## Unit 1 – Introduction to BI and methodology

# BI First class: An introduction

- **What is BI?**
- **Why is it important?**
- **How does it relate to Big Data?**
- **Where is it used?**
- **BI Tools**



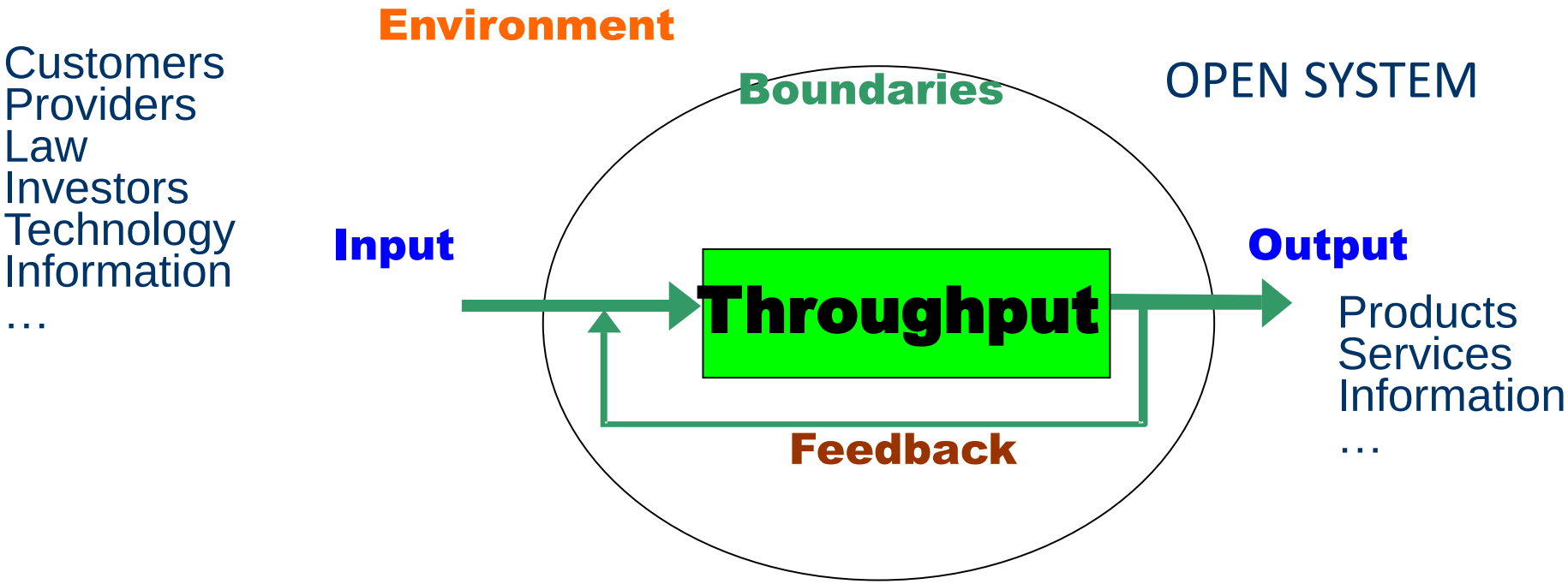
Source: Skillsoft

# Information systems



- **System:**
  - System is a set of interrelated things that contribute to a certain goal or perform a specific function.
  - A system consists of a set of interrelated elements, operating in a changing environment and with specific objectives.
- **Basic elements:**
  - System components.
  - The relations between them, which determine the structure of the system.
  - The objective of the system.

# Information systems

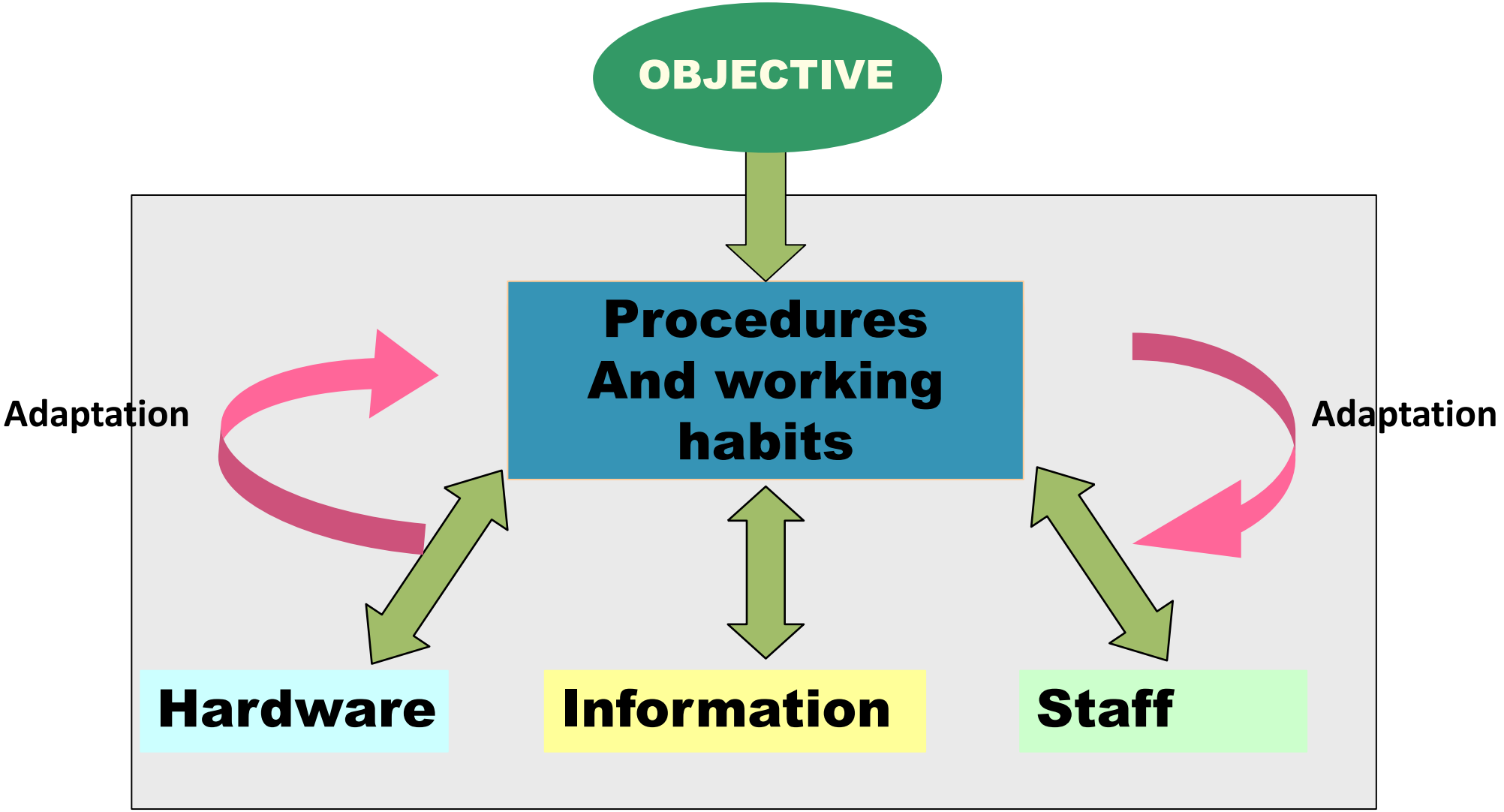



- The system **environment**: his surroundings, within which it is located.
- The system **boundaries**: the boundary between what the system is and what constitutes the environment.
- **Feedback**: In many systems the output influences the input of the system.
- **Throughput**: Transformation processes.

- Definition:
  - A formal set of processes that, operating on a collection of data structured according to the needs of the company, collect, process and distribute the information (or any part of it) necessary for the daily operations of the company and for the direction and control (decisions) to carry out their activities in accordance with its business strategy.

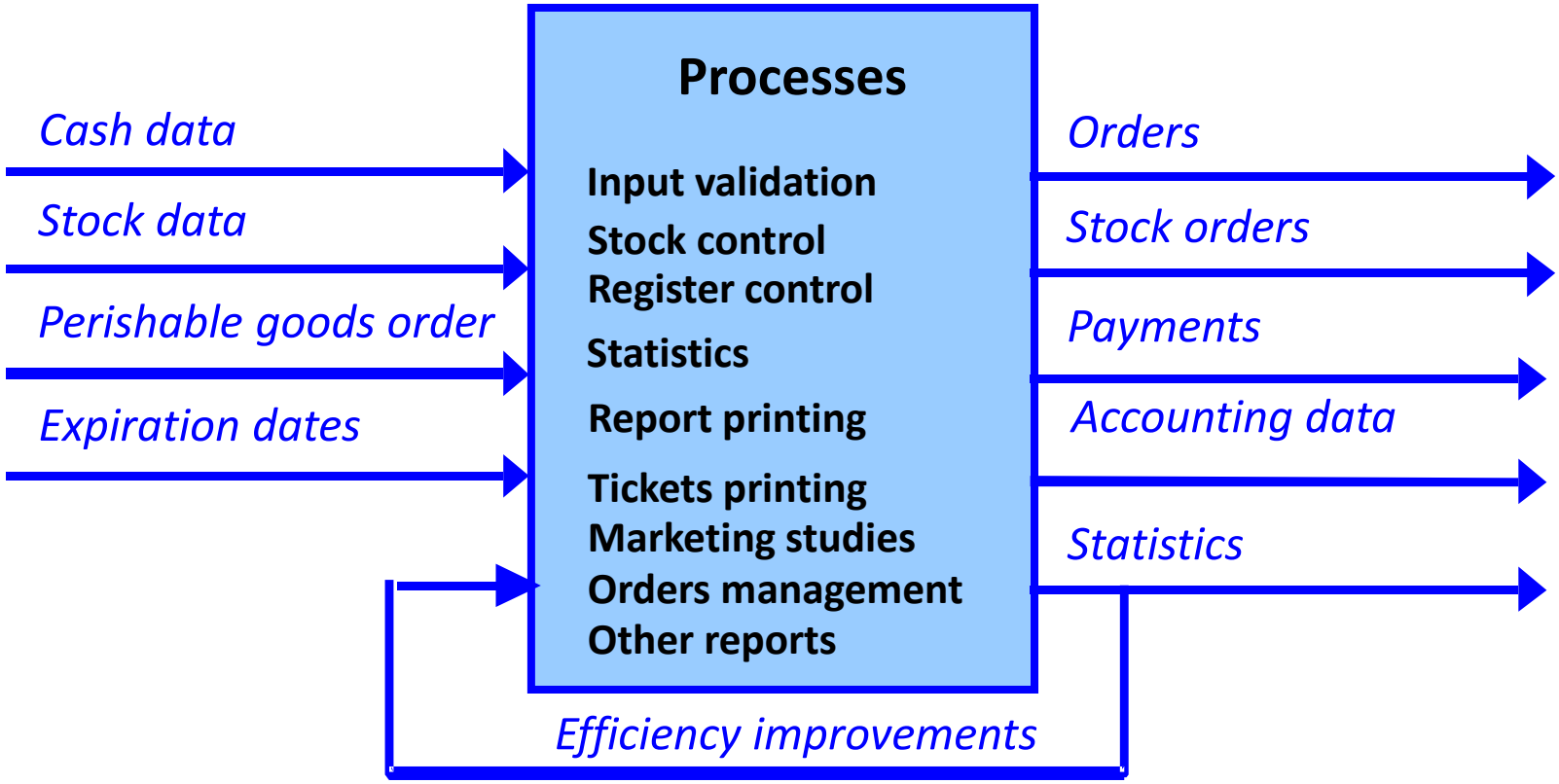
- Other definitions do emphasize that the goal is to provide quality information:
  - The goal of SI is to help the performance of activities at all levels of the organization, by providing the right information, with sufficient quality, to the right person at the right time and place, and with the most useful format to the receiver.

