

# Business intelligence

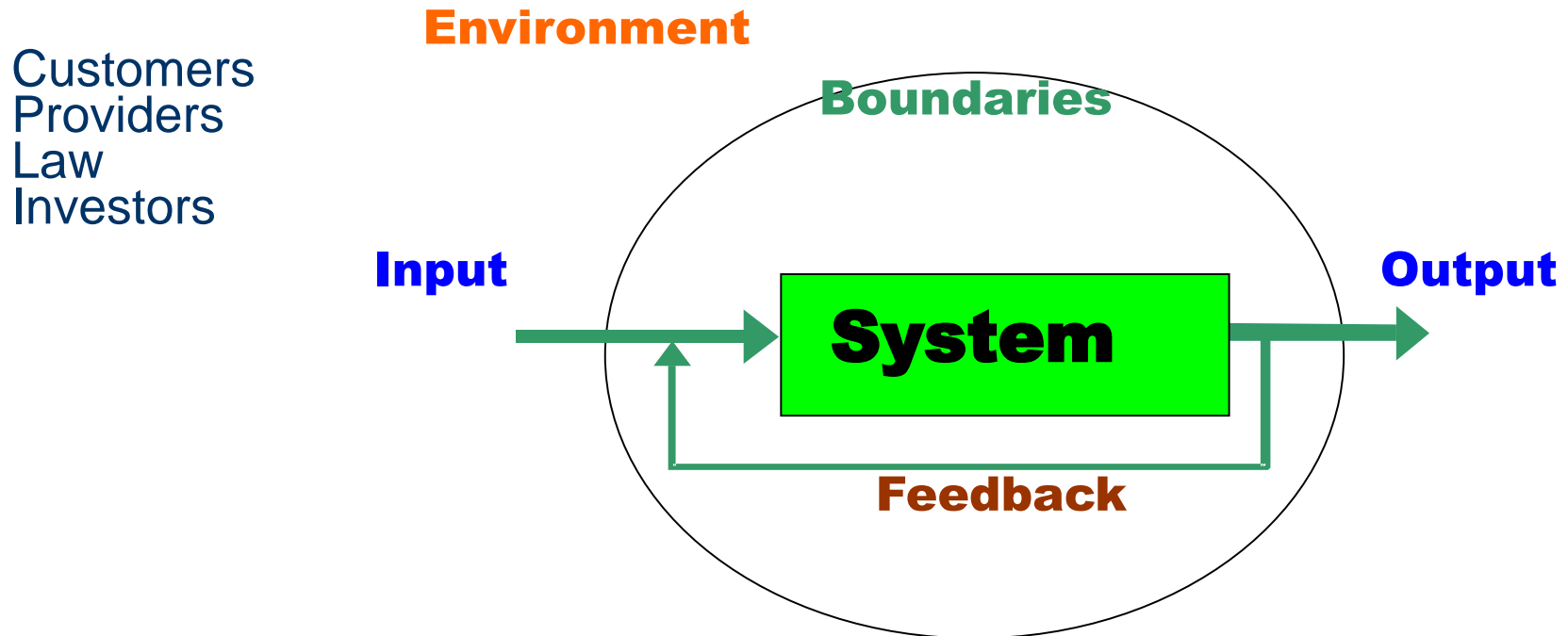
## Unit 1 – Introduction to BI and methodology

- **System:**

- System is a set of interrelated things that contribute to a certain goal.
- A system consists of a set of interrelated elements, operating in a changing environment and with specific objectives.