# IS: basic elements and relations (Piattini et al. 96)

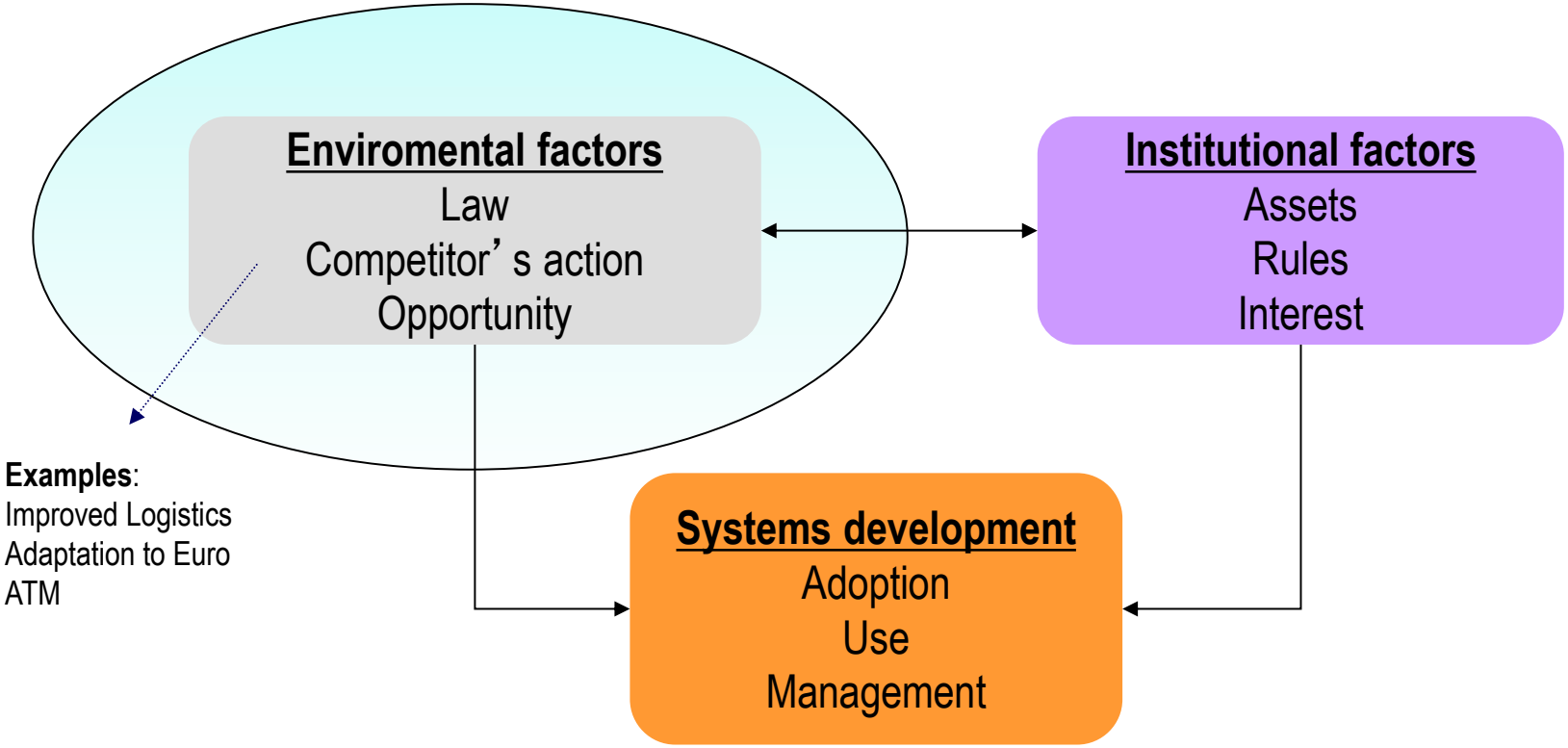


# Information systems



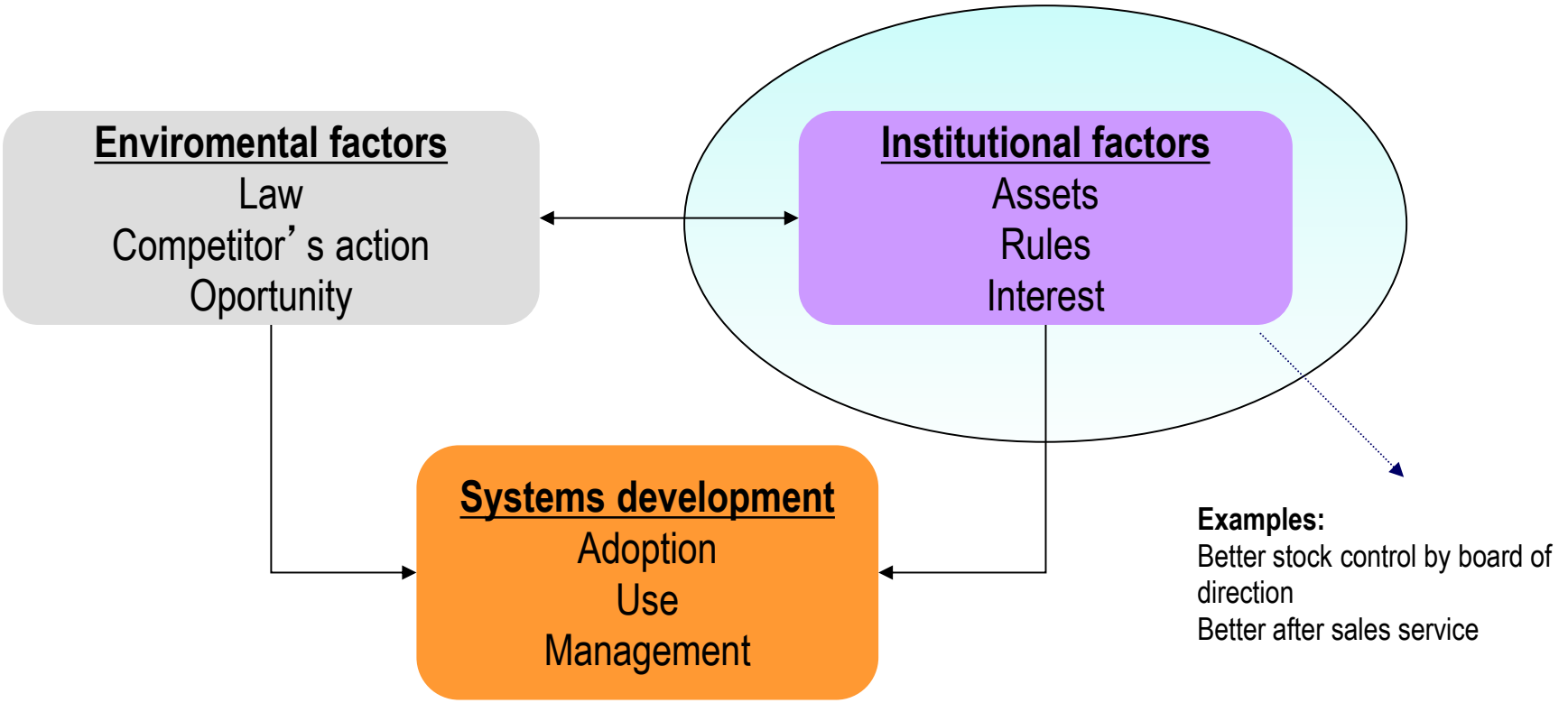


# Information systems



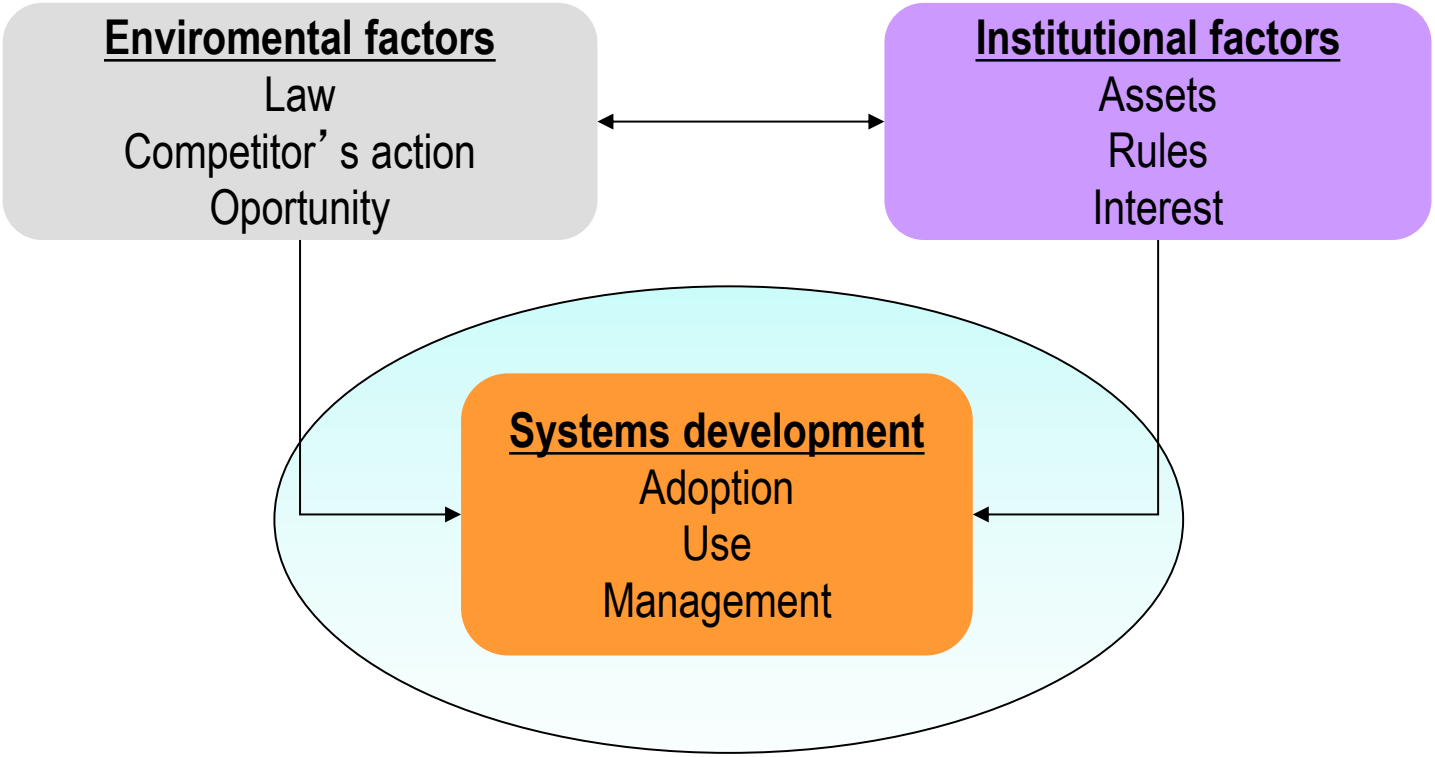

- From a business perspective, an information system is a solution for business organization and management based on information technology whose aim is to deal with an emerging challenge in the business context

# Information systems




- Strategic information systems: at any organizational level there are changes in goals, operations, products, procedures or relationships with the environment to gain a competitive advantage

# Information systems




- The value of an IS depends on its effectiveness, its scope, its acceptance by those who use it, its cost, the quality of information that is produced, etc..

# Information systems
# An example

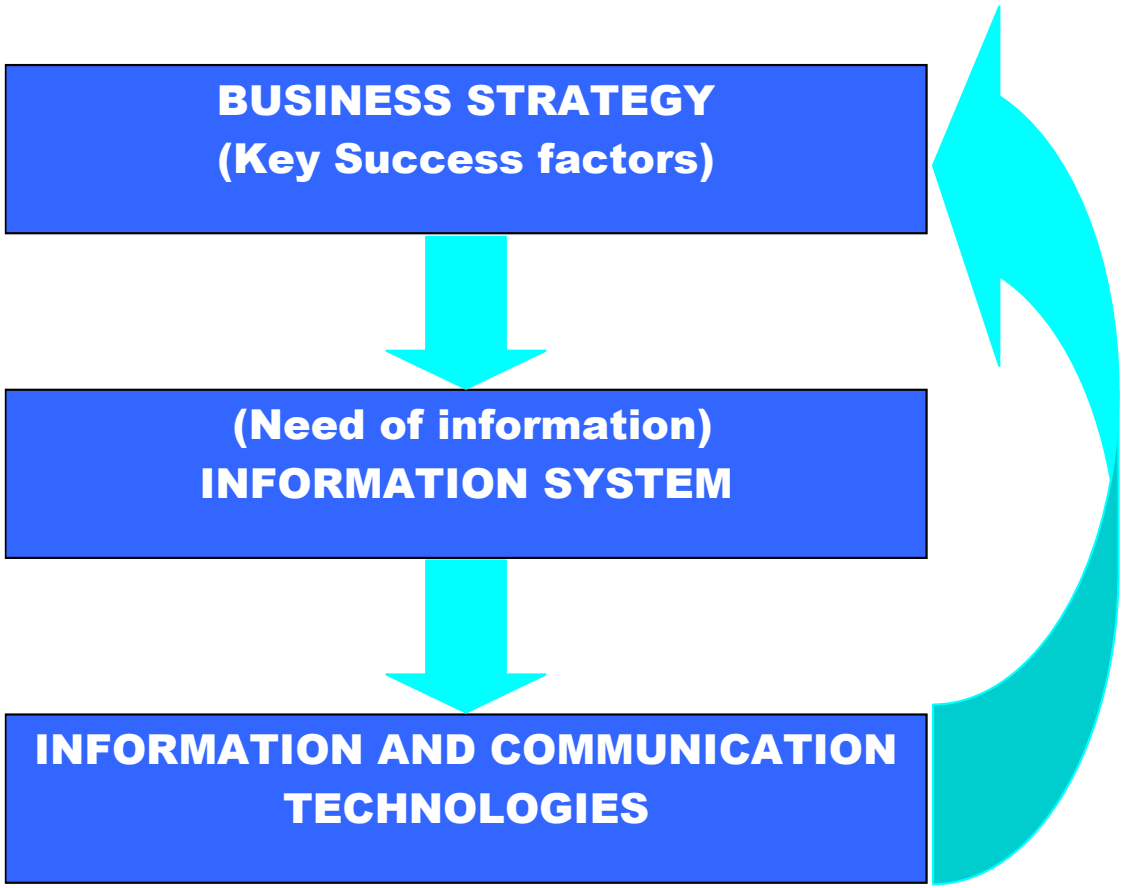
- Mussel farming IS
  - Some Environmental factors:

| | |
|---|---|
| EU, national and regional regulations (food safety and hygiene) | Certifications |
| Climate change | Economic landscape |
| Technology improvements | Public administration support |

  - Some Institutional Factors:

| | |
|---|---|
| Technology and Infrastructure | Human resources |
| Financial resources | Internal procedures |
| Interest in maximizing profits | Interest in expanding market |

# Enterprise information systems



# Enterprise IS

- “"The Enterprise IS coordinates the information flow and records necessary to carry out the functions of a company according to its business approach or strategy”"
    - Business strategy is the key.

- Enterprise-resource planning ERP
- Workflow management systems
- Groupware systems
- E-commerce systems
- Electronic Data Interchange, EDI

# IS Evolution

| Decade | Required Information | Type of Information Systems | Objectives |
|---|---|---|---|
| 1950s | Basic Computational Data | Batch Processing Systems | Automate complex calculations and data processing tasks |
| 1960s | Transactional Data | TPS (Transaction Processing Systems) | Automate routine business operations and record-keeping |
| 1970s | Operational and Tactical Data | MIS (Management Information Systems) | Improve administrative control and reporting |
| 1980s | Analytical Data | DSS (Decision Support Systems) | Support decision-making with data modeling and analysis tools |
| 1990s | Strategic and Competitive Data | EIS (Executive Information Systems) | Enable strategic decision-making and real-time monitoring |
| 2000s | Enterprise-Wide Data | ERP (Enterprise Resource Planning Systems) | Integrate business functions for efficiency and process automation |
| 2010s | Big Data, Predictive Analytics | BI (Business Intelligence Systems) | Data-driven decision-making and predictive analytics |
| 2020s | Real-Time and AI-Enhanced Data | AI Systems and Cloud-Based Analytics | Automate decisions with AI, scalable cloud infrastructure |

# IS Types



Systems **operational level**: They monitor the activities, operations and basic transactions of the organization.

MIS and decision support : They support the monitoring, control and decision-making and administrative activities at management level.

Strategic level systems: They support long-term planning at management level in order to gain competitive advantage.

A fourth type: Knowledge level systems: support knowledge and information workers of the institution.

# IS Types



Example USC:

**Operational-** Student enrollments;

**MIS**-monthly reports on student performance

**DSS**-Analyzes trends in student enrollment to support decision making

**Strategic**- USC takes long-term decisions on new degrees and research lines

**Knowledge**- digital repository with theses, papers, ...

# Operations IS

| Its main features are: |
| --- |
| Performs and records routine daily operations necessary for the operation of the company. |
| Designed to increase productivity |
| Investment in them is easy to justify to the directive board, as its benefits are visible and palpable. |
| They are often the first type of IS that is implanted in organizations. It starts supporting efforts at the operational level of the organization. |
| They are intensive in data input and output, their calculations and processes are usually simple and unsophisticated. |
| They provide administrators with reports and on-line access to historical and daily records |
| They are the main generators of information for other types of systems. |
| Examples: billing, plant scheduling, payroll, inventory, … |

# Decision support systems



| • **Its main features are:** |
| --- |
| They are always included after having implemented more relevant operational IS, since the latter are its information providers. |
| Calculations are usually intensive, while outputs are scarce. |
| The information generated provides support to middle and high management in the decision making process. |
| They combine changing information with sophisticated analytical models to support unstructured and semistructured decision-making. |
| They tend to be interactive, visual and friendly, and they are focused to the end user. |
| They do not intend to save work. As a result, the economic justification for investment in these systems is difficult because the direct benefits of the project are not known. |
| Involve analytical models, what-if analysis, simulations and forecasts. |

# Strategic systems



- **Its main features are:**

They incorporate external information and get summarized information from operational and decision support systems.

They support the introduction of products and processes within the organization because they aim to gain advantage over competitors by innovating.

They function is to achieve advantages that competitors do not have, such as cost advantages and differentiated services to customers and suppliers. In this context, the Strategic System are creators of entry barriers to the business.

- For example, ATM, Internet banking, recommendation systems

They used to be developed ad hoc within the organization.

Aimed to achieve long-term strategic goals (e.g. expanding market share, créate new markets).

# Strategic systems



- **Examples:**
  - Sales forecasts
    - Planning marketing campaing
  - MRP (Manufacturing Planning Resource) focused on increasing productivity in a manufacturing process
  - Product discovery and launching in banking: types of mortgages, types of accounts, ... with the purpose of achieving business goals:
    - Attracting new customers
    - Customer Loyalty
  - Television schedule grid
    - Profile of viewers and appropriate advertising
    - Duration of advertising

# Features

| System | Input | Processes | Output | User |
|---|---|---|---|---|
| Operational | Transactions, | Store, report, merge, | reports, summaries, transactions. Invoice, payroll | Operative staff, supervisor |
| DSS, tactical | Summarized operational information; high volumen of data; Simple models | Model; simulations; analysis Optimization What-if analysis | Analytical reports, Decision Models | Technical staff, data analysts |
| Strategic systems | Agregated information; External information; | Strategic analysis, scenario planning CI | Strategic plans, Forecasts | Directive |

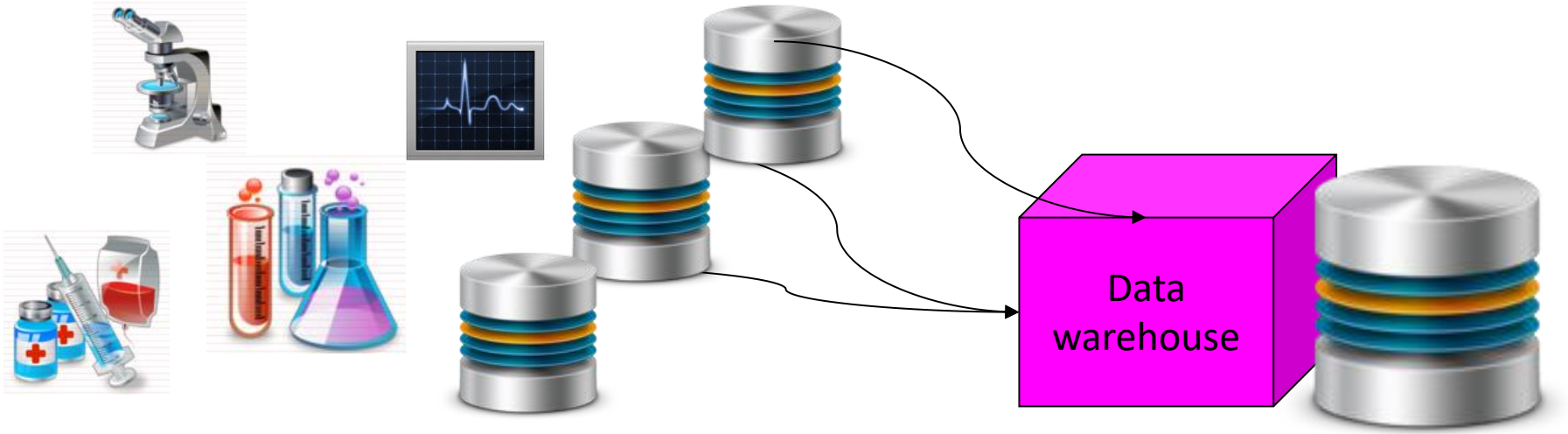
# Business intelligence
# The next step in data-driven decision making

- This term refers to the management of information in a **specific** business or business area.
- It consists of a set of strategies and tools focused on knowledge management by **analyzing existing data** in an organization or company.
- In the business intelligence we focus on:
  - Setting business goals
  - Determine needs of data, information and knowledge to meet business objectives
  - Keywords: Integration, accessibility and data-driven decisions.

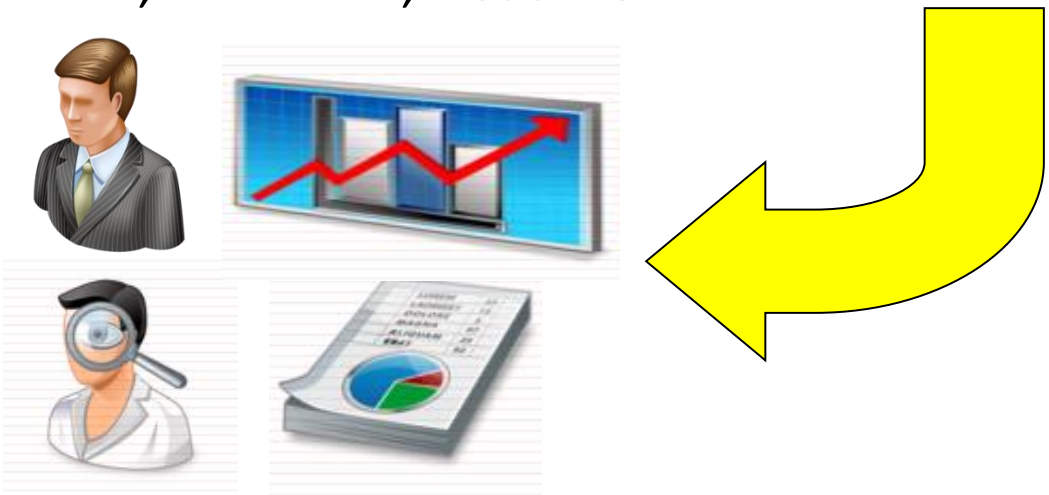
# Business intelligence



- Operative IS: gather and organize data




- **Tactical IS: Analyze**, summarize, transform, **visualize**



# Business intelligence



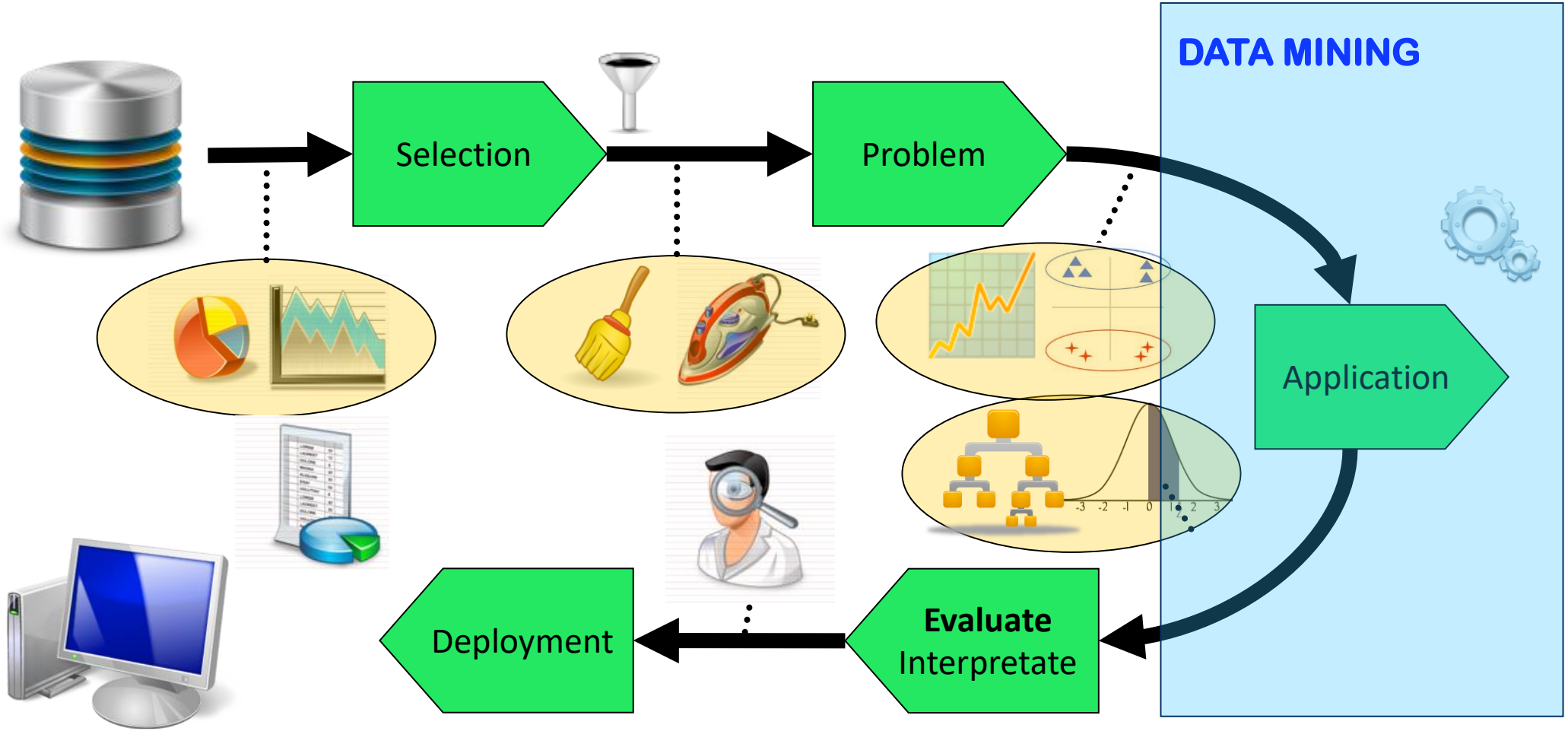
- Strategic IS: **Acquire, discover**, **evaluate** and **use** knowledge

**Knowledge acquistion**

**Knowledge discovery**

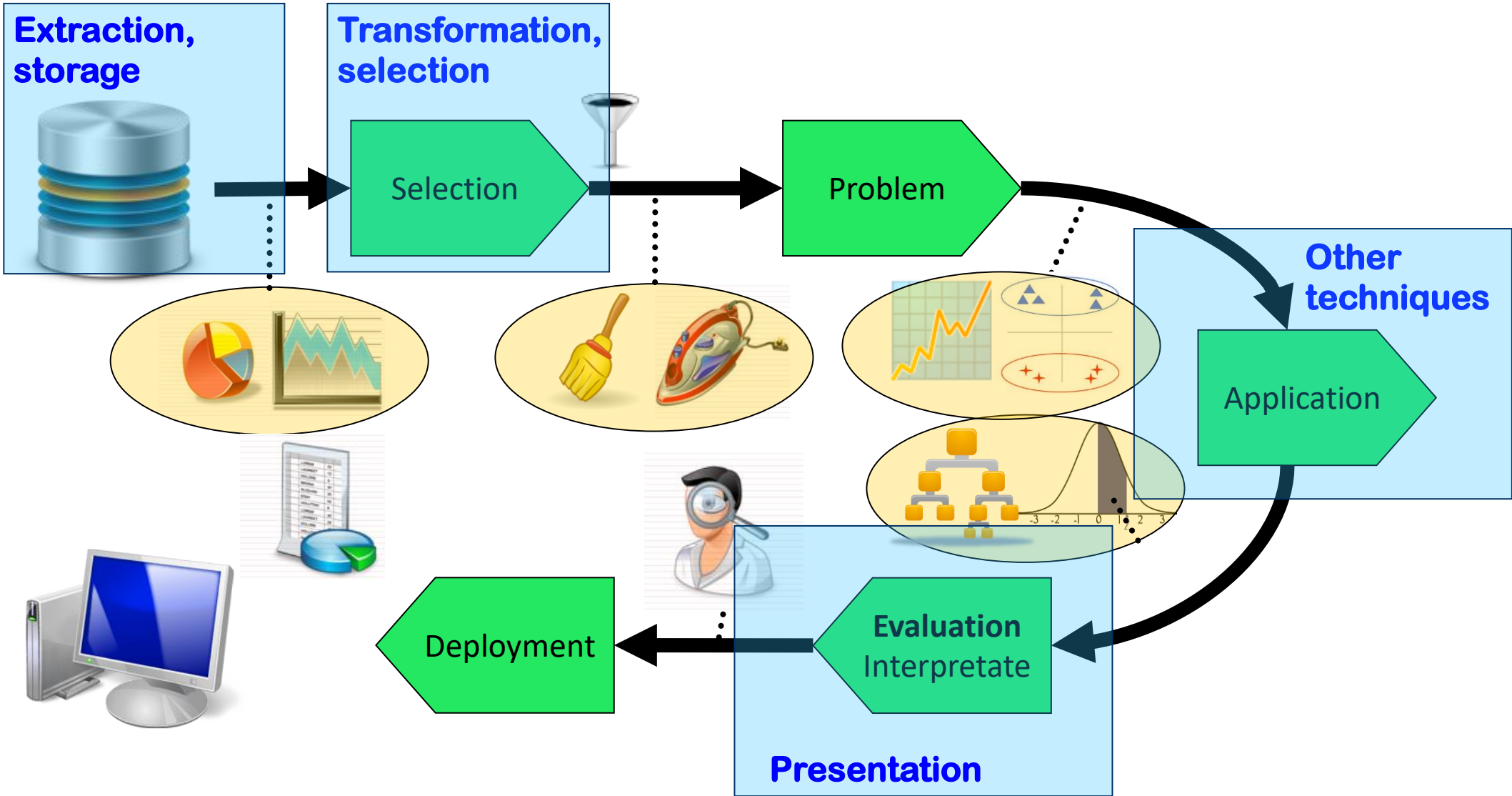
Expert system, dss, knowledge management, data mining, …

- In the other courses you will focus on:



# Business intelligence process

- Now we will focus on the complete process:



# Maturity models

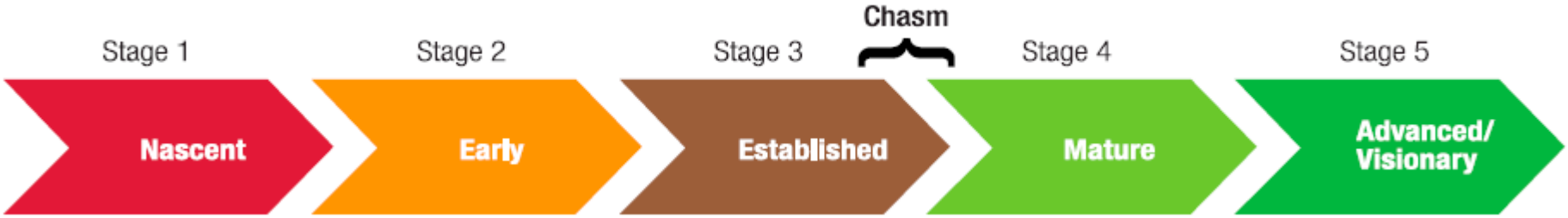
- Maturity models define levels of definition, efficiency, manageability and measurement of the monitored environment.
- The maturity model for Business Intelligence helps organizations understand where they are and how they can improve.
- Several MM:
  - TDWI MM
  - HP
  - Gartner
  - LOBI, Forrester's, …

•Each model has at least 5 stages of maturity
•Each model starts with operational / one-off reporting and culminates in pervasive BI
•The models do not focus on technology alone and hinge on the involvement of people and process as well.
•Provide current capabilities, roadmap for improvement, benchmarking, alignment with business goals, improves data-driven culture, optimize resource allocation and support strategic planning.

# The Data Warehousing Institute (TDWI) Analytics MM

## Model Dimensions

| Organizational Maturity | Resource Maturity | Data Infrastructure Maturity | Analytics Maturity | Governance Maturity |
|---|---|---|---|---|
| • Leadership<br>• Culture<br>• Impact<br>• Strategy | • Funding<br>• Talent/skills<br>• Roles/responsibilities<br>• Training | • Diversity, volume, and speed<br>• Data access<br>• Data integration and management<br>• Data architecture | • Scope of capabilities<br>• Automation/ augmented<br>• Deployment and delivery approaches<br>• Innovation | • Data governance processes and tooling<br>• Model governance processes and tooling<br>• Governance roles<br>• Security/privacy |



# The Data Warehousing Institute (TDWI) MM

Stage 1: Nascent– Pre-analytical environment. Culture is not data-driven and no data-driven decisions. No formal BI infrastructure. Data scattered and in silos. Poor data quality. IT and business don't work together. Excel or manual reports.

Stage 2: Early- Basic BI tools. Starting to undertand the value of analytics. The organization realizes some data infrastructure is needed to support data integration for analysis. Rudimentary analytics but advancing.

Stage 3: Established- A data warehouse has been implemented. There is a group that is reponsible for analytics in IT. A data governance team has been formed. Data visualization tools are used.

Stage 4: Mature- End users integrate analytics into decision-making and operations. Culture of innovation and collaboration between IT and business. Use of diverse sources, including semi and unstructured data. Worforce include data scientitsts, data engineers, MLOps and roles like Chief Analysics Officer.

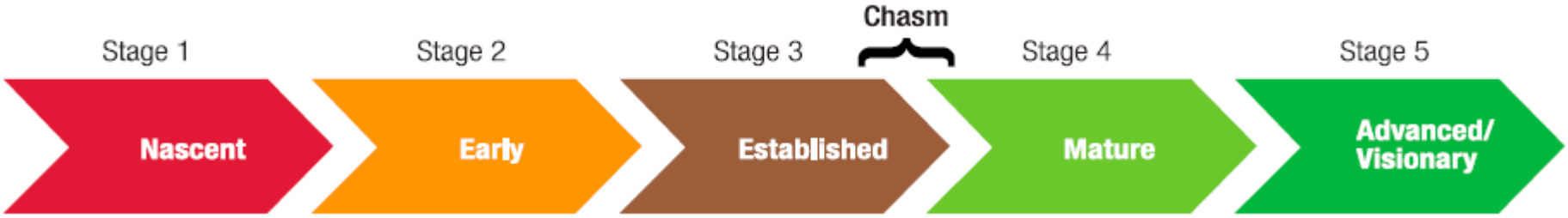
Stage 5- Advanced/Visionary- Self service analytics. Organization considers analytics a critical, competitive weapon. Operational, flexible and scalable infrastructure (data lakes, DWH, cloud), ML, AI is used in NRT for decision-making.

# The Data Warehousing Institute (TDWI) Analytics MM

## Model Dimensions

| Organizational Maturity | Resource Maturity | Data Infrastructure Maturity | Analytics Maturity | Governance Maturity |
|---|---|---|---|---|
| • Leadership<br>• Culture<br>• Impact<br>• Strategy | • Funding<br>• Talent/skills<br>• Roles/responsibilities<br>• Training | • Diversity, volume, and speed<br>• Data access<br>• Data integration and management<br>• Data architecture | • Scope of capabilities<br>• Automation/ augmented<br>• Deployment and delivery approaches<br>• Innovation | • Data governance processes and tooling<br>• Model governance processes and tooling<br>• Governance roles<br>• Security/privacy |

Chasm- Organizations need to cross the Chasm, ensuring the right governance, data architecture,  skills, data culture, is in place.



# The Data Warehousing Institute (TDWI) Analytics MM

## Leadership

* Your leadership supports and evangelizes analytics across the company.
  - ◯ Not at all
  - ◯ They seem ambivalent about analytics, and they don't really evangelize it
  - ◯ They support analytics efforts and are starting to evangelize it
  - ◯ They firmly support analytics efforts, they use analytics to make decisions, and they evangelize it across the company

* Your company has a Chief Analytics Officer (CAO) who is in charge of your analytics efforts.
  - ◯ We don't have anyone in charge of analytics in the organization
  - ◯ Analytics is controlled by IT in my company
  - ◯ We have a VP or Director of Analtyics in my company, who is in charge of analtyics
  - ◯ Yes, have a Chief Analytics Officer

## Strategy

* Your company has a strong strategy in place to support its data and analytics efforts.
  - ◯ No and we have no plans to do so
  - ◯ No, but we plan to do so in the next year
  - ◯ Yes, we are in the process of putting a strategy together
  - ◯ Yes, we have a solid strategy in place for analytics

* Analytics is an important part of your company's digital transformation strategy
  - ◯ No, we do not have an digital transformation strategy
  - ◯ Yes, we are in the process of putting our digital transformation strategy in place and analytics will play an important role
  - ◯ Yes, analytics is an important part of my company's digital transformation strategy

# The Data Warehousing Institute (TDWI) Analytics MM

## Impact

* What % of business units in your company use analytics for day to day decision making
  - ◯ less than 25%
  - ◯ 26-40%
  - ◯ 41-55%
  - ◯ 56-70%
  - ◯ Greater than 70%

* Your organization has measured an impact with its analytics
  - ◯ No
  - ◯ No, but I think we've gained value
  - ◯ Yes, we've measured a top or bottom line impact

## Culture

* Your organization uses analytics to take action.
  - ◯ Strongly disagree
  - ◯ Disagree
  - ◯ Neutral
  - ◯ Agree
  - ◯ Strongly agree

* There is a culture of trust in analytics across your company.
  - ◯ Strongly disagree
  - ◯ Disagree
  - ◯ Neutral
  - ◯ Agree
  - ◯ Strongly agree

## Interpretation

Once you complete the survey, you will receive your results. The breakdown of scores for each dimension is as follows:

| SCORE PER DIMENSION | STAGE |
|---|---|
| ≤ 5 | Nascent |
| 6-10 | Early |
| 11-15 | Established |
| 16-19 | Mature |
| 20 | Visionary |

For instance, if you receive a score of 11 in the Organization dimension of the assessment, you are in the Established stage for that particular dimension. You should expect to see your scores vary for the different dimensions. Analytics programs don't necessarily evolve at the same rate across all of the dimensions. For example, your company might be more advanced in terms of bringing data sources together than it is in analyzing them or governing this data.

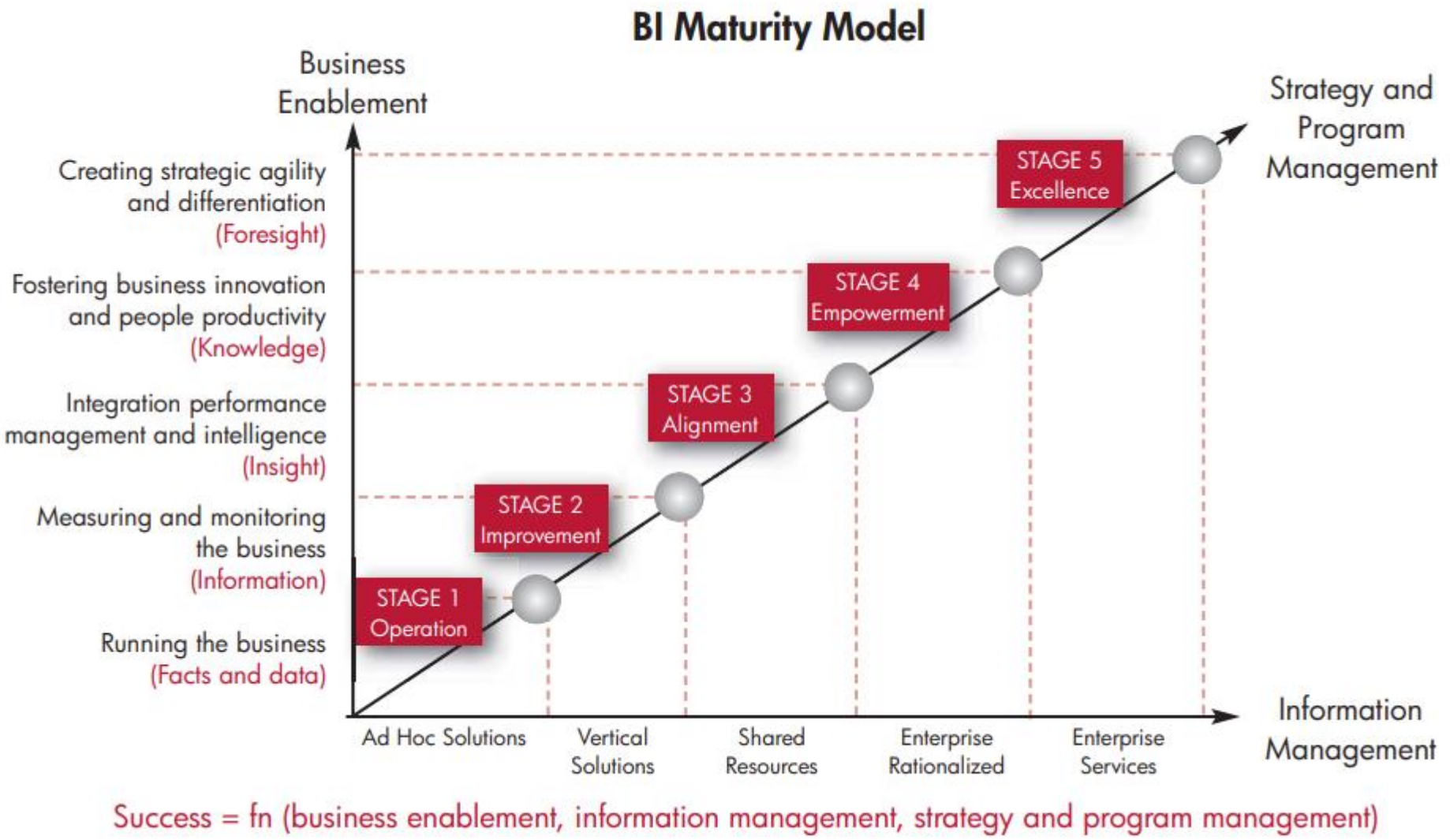
The Chasm can be overlaid between the Established and Mature stages.

When you complete the assessment, you might see scores like this:

| DIMENSION | SCORE | STAGE |
|---|---|---|
| Organization | 10 | Early |
| Resources | 7 | Early |
| Data Infrastructure | 11 | Established |
| Analytics | 4 | Nascent |
| Governance | 7 | Early |

Total Score: 39

# HP BI MM



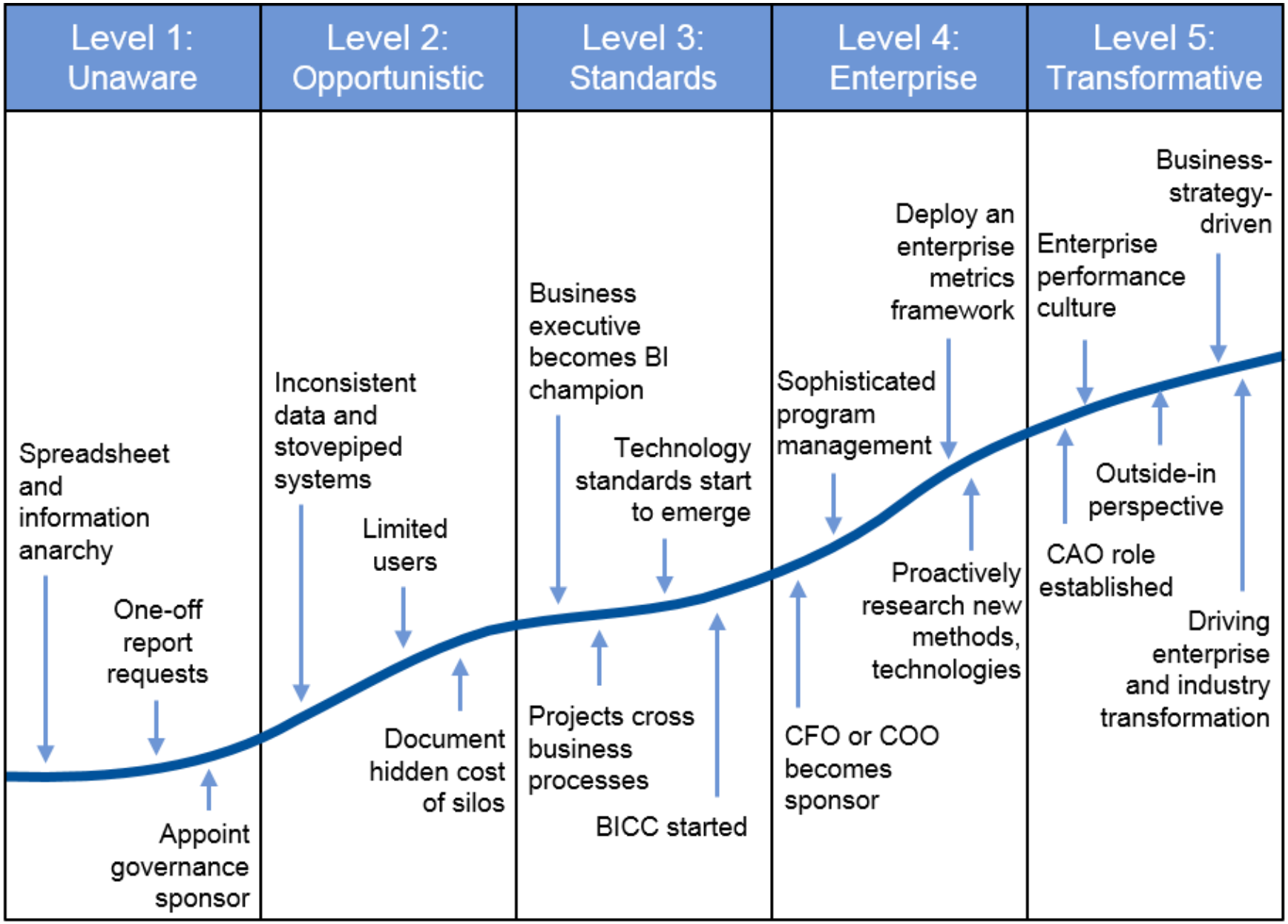

Success = fn (business enablement, information management, strategy and program management)

# HP BI MM



| |
|---|
| **Stage 1 –** Operation (Running the business) – involves ad-hoc solutions focused at local demand. No formal BI strategy or analytics. |
| **Stage 2 –** Improvement (Measuring and monitoring the business)-BI more structured. Analytics are being used. DW are still focused on specific business units. |
| **Stage 3 –** Alignment-BI aligned with strategic goals. Information integration across subject areas. Data quality and governance are increasing their importance. Organization evolved from BI project management to BI program management. |
| **Stage 4 –** Empowerment – Self-service tools being used at all levels. Advances analytics are implemented. Single version of truth across the organization.BI is critical. |
| **Stage 5 –** Excellence (Change the business) – predictive analytics for most business decisions. SOA for information delivery. Analytics to users are delivered quickly. |

# Gartner BI Maturity Model




BI = Business intelligence

BICC = BI competency center

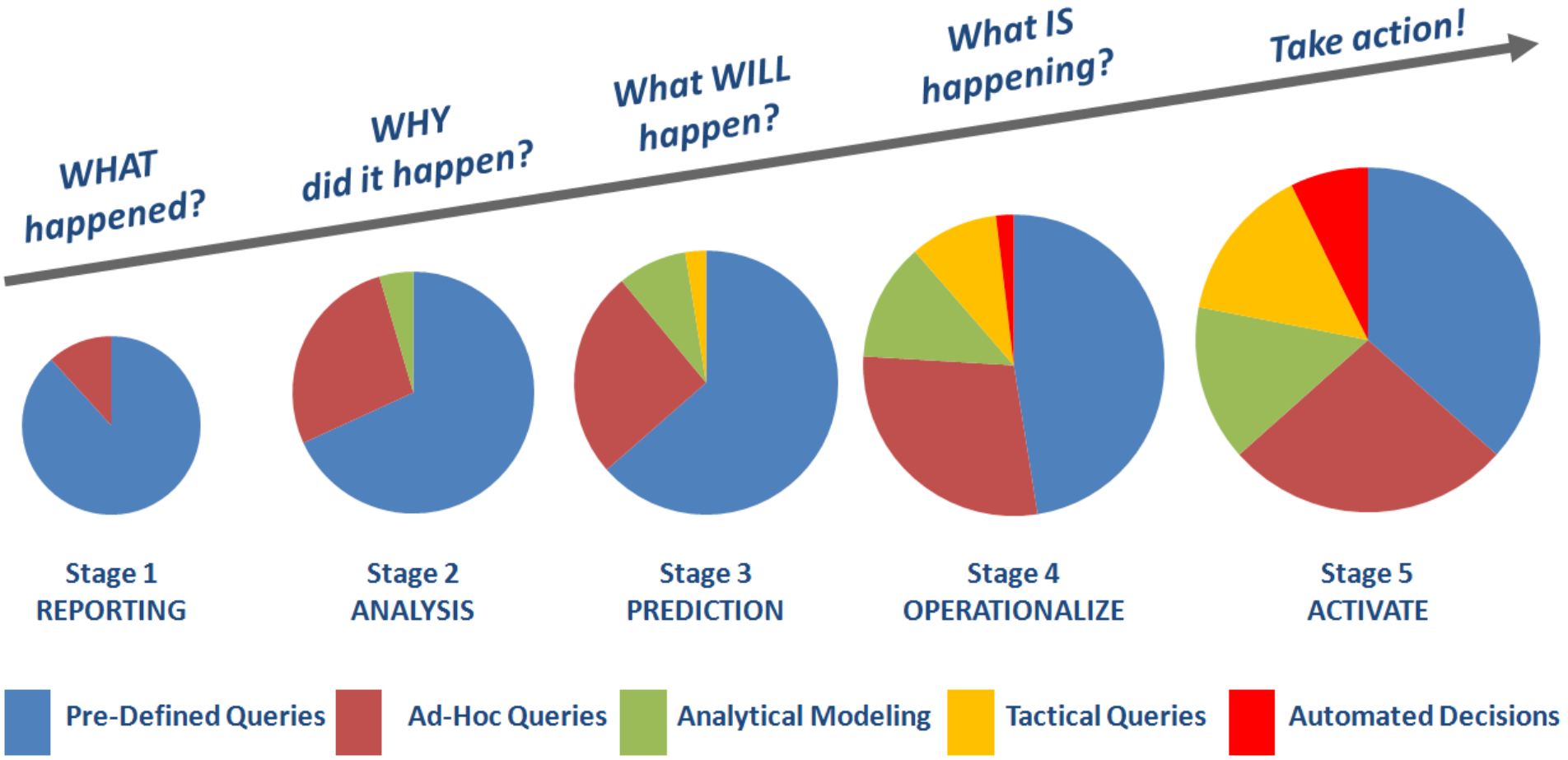
# Gartner BI MM

| |
|---|
| **Level 1 –** Unaware– ad-hoc BI and analytics. No formal decision-making process. No IT infrastructure. |
| **Level 2 –** Opportunistic- business units use BI individually. Each project/unit has its own infrastructure. |
| **Level 3 –** Standards- Coordination improves. Projects across multiple business processes. A BI competence center is created. Technology standards start to emerge for infrastructure, DW and BI platforms. |
| **Level 4 –** Enterprise – Top executives sponsor BI. Performance metrics guiding strategy have been defined. BI supports companywide decision processes. |
| **Level 5 –** Transformative- BI and analytics are a strategic initiative run and supported by business and IT. They are used to generate revenue. Users across all levels, including customers and partners. |

# Maturity Models



## Evolution of BI capabilities



# Methodologies



- **What is it and why?**
  - Every process needs to be repeatable.
  - It accumulates experience in a standard process.
  - Created by experts.
  - It helps in planning and process management.
  - Reduces initial fear
    - There is a standardized process
    - Reduce dependency on specific staff

# Methodology

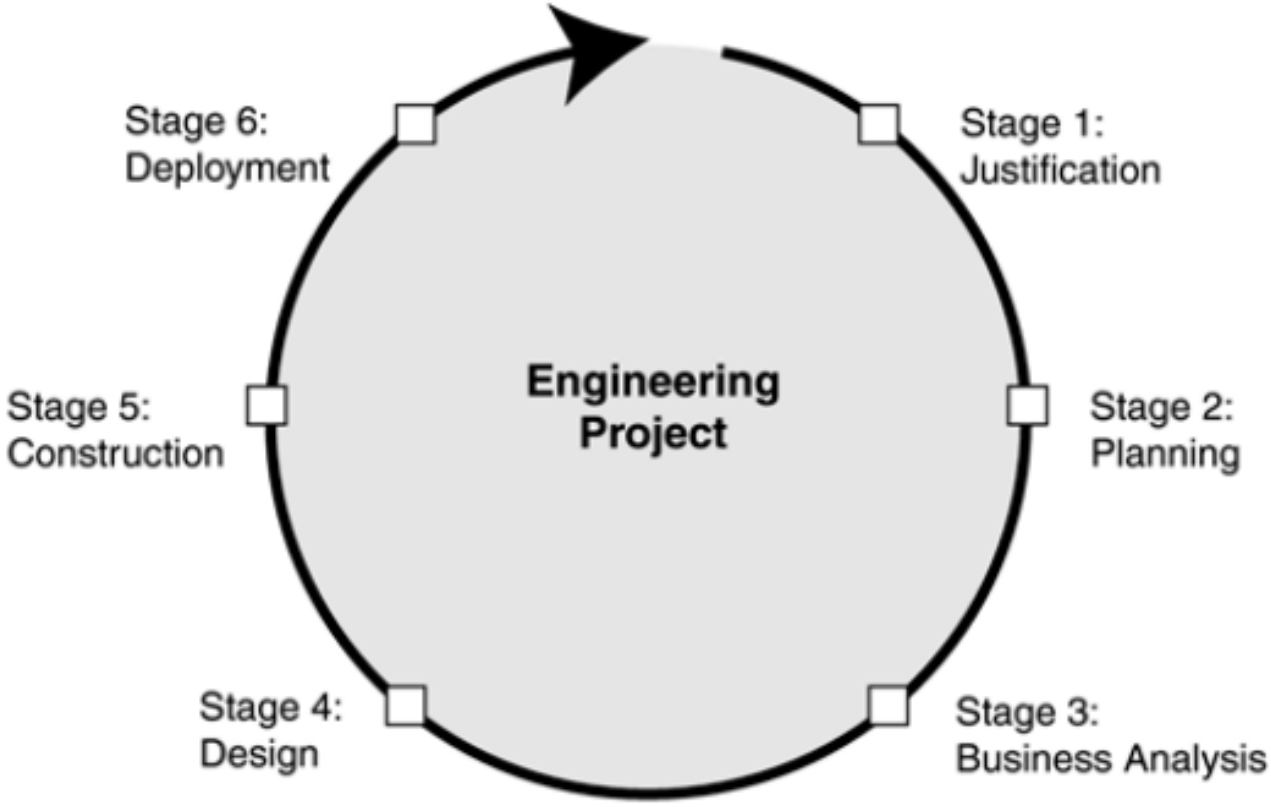


- **Focus on management and development**
  - Management: PMBok, Agile, ...
  - Management+Development: Larissa, Kimball, Inmon, SAFE, QPM (QlikView), ASAP (SAP), … other tools
- **CRISP-DM: Generic tasks for data mining**
  - Cross Industry Standard Process for DM
  - Proposed by a consortium (SPSS, NCR, AG, OHRA)
  - Other focus on DM:
    - 5A'S, critikal, CAT (Clementine Application Template), SEMMA (SAS, Sample, Explore, Modify, Model, Assess) , associated to tools

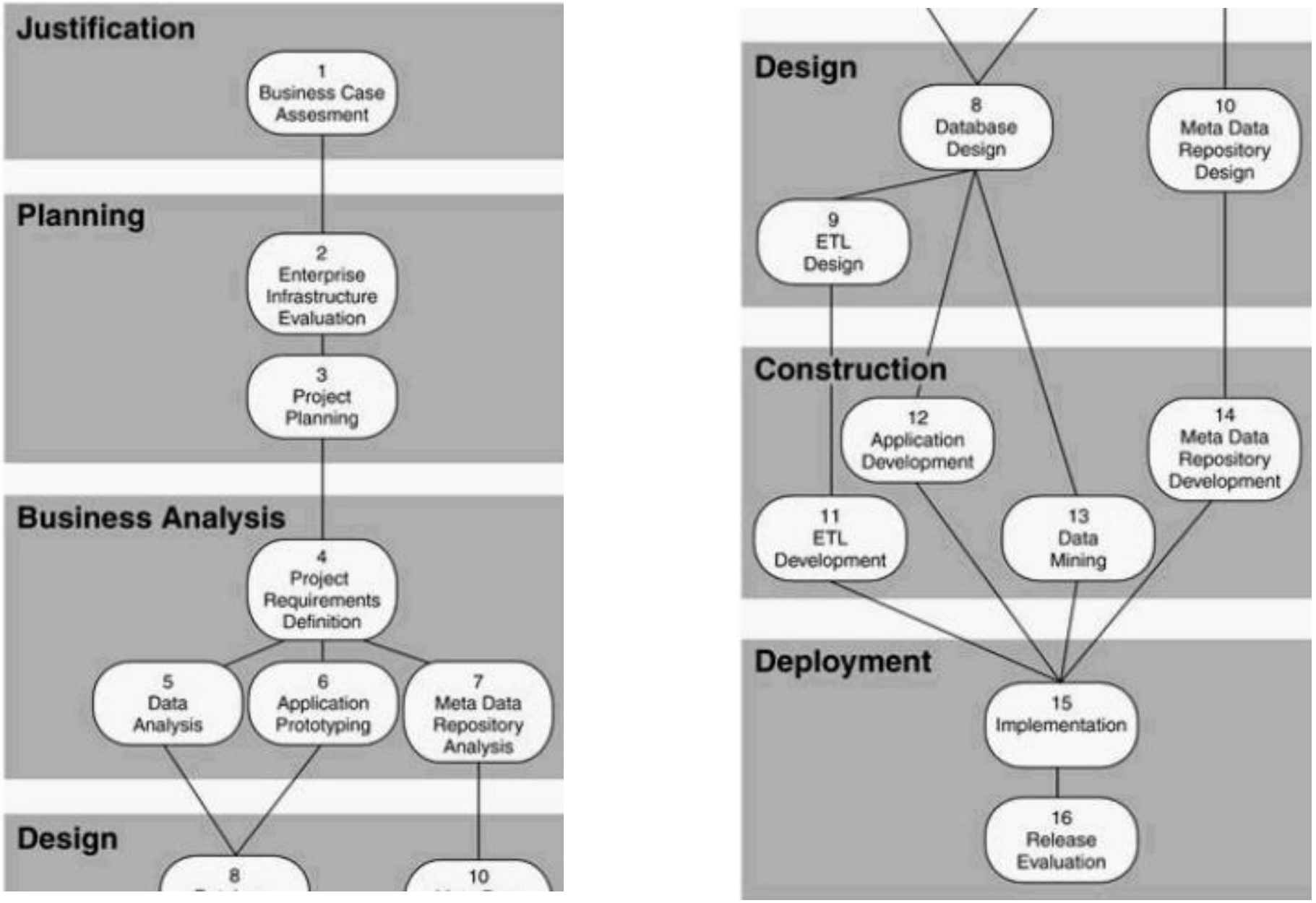
# Larissa Moss: Roadmap to BI with 6 stages and 16 steps

- Defines stages, steps, roles, standards and deliverables.
- 6 iterative stages from inception to deployment.
- Agile and adaptive, promotes subprojects.

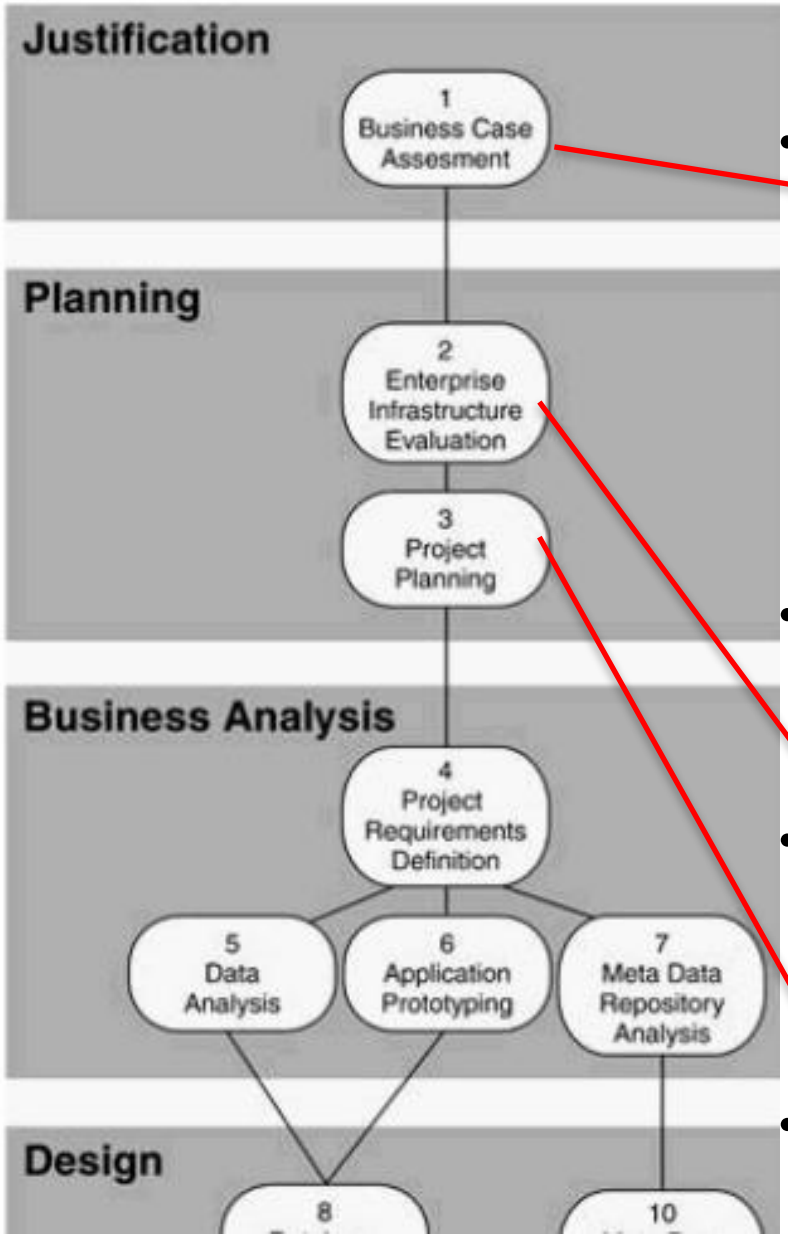
**Figure 0.1. Engineering Stages**



# Larissa Moss: Roadmap to BI with 6 stages and 16 steps

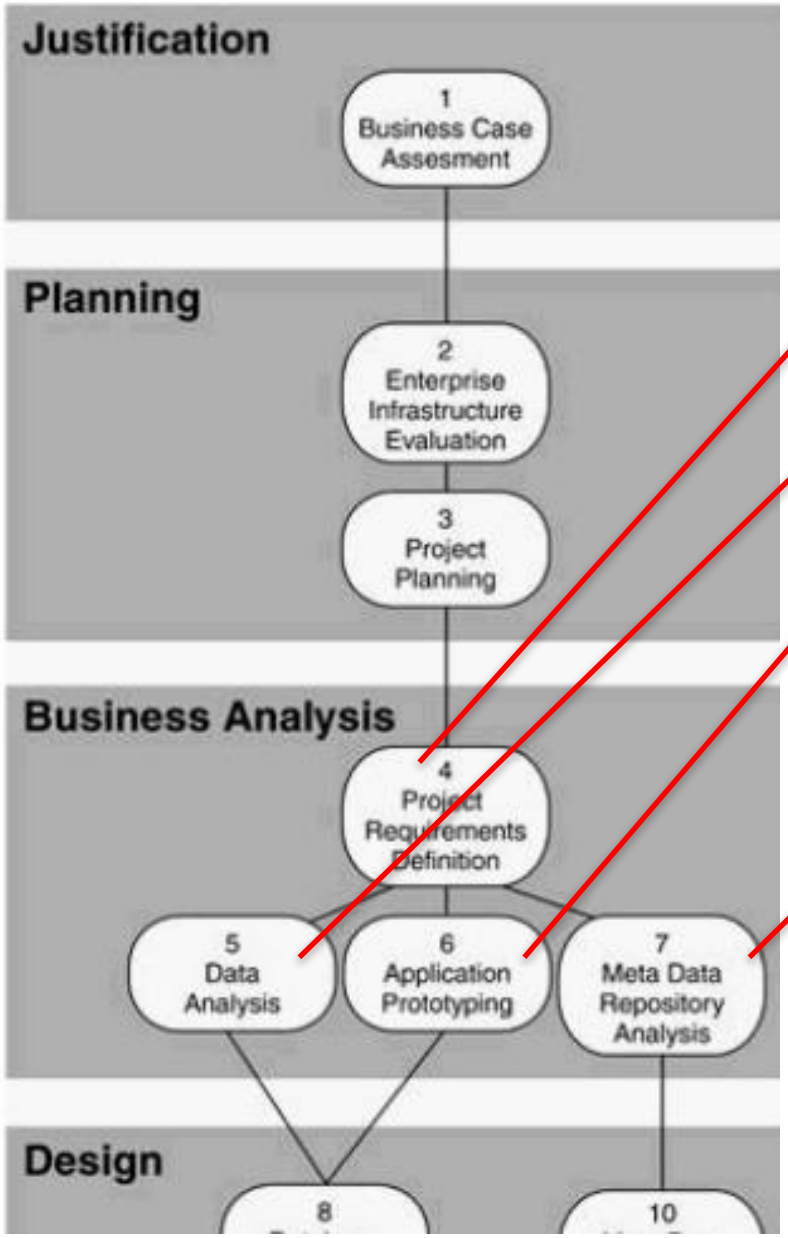


# Larissa Moss: Roadmap to BI with 6 stages and 16 steps



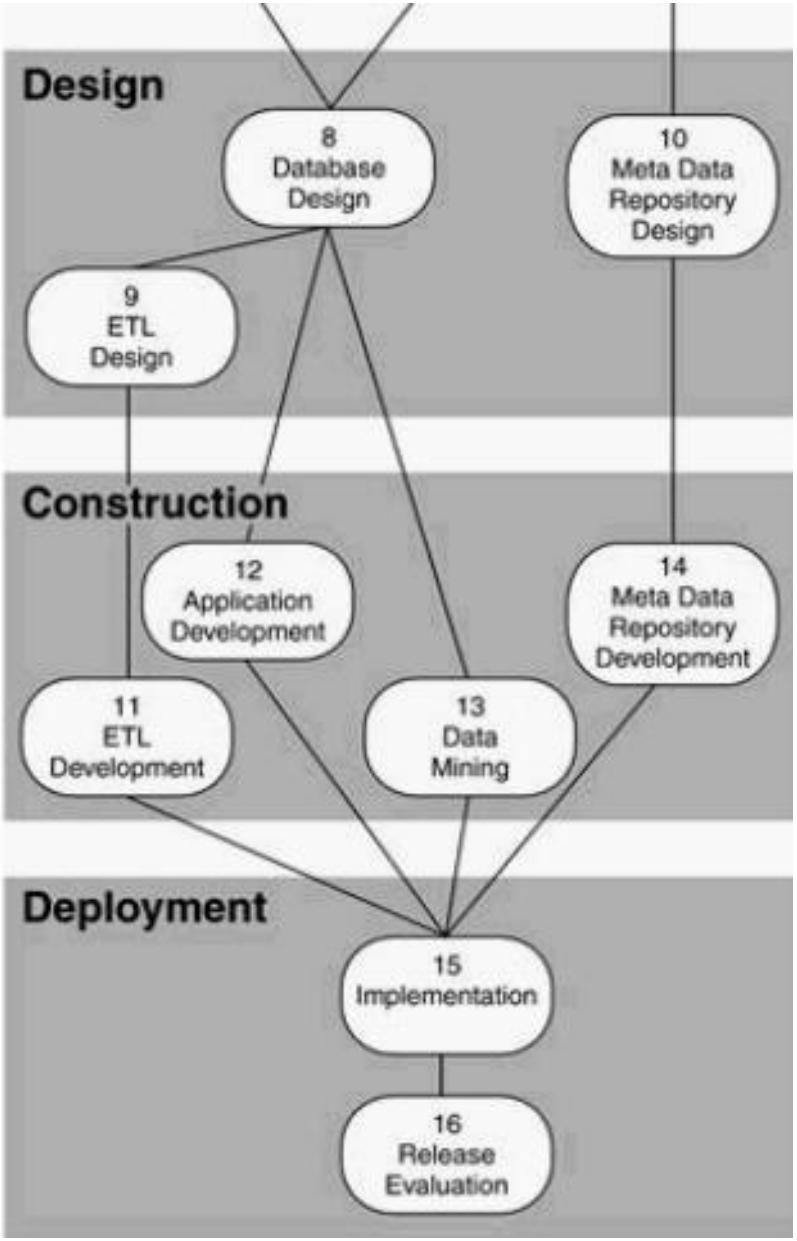

- **1. Justification:** Assess the business need that gives rise to the new engineering project.
- Step 1: Business Case Assessment
  - Defines business problem or business opportunity and proposes a BI solution.
  - Each BI application release should be cost-justified and should clearly define the benefits.
- **2. Planning:** Develop strategic and tactical plans, which lay out how the engineering project will be accomplished and deployed.
- Step 2: Enterprise Infrastructure Evaluation
  - Technical: hardware, software, networks, …
  - Non-technical: procedures, methodologies, …
- Step 3: Project planning: scope, staff, budget, technology, business representatives, …

# Larissa Moss: Roadmap to BI with 6 stages and 16 steps




- **3. Business analysis:** Perform detailed analysis of the business problem or business opportunity to gain a solid understanding of the business requirements for a potential solution (product).
- Step 4: Project Requirement s Definition
  - Managing and specifying scope, user needs, …
- Step 5: Data analysis, avaliability, quality, …
- Step 6: Application prototyping
  - Helps in the requirements definition and avoid risks
- Step 7: Meta Data Repository analysis
  - Technical meta data needs to be mapped to the business meta data.
  - Meta data describes an organization in terms of its business activities, the business objects, and rules on which the business activities are performed.
    - Ex: definitions, units, relationships, sources….

# Larissa Moss: Roadmap to BI with 6 stages and 16 steps




- **4. Design:** Conceive a product that solves the business problem or enables the business opportunity.
- **5. Construction:** Build the product, which should provide a return on investment within a predefined time frame.
- **6. Deployment:** Implement or sell the finished product, then measure its effectiveness to determine whether the solution meets, exceeds, or fails to meet the expected return on investment.

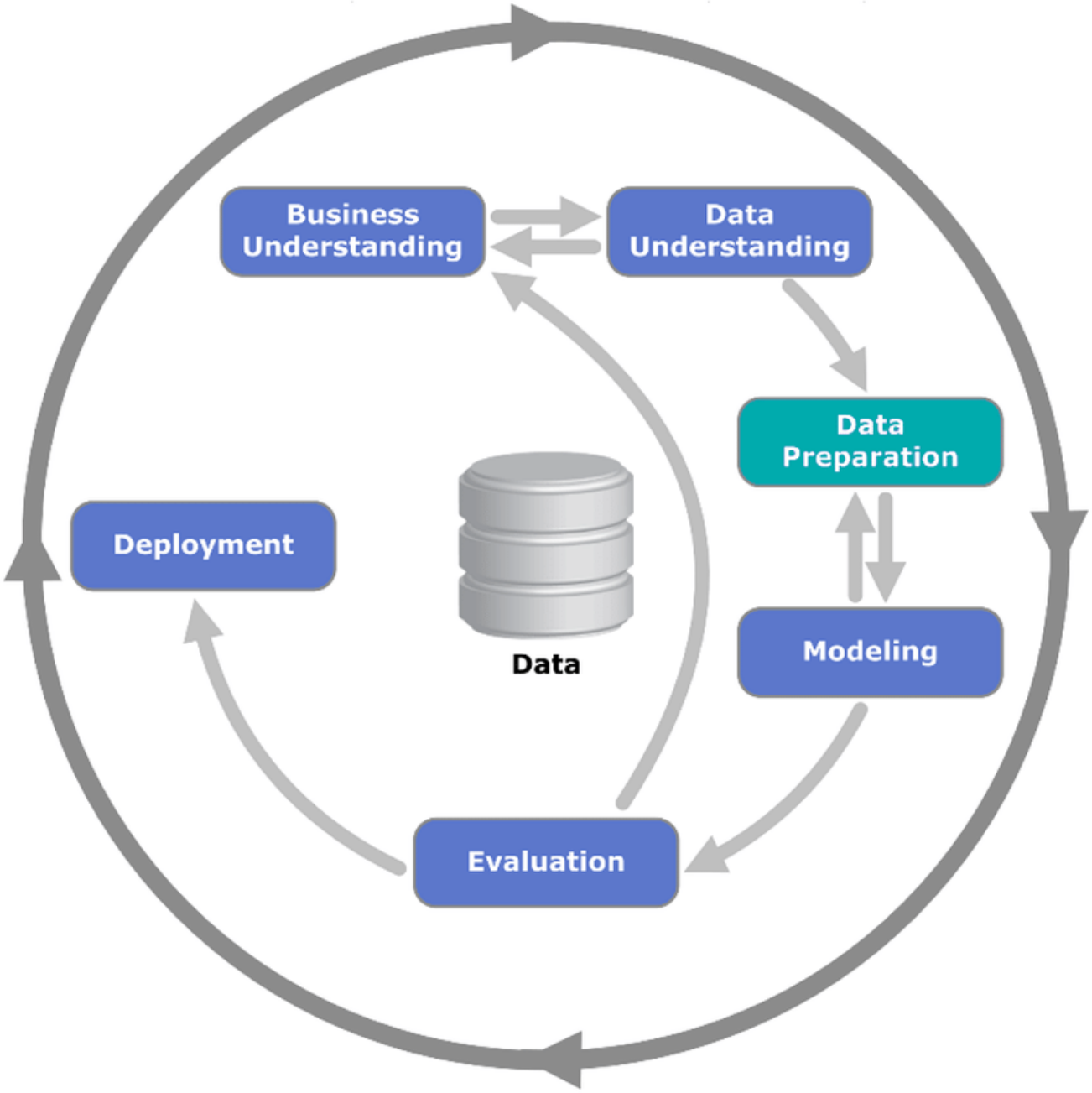
# Methodology – CRISP-DM

- Process without owner
- Application-independent and
- Context-independent
- Tool-independent
- It is a guide that consider the problem and techniques
- Based on experience (developed in several workshops)


- 2 documents:
  - Reference model: describes phases, tasks and outputs.
  - User guide: practical application tips, check list

# Methodology – CRISP-DM

## Phases



# Methodology – CRISP-DM

Data Science Process Alliance

Frameworks | Blog | Consulting | About

# Executing Data Science Projects

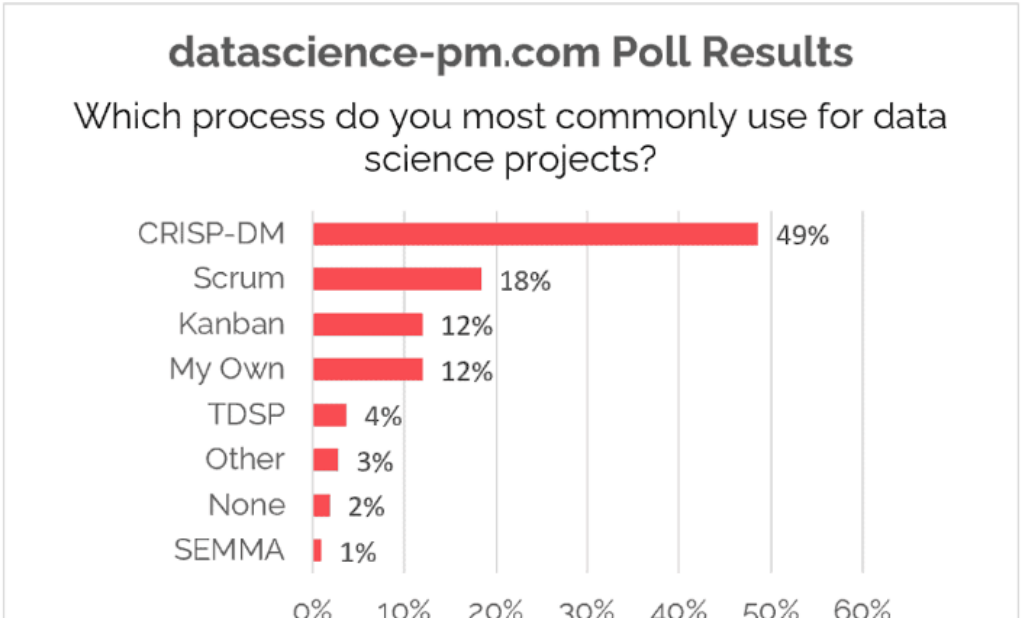
BY JEFF SALTZ | LAST UPDATED APR 10, 2024 | LIFE CYCLE

Table of Contents



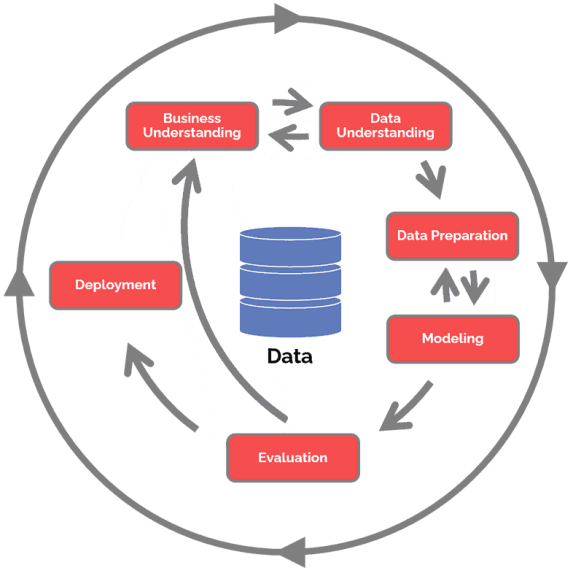
During the past few months, we conducted a poll to see what project management framework teams used to help execute their data science projects.

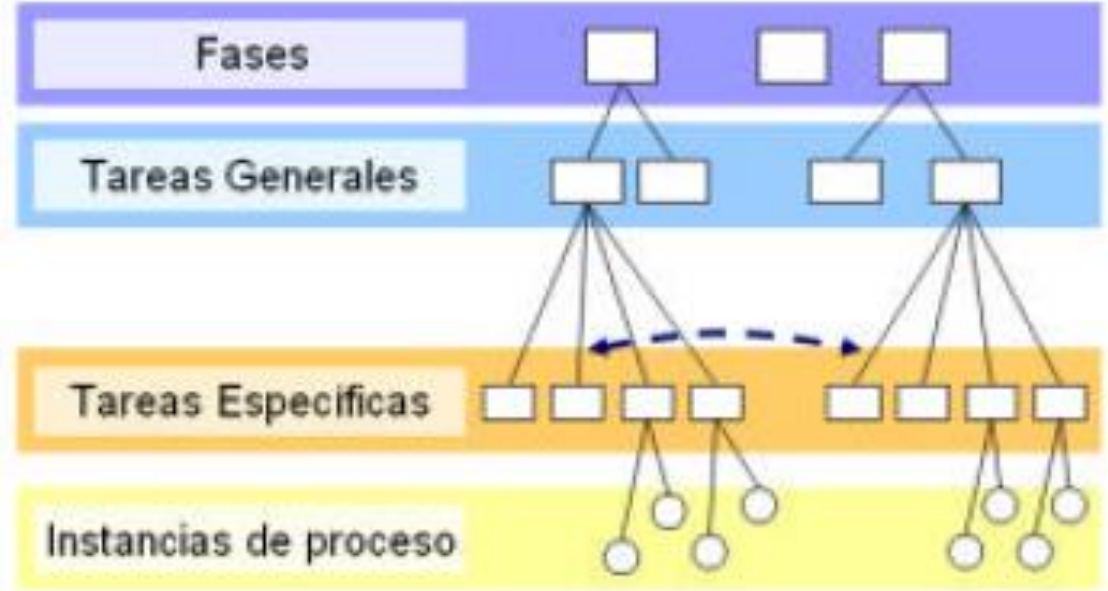
Based on our survey of 109 respondents, nearly half of the respondents most commonly use CRISP-DM. This was followed by Scrum, Kanban and "My Own". See results below.



# Methodology – CRISP-DM

- **4 hierarchical levels**
  - General task; e.g.: cleansing
  - Spectific task; e.g.: null cleansing



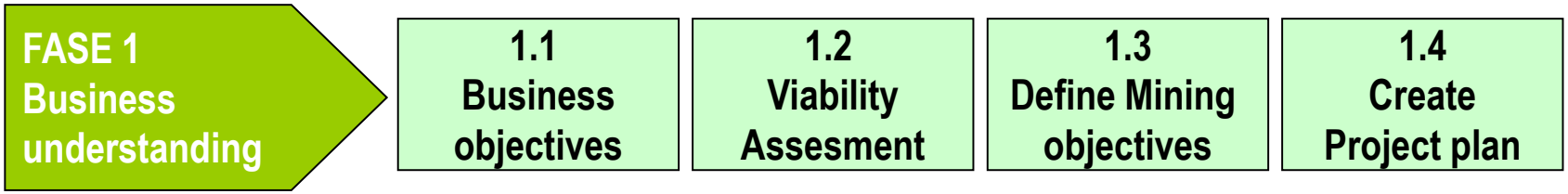

# Methodology – CRISP-DM

3 primeras fases repsentan el 80% del proyecto aprox, la tercera es la que mas tiempo lleva

| | | | | | |
|---|---|---|---|---|---|
| **PHASE 1 Business understanding** | 1.1 Business objectives | 1.2 Viability Assesment | 1.3 Define Mining objectives | 1.4 Create Project plan | |
| **PHASE 2 Data understanding** | 2.1 Gathering | 2.2 Description | 2.3 Exploration | 2.4 Quality checking | |
| **PHASE 3 Data preparation** | 3.1 Select | 3.2 Clean | 3.3 Modify | 3.4 Integrate | 3.5 Format |
| **PHASE 4 Modeling** | 4.1 Select techniques | 4.2 Generate Test design | 4.3 Build model | 4.4 Asses Model | |
| **PHASE 5 Evaluation** | 5.1 Interpretation Of results | 5.2 Review the process | 5.3 Defining Next steps | | |
| **PHASE 6 Deployment** | 6.1 Plan Deployment | 6.2 Plan Monitoring and maintenance | 6.3 Generate Final informe | 6.4 Review project | |

# Methodology – CRISP-DM

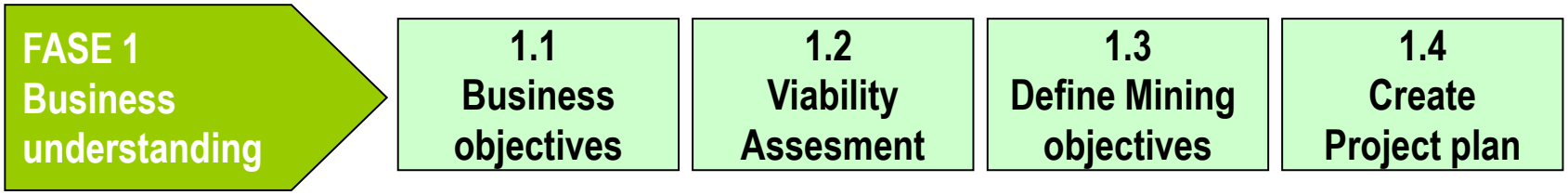
un gran fallo que se puede tener es considerar que comprender el negocio es un objetivo secundario, cuando es de los aspectos más importantes




- 1.1 Defining business objectives
  - Knowing what the customer wants from the business point of view
    - Is it to increase profitability campaign?
    - Is it to improve inventory management?
  - Show factors affecting the project
  - Establish success criteria
    - What does make the project successful?: Indicate objective measures
- 1.2 Viability Assesment analisis dafo - debilidades, fortalezas, amenazas, oportunidades
  - Indicate resources (people, data, software, ...), constraints, assumptions, requirements, and other factors: ARE WE ALLOWED TO USE DATA?
  - Cost-benefit.
    - Equipment costs, human resources, etc.
    - ROI Benefits ...
  - Risks: Data not available, not available knowledge, tools

# Methodology – CRISP-DM



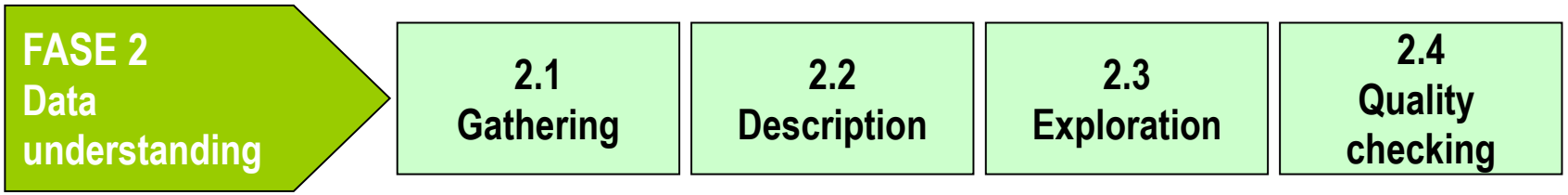



- 1.3 Define mining objectives
  - Specific objectives of the problem:
    - Translate customer target into mining goals.
    - Eg segmentation, sequential patterns, and so on.
  - Establish technical objectives
    - Translate goals into parameters of the output: rate of success, failure prediction, error cost, etc.

- 1.4 Create project plan
  - Establish steps:
    - Step: duration, resources required, inputs, outputs.
    - Scheduling and risks management
  - Initial selection of techniques and tools

# Methodology – CRISP-DM






- 2.1 Gather
  - Define sources, integration, localization, problems and solutions, …

la parte de integracion suele ser muy complicada

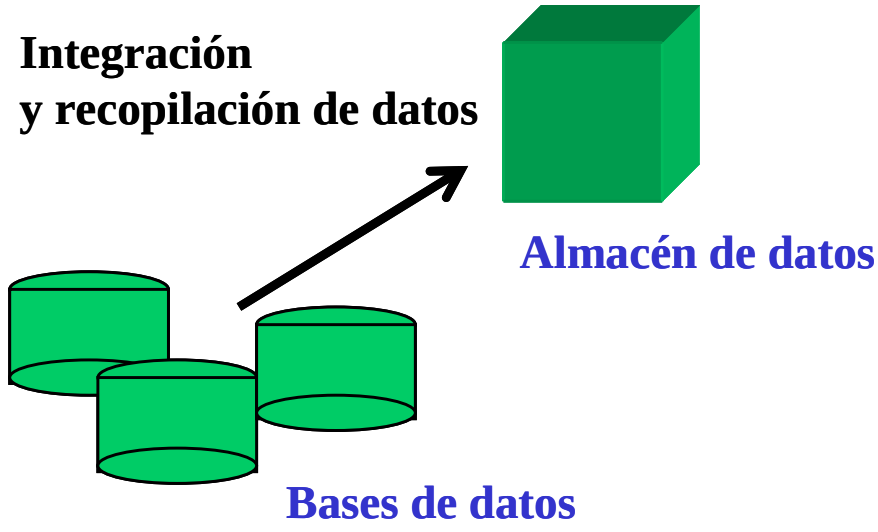
- 2.2 Describe
  - Formats, keys, amount…
- 2.3 Explore
  - Query, visualize, report
  - Value distributions, aggregations of data, statistics (properties, mean, std,...)
- 2.4 Quality check
  - All type of data? All classes represented? Enough data? Complete data? Null or impossible values?

# Methodology – CRISP-DM



| FASE 3 Data preparation | 3.1 Select | 3.2 Clean | 3.3 Modify | 3.4 Integrate | 3.5 Format |
|---|---|---|---|---|---|

- It involves most of the total time
- 3.1 Selecting Data
  - Selecting tables, attributes and rows
  - Volumen suitable for tools
  - selection:
    - Partitioning the data set:
    - Training data
    - Test Data
    - Validation Data
  - Sampling Data



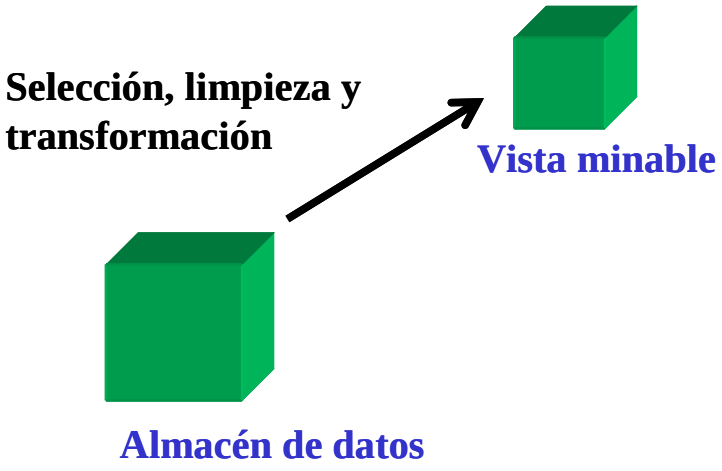

- 3.2 Cleaning: incomplete and erroneous data records:
  - Improving data quality
  - Select subset of data, replace null
  - Complete, delete or ignore

# Methodology – CRISP-DM

esto es como un 50% del total del proyecto

| FASE 3 Data preparation | 3.1 Select | 3.2 Clean | 3.3 Modify | 3.4 Integrate | 3.5 Format |
|---|---|---|---|---|---|

- 3.3 Transformation:
  - Choose more relevant attributes
  - Deriving most significant new features from the original
  - Discretize, map, normalize, ...
- 3.4 Integrate: **DATAWAREHOUSE**
- 3.5 Format: if needed by techniques
- Known as ETL
  - Extraction, Transformation, Load
  - Often uses a graphical model
  - Frecuency of execution
  - It involves:
    - Execution monitoring
    - Recording
    - Exceptions and errors

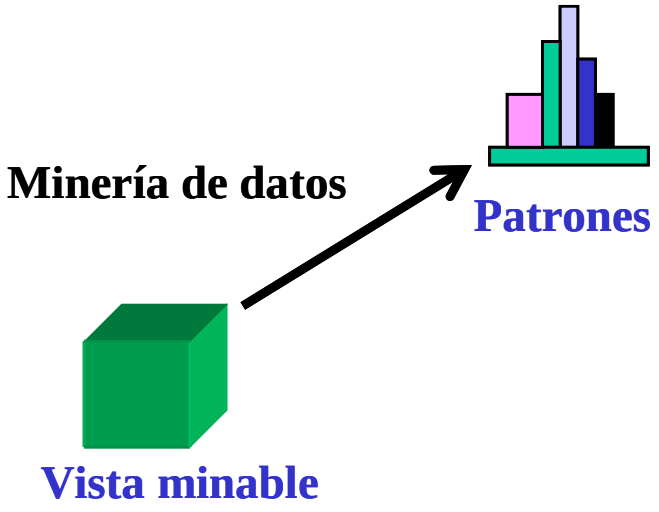


# Methodology – CRISP-DM



| FASE 4 Modeling | 4.1 Select techniques | 4.2 Generate Test design | 4.3 Build model | 4.4 Asses Model |
|---|---|---|---|---|

- We assume a minable view: A table containing the relevant attributes, labeled as inputs or outputs
- 4.1 Select techniques: considering
  - Appropriateness to the problem
    - Classification, prediction, clustering, association, dependencies
  - With adequate data
  - Meeting the requirements of the problem
  - Executable in time
  - Knowledge of the technique

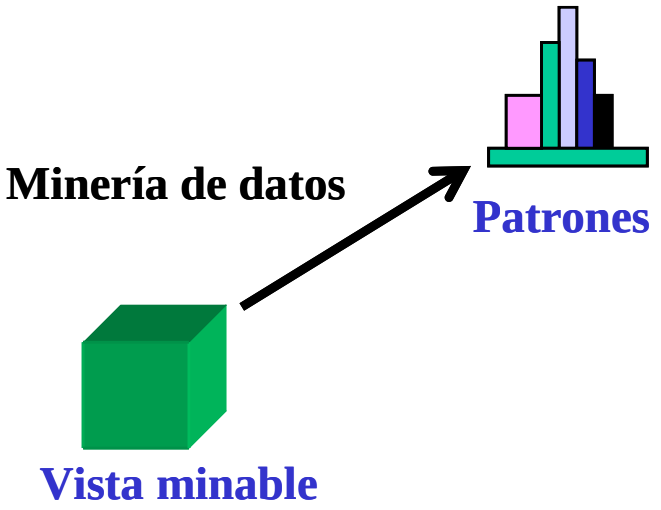


# Methodology – CRISP-DM



| FASE 4 Modeling | 4.1 Select techniques | 4.2 Generate Test design | 4.3 Build model | 4.4 Asses Model |
|---|---|---|---|---|

- 4.2 Design tests
  - Establish training, testing and validation
  - Establish criteria for goodness of models
- 4.3 Creating model
  - Set parameters
  - run
- 4.4 Evaluate model
  - Meets goodness?



# Methodology – CRISP-DM

**FASE 5**
**Evaluation**

| 5.1 Interpretation Of results | 5.2 Review the process | 5.3 Defining Next steps |
|---|---|---|

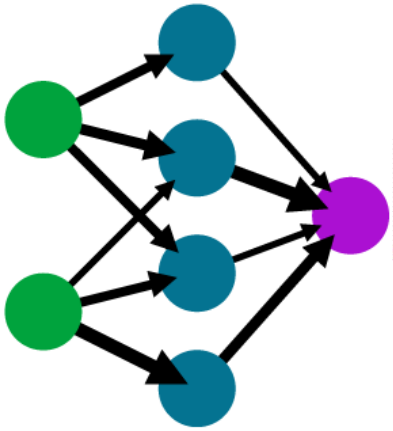
- 5.1 Interpretate the results
  - Is the problem solved?
  - Is the answer appropiate?
  - Is it valid?
  - Is the business objective well defined?
  - Is the knowledge new?
  - Is the model useful?
  - Is it better than we had?
  - Are there too many patterns?
- 5.2 Review the process
  - Any error from the technical point of view?
  - Have we overoolked anything?
- 5.3 Define next steps
  - Iterate? Deploy? Rebuild?

**[38,44]sem**

**[32,36]sem**

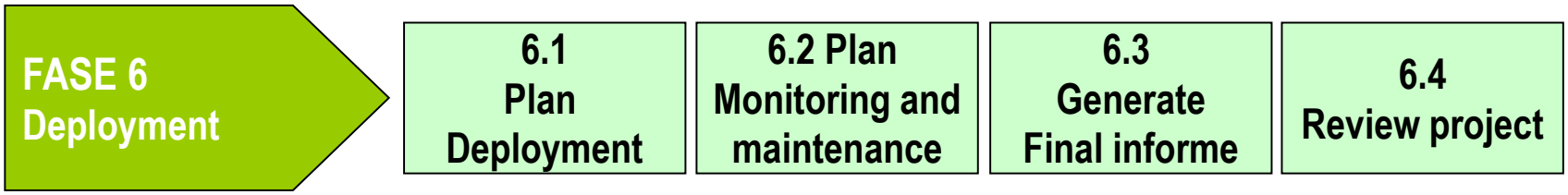
**(hair=blonde)**
**(profession=actress)**

**carrier=no**



# Methodology – CRISP-DM






- 6.1 Plan deployment
  - Who are the users?
  - How and when the model will be used?
  - How is it deployed? As a tool?
  - A computer program is necessary? Paper?
  - Strategic system for a doctor: hypothesis proposal
    - Screening, diagnose, prognose
- 6.2 Plan monitoring and maintenance
  - Is it being used? Is it properly used?
  - Updatable models?
- 6.3 Generate final report
- 6.4 Review the project
  - Strong and weak points, any aspect that can be improved?…

# Methodology – CRISP-DM
# Practical example

- **A Practical example: Research center focused on using remote sensing for assessing water quality in the Rías**
  - Business Understanding
    - Objective: Monitor and forecast water quality in the Galician estuaries using satellite data (business objective).
    - Viability Assessment:
      - Does this project align with the lab's strategy? + estimated costs & expected ROI + human and technical resources + estimated time + risks + expected outcomes.
    - Data mining objectives: what we want to obtain?: descriptions, predictions, classifications, detect anomalies, time-series,…
    - Create project plan:

      work breakdown structure

      - Scope + define region of interest and timeframe + WBS & Tasks + Resources + milestones + internal and external communication, …

# Methodology – CRISP-DM
# Practical example



- **A Practical example: Research center focused on using remote sensing for assessing water quality in the Rías**
  - Data Understanding
    - Gathering:
      - Identify data sources: satellite, in-situ, external (GIS layers, climatologies), etc.
      - Obtain the data
    - Description: Formats, quality, we have to understand the data.
    - Exploration: statistical summaries, visual inspection, explore correlations, etc.
    - Data quality: outliers, missing data, does satellite and in-situ data correlate? Are the different datasets consistent?

# Methodology – CRISP-DM
# Practical example

- **A Practical example: Research center focused on using remote sensing for assessing water quality in the Rías**
  - Data Preparation (overlap with ETL):
    - Select: spatial, temporal + satellite bands
    - Clean: what we do with missing data and outliers? How to remove/reduce noise?
    - Modify: transform data (e.g. atmospheric correction to satellite data), normalize, feature engineering (e.g. to generate value added products), …
    - Integrate: different data sources, spatial and temporal integration,…
    - Format: ensure we are using the same coordinate systems, raster vs vector, storage formats, …

# Methodology – CRISP-DM
# Practical example



- **A Practical example: Research center focused on using remote sensing for assessing water quality in the Rías**
  - Modeling (usually very iterative):
    - Select techniques: based on the objectives, available data and the desired outputs: classification, regression, clustering, association, …
    - Generate test design: training, validation and test datasets + what metrics are we going to use?
    - Build model: train the model and adjust hyperparameters
    - Assess model: assess performance + understand and interpret the model (e.g. when it performs poorly)

# Methodology – CRISP-DM
# Practical example

- **A Practical example: Research center focused on using remote sensing for assessing water quality in the Rías**
  - Evaluation:
    - Interpretation of the results: how good is the model for achieving the objectives of the project? Do we detect spatial/temporal patterns? Are there factors such as population/weather affecting water quality? Do the experts agree with the outcomes? How do the results compare to baselines?
    - Review of the process: we review the data, the techniques, the metrics, the feedback,…
    - Defining next steps: Does the model needs improvements? If it works well, deploy? Do we need to retrain the model with new data? Update the algorithm? Expand it to other regions? …

# Methodology – CRISP-DM
# Practical example



- **A Practical example: Research center focused on using remote sensing for assessing water quality in the Rías**
  - Deployment:
    - Plan deployment: do we need to integrate the model into an existing system (e.g. in the cloud, GIS)? How do we plan to access the data to feed the model? What are the outputs and in which format?
    - Plan monitor and maintenance: how often we validate our system? We need to define metrics to assess model performance and trigger retraining. What happen in case or errors? Do we need a data backup?
    - Generate final report: describe project, data sources, procedures, results, recommendations, etc.
    - Review project: did we learn anything? Things that went well, challenges, bad things. We need to use this experience to improve future efforts.

# Methodology for DM- Proposals



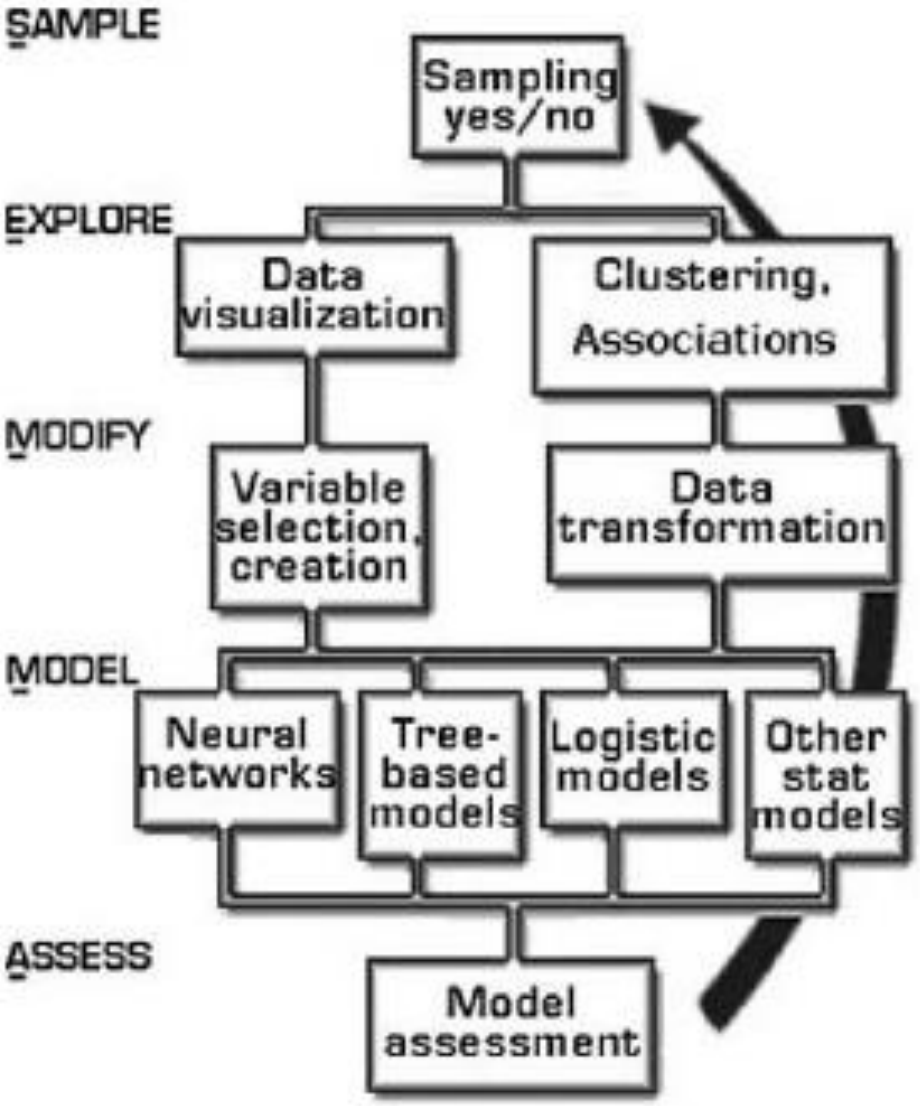
- **CATs: Clementine Application Templates: [CATs]**
  - Specific libraries with best practices for specific applications: mapping to a project type
  - Following the CRISP-DM standard.
  - Each flow CAT is assigned to a phase of CRISP-DM.
- **Templates:**

ejemplos de plantillas

- Telco CAT - loyalty and customer retention for telecoms
- CRM CAT - understand and predict customer migration between segments,
- Microarray CAT - specific functions for biological applications, finding genes for therapeutic purposes, predict genetic diseases
- Fraud CAT - predict and detect fraudulent transactions, claims, taxes and so forth.
- Web CAT - predict visitor behavior, access and merge web log data
- Crime CAT- Identify and predict areas of high crime, offender type identification

# Methodology for DM- Proposals

- SEMMA: SAS Entreprise Miner proposal for data mining projects
- Sampling:
  - A good sampling strategy almost garantees good patterns
  - Rare patterns are statistically discovered
  - Subsets: training, validation, test
- Explore: simplify the problem




**What main methodology are you using for your analytics, data mining, or data science projects ?** [200 votes total]

2014 poll 2007 poll

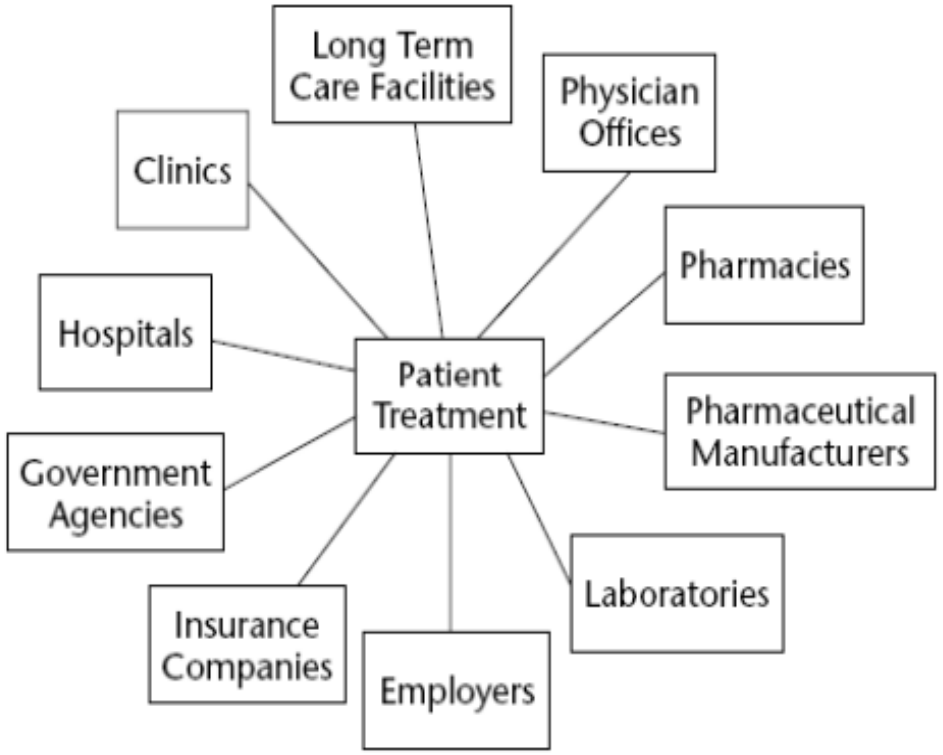
| | |
|---|---|
| CRISP-DM (86) | 43% / 42% |
| My own (55) | 27.5% / 19% |
| SEMMA (17) | 8.5% / 13% |
| Other, not domain-specific (16) | 8% / 4% |
| KDD Process (15) | 7.5% / 7.3% |
| My organizations' (7) | 3.5% / 5.3% |
| A domain-specific methodology (4) | 2% / 4.7% |
| None (0) | 0% / 4.7% |

Source: Kdnuggets.com

# BI applications in health

ejemplo:

- **Application**
  - Definition of Health policies
  - Detection of inefficient services
  - Fraud detection by health providers




como mapa mental pero bien estructurado

# Example domain: Business Intelligence in Medicine

PATIENT CARE

MANAGEMENT

RESEARCH

# Example domain: Business Intelligence in Medicine

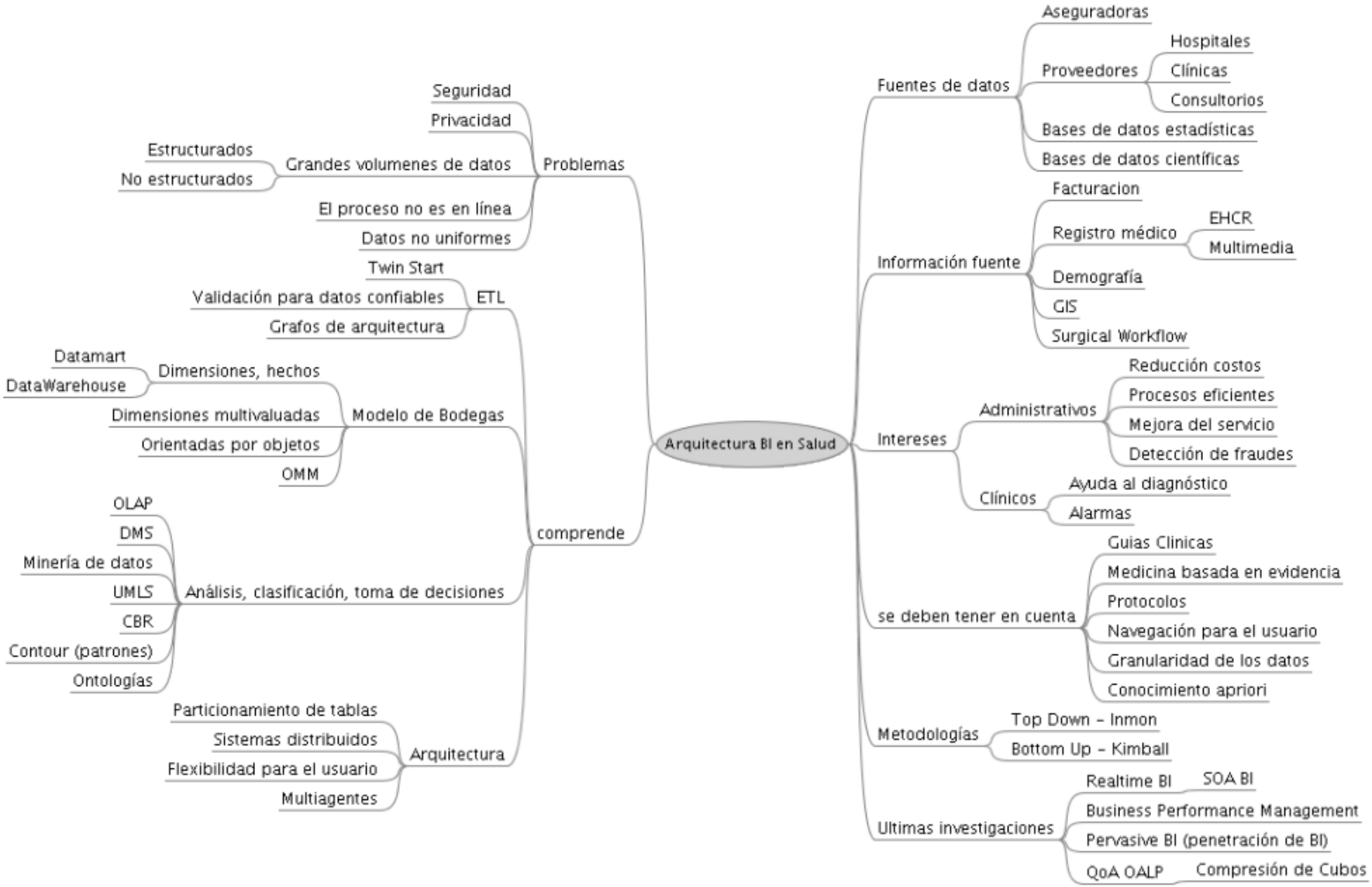
| FASE 3 Data preparation | 3.1 Select | 3.2 Clean | 3.3 Modify | 3.4 Integrate | 3.5 Format |
|---|---|---|---|---|---|

- Besides:
  - Imprecise data, noise, imcomplete data
  - Subjective data due tu human interaction
  - High volumen of data
  - Complex and dynamic data
  - Numeric, nominal, time series, images, video, 3d
  - Text: reports, interpretation, ambiguous expressions, …
  - Difficult to characterize mathematically

- We find
  - Propietary EHR
  - Too many standards
  - Low interoperability

- It implies:
  - Big difficulty to acquire and consolidate changing data

# BI applications in Health [Sánchez,08]

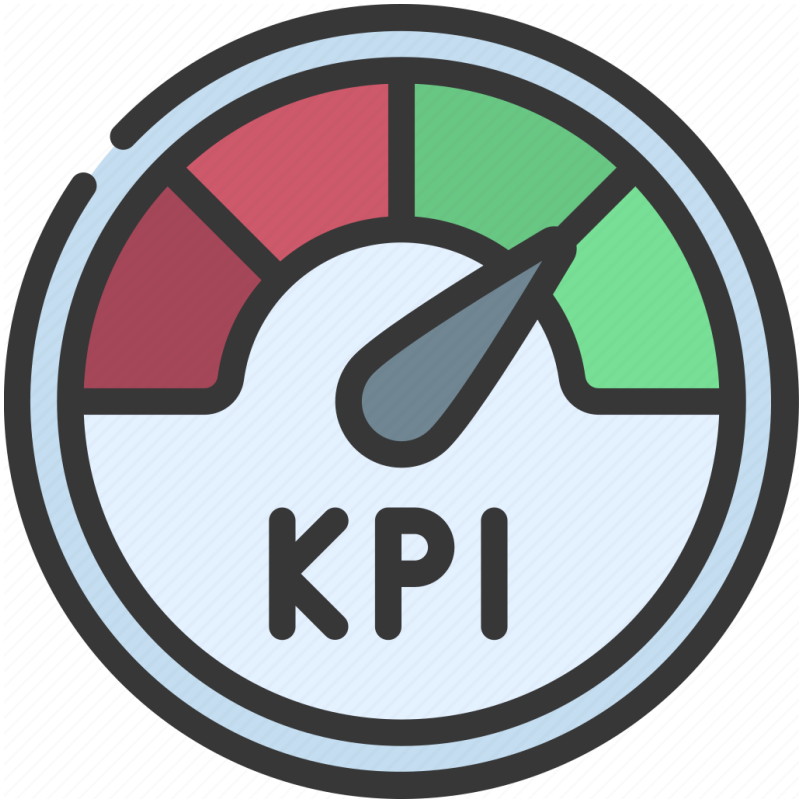


# Business intelligence

## Unit 1b – Key Performance Indicators

# Key Performance Indicator (KPI) definitions

- A lot of words for Indicator, but mainly:
- KPIs represent a set of measures focusing on those aspects of organizational performance that are the most critical for the current and future success of the organization (Parmenter)
- Lets see this in detail



# Goals, success factor, indicator

ejemplo: ganar la champions

ganar la champions

- KGIs, **Key Goal Indicators**, are defined, as measurable targets based on broad objectives such as increasing sales, improving customer satisfaction, raising availability, and raising employees' skill level. KGIs are defined as concrete goals.

buena politica de fichajes, buena cantera, gran numero de abonados, …

- CSFs, **Critical Success Factor**, These are the essential elements needed to meet the KGIs. They identify the most important areas that need attention to ensure that goals are achieved. CSFs clarify what needs to be focused on to reach the KGIs.

define lo bien que van los CSF, que avancemos en ello, kpis asociados a cada csf

- KPIs to measure the current status of business processes related to CSFs. Usually a % or rate. Help assess progress towards meeting the goals and CSFs.

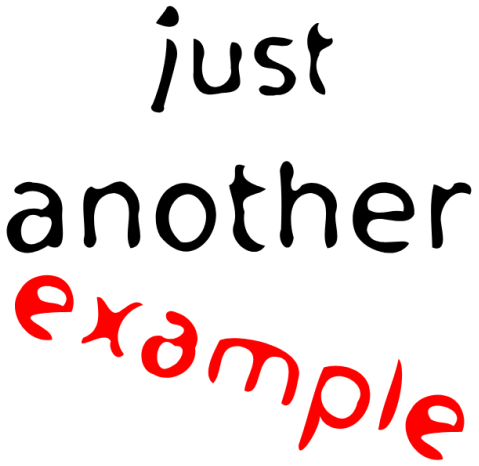
# Goals, success factor, indicator

- Example:

KGI: tener más beneficio
CSFs: publicidad, reducir costes y mejorar eficiencia,
KPIs: llevar desglose de costes, ver idariamente defectos en las latas, …

  - Objective: good health
  - KGI: Improve results medical tests
  - CSF1: Balanced diet  CSF2:Regular physical activity CSF3:Emotional well-being
  - KPI for CSF1: Number days per week following diet
  - KPI for CSF2: Number of days per week doing at least 30min exercise
  - KPI for CSF3: Number of hours of sleep per day.

# Goals, success factor, indicator




- **Example**:
  - KGI: 10% increase in sales
  - CSF: Increase net sales per customer
  - KPI: the customers of accessories of more than $10 >30%
- **Example**:
  - KGI: Increase 10% customer satisfaction
  - CSF:Provide positive customer experience
  - KPI: Call Abandon rate
  - KPI: Call Waiting rate
  - KPI: Service Desk availability

# Target, action

Once KPIs are defined, proceed to the definition of desired targets, decision about concrete actions, and so on. **Critical success factor**: get new customers

**Indicator 1 New sales**

monitorizamos su estado pero no permite tomar decisiones para que haya mejora el mes que viene

**Type**: Outcome

**Target**: 5 per month

**Action**: Follow-up

refrescarlo para ver como va

**Indicator 2 New offers and proposals**

se usa para estimar futuras metas

**Type**: Driver

**Target**: 60 per month

**Action**: Publish in social-media

# Indicators

ejemplo: he vendido 5 apps: eso es una medida
metrica: he vendido 5 apps por semana

- Indicator vs Measure: indicators are obtained from metrics, which are composed of measures. Indicators often aggregates or combines multiple metrics.

- [Kaplan&Norton] There are two fundamental types of KPIs: outcomes and drivers:

mas faciles de definir

  - Outcome KPIs—sometimes known as lagging indicators—measure the output of past activity.
  - Driver KPIs—sometimes known as leading indicators or value drivers—measure activities that have a significant impact on outcome KPIs.

# Indicators: Example

- **Objective**.:Arrive at a meeting at 12.00.
- I'm in the car, what do I want to know? Which of the answers to these questions help me to act in order to fulfill the objective?

un leading para un objetivo puede ser lagging para otro !!!!!!!

  - Will I be late if I continue at the current speed? leading
  - How many kilometers do I have left to reach the destination? leading
  - How much time do I have left to reach the destination? leading
  - How many kilometers have I traveled? lagging
  - How long have I been traveling? lagging
  - What is the current fuel consumption? lagging
  - How many liters are left in the tank? leading
  - How many kilometers can I travel without refueling? leading

# Indicators: Example

- **Objective**.:Arrive at a meeting at 12.00.
- What if there are more goals? For example: Lower cost.
  - Can I throw the bags so that the car weighs less and spend less fuel to reduce cost while maintaining speed?
  - Can I bribe the cops so they don't fine me if I get up the speed?

# Indicators

- More examples:
  - **Mortality rate:**
    - Action is possible?
    - Can I “correct” it?
    - If not, It is a KRI not a KPI.
  - **Length of stay:**
    - Action is possible?
    - Can I just discharge the patients early to improve this?
    - It is possible, but other KRI such as readmissions will increase.

# Indicators, Recommendations for SME [Parmenter]

igual que KRIs y KPIs, pero sin ser KEYS

Parmenter - autor muy reconocido

- **RESULT indicators (RIs)** – are summary and retrospective indicators, that provide information about what has been done. They summarize **past** performance.
  - these can be measured daily, weekly, monthly, or quarterly.
- **PERFORMANCE indicators (PIs)** – track specific aspects of performance that are important but not critical for meeting strategic objectives. They are typically used for more focused areas of operation, providing information that can help teams improve day-to-day processes but are not as high-level or impactful as KPIs.
  - These are measured daily, weekly, or monthly and are not as important as the KPIs.
  - **Number:** 30-50 PIs and RIs.

# Indicators, Recommendations for SME [Parmenter]

10(KPIs max) - 80 (Pis + KIs maxs) - 10(KRIs max)

- **KEY RESULT indicators (KRIs)** – high-level summary indicators that provide broad insights into how well an organization or department has performed in achieving key goals. They help communicate to the board or senior leaders how effective management has been in executing the strategy, but they are not detailed enough to show how to improve performance in real-time.
  - Typically these are measured **monthly** or quarterly. **Number**: 6-10.
- **KEY PERFORMANCE indicators (KPIs)** –are the most critical indicators for driving strategic objectives. They give real-time, actionable insight into whether the organization is on track to meet its goals. dramatically.
  - These are measured either 24/7, daily, or weekly. **Number: 6-**

# Indicators

- KPIs characteristics [Parmenter] segun el pibe, no es la biblia
  - Usually non financial measures (not expressed in €, $,…)
  - Measured frequently e.g. 24 by 7, daily or weekly
  - Acted upon by the CEO and senior management team
  - All staff understand the measure and what corrective action is required
  - Responsibility can be tied down to a team
  - Significant impact e.g. it impacts on more than one of top CSFs and more than one balanced scorecard perspective
  - It is imperative that the importance and impact of these measures are well understood by all, and the corrective action required if the performance is not reaching targets.

# Indicators

- Responsibility can be tied down to a team
  - “Within 12 months, every employee will be able to check their email, calendar and their personal top 10 KPIs at the breakfast table using their smartphone or tablet.”
  - Performance management needs to become personal, simple and directly effective. We need to put business metrics in the hands of the people who do the work and drive performance and make sure that KPIs leave the ivory tower called the boardroom.

# Indicators

- Other classification depending on what you are measuring

indicadores complementarios a lo que ya hemos visto

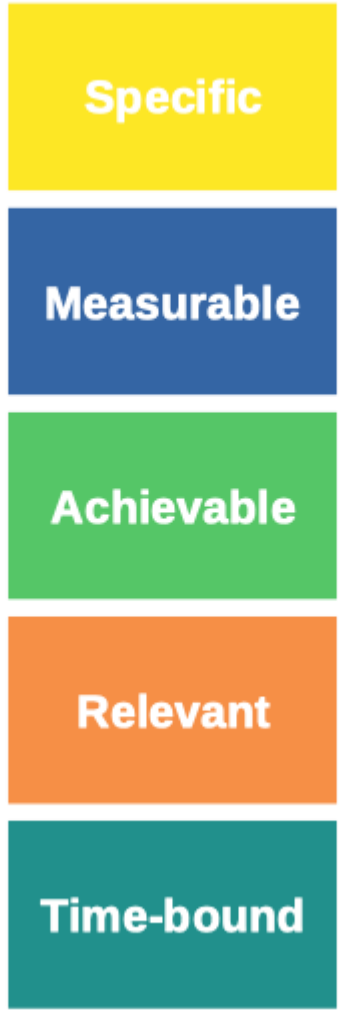
  - Risk Indicators show the risk to exceed the defined risk appetite in the future and should be able to accurately predict losses. EJEMPLO: KPI de reducir un % costes medido como número de catéteres usados incrementa el KRI de tendencia reducida en número de infecciones de bacteriemia. Una tendencia sube y la otra baja, es una mirada al futuro.
  - Control Indicators, set the desired internal control effectiveness of an organization. EJEMPLO: adherence to security standards to prevent data losses.
  - Lead Indicators are being increasingly used to measure the achievement of strategy goals (for instance, in terms of customer satisfaction).
  - DI: Diagnostic indicators help analyze and understand past performance

# Indicators

- Qualitative vs Quantitative.
- Quantitative data measures activity by counting, adding, or averaging numbers.
  - E.g.: inventory, purchasing, orders, accounting, employee injuries, number of training classes, …
  - Quantitative data forms the backbone of most KPIs.
- The most common qualitative ones gather customer or employee satisfaction through surveys.
  - While the survey data itself is quantitative, the measures are based on a subjective interpretation of a customer's or employee's opinion on various issues.
  - These opinions can help explain why performance is dropping when all other indicators seem fine.

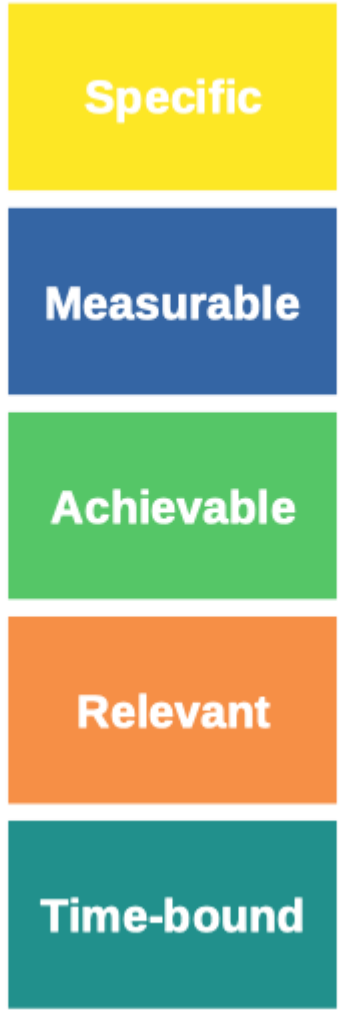
# KPI features: S.M.A.R.T. (and others)

- **Specific:** it has to be clear what the KPI exactly describes and the context within which it is defined; Target an specific area rather than a general one
  - What you want to achieve?
  - Who is involved?
  - Why is important?
- **Measurable:** Must be quantificable
  - Performance can be measured
  - How much? How many? What is the baseline? What is the timeframe?



# KPI features: S.M.A.R.T. (and others)

- **Achievable:** realistic and can be accomplished with the resources, constraints and time we have.
  - We need to assess resources
  - Take into account stakeholders, external factors and past performance
- **Relevant**, align with organization's goals.
  - It should matter to the organization.
- **Time-bound** – specify when the result(s) can be achieved
  - Measured each week? Time scale suitable.
  - Facilitates planning and tracking.



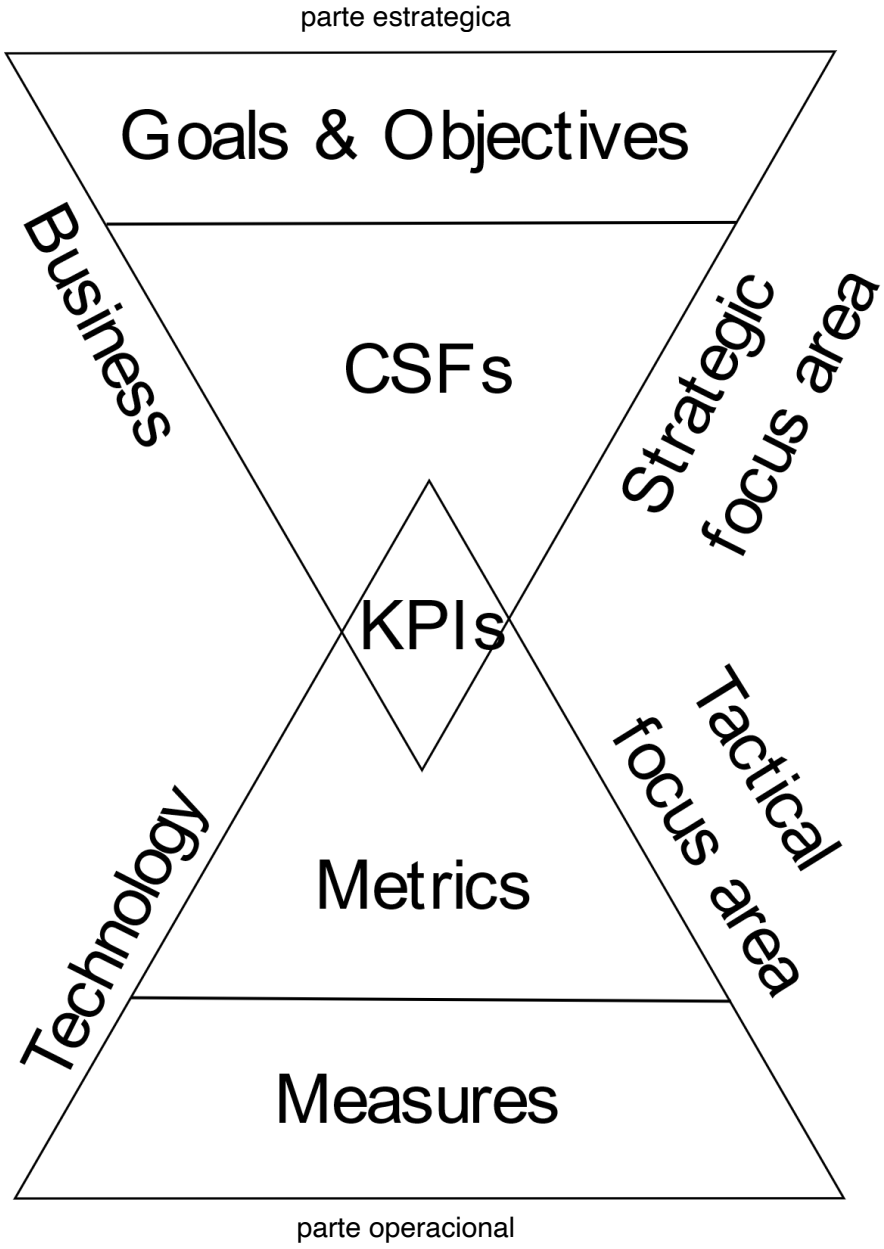
# KPI features: S.M.A.R.T. (and others)

- **Others**
  - Comparable
    - Used to compare different organization and data available
  - Verifiable
    - For an effective data collection
    - Process reapetable with robust data
  - Cost-effective
    - The cost of measuring and collecting information must be balanced with benefits

# KPIs: why we need them?

con los KPIs podemos:

- Evaluate status/progress
- Identify changes
- Show if we are meeting our strategic business objectives
- Drivers of improvement
- Save time and increase efficiency
- Improve decision making
- Support Benchmarking



# Just thinking a little bit

- **KPI in HealthCare**
  - Exitus rate
  - Admission and readmission rates
    - ICU without walls
  - Patient Wait Time: before check-in
  - Patient Satisfaction
- What others?
  - **For next week: You have to search for national/local KPI for healthcare system or other field.**
  - **Eg. Vaccination rate.**


traer 3 KPIs apuntados para la semana que viene

fallos tipicos de los KPIs: fallos de comunicacion, de trabajo en equipo, de compromiso, …

# Just thinking a little bit

- KPI for SDG
  - Example: SDG#6-Water and Sanitation
    - Target 6.5: By 2030, implement integrated water resources management at all levels, including through transboundary cooperation as appropriate
    - Indicators:
      - 6.5.1- Degree of integrated water resources management implementation (0-100)
      - 6.5.2- Proportion of transboundary basin area with an operational arrangement for water cooperation
  - SDG#1- No Poverty
    - Target 1.2: By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions
      - Ind 1.2.1- Proportion of population living below the national poverty line, by sex and age
      - Ind 1.2.2- Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions

# Just thinking a little bit

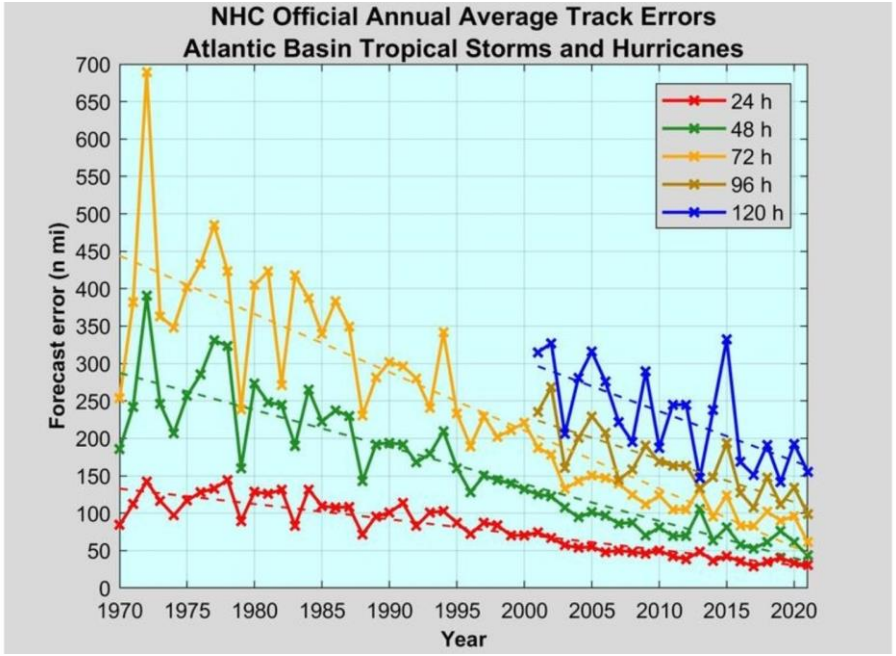
## KPI - Key Performance Indicators, global DU

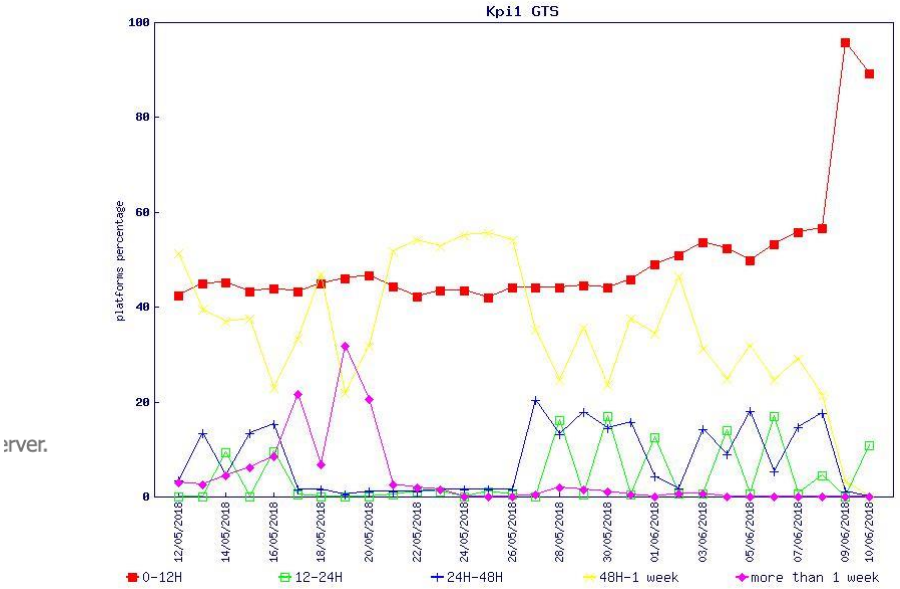
The Key Performance Indicators of the global in-situ TAC are daily calculated.
They provide indicators on avalability and quality of in-situ data distributed by CMEMS.

### KPI-1 : data availability

KPI-1 monitors the delay between the last date of observation and the first date of availability on CMEMS ftp server.

- Drifting buoys
- Gliders
- GTS
- Moorings
- Profiling floats
- Vessels




**NOAA Annual Average Track Errors for Tropical Cyclones in the Tropical Atlantic**