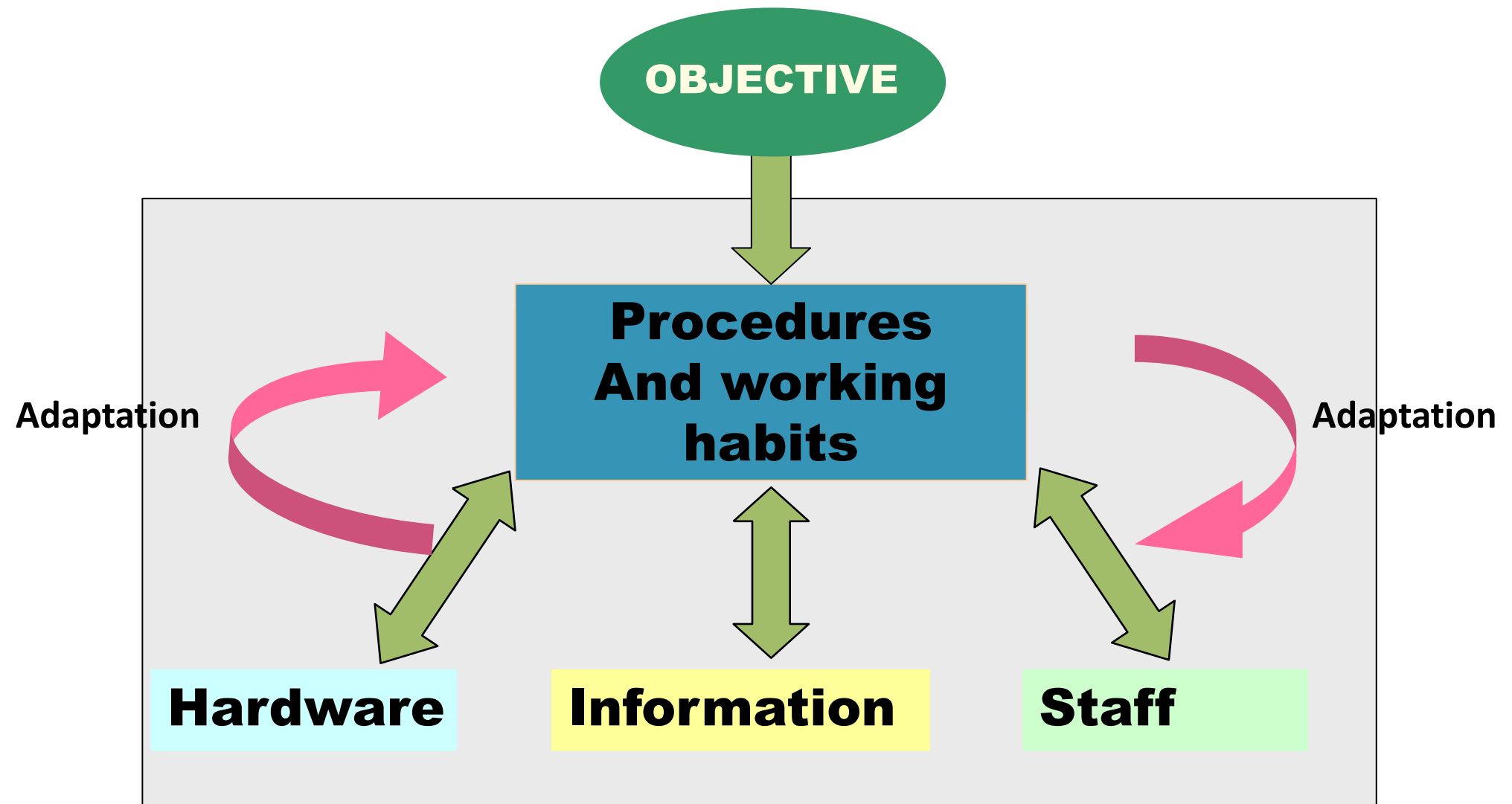
- **Basic elements:**

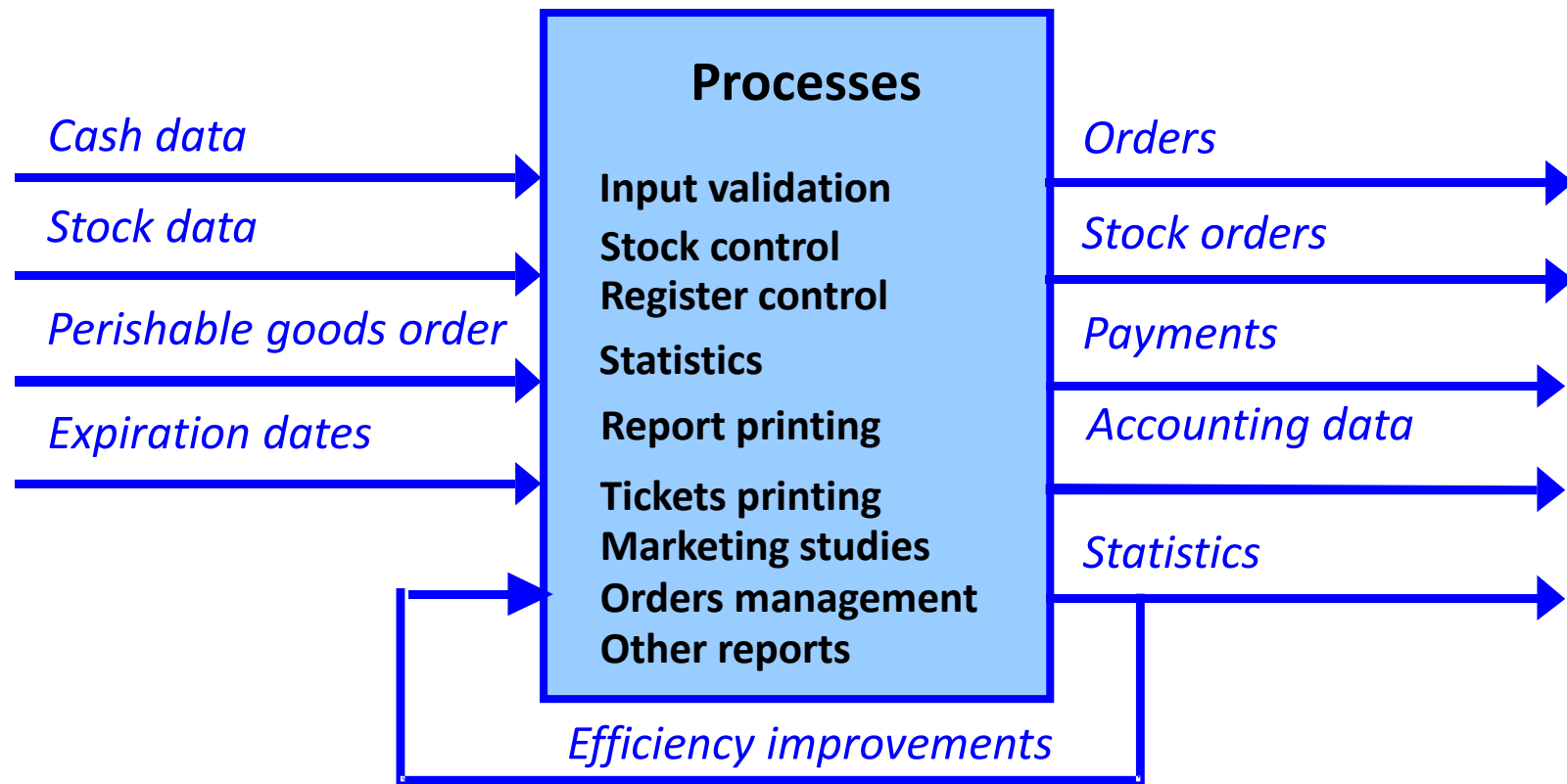
- System components.
- The relations between them, which determine the structure of the system.
- The objective of the system.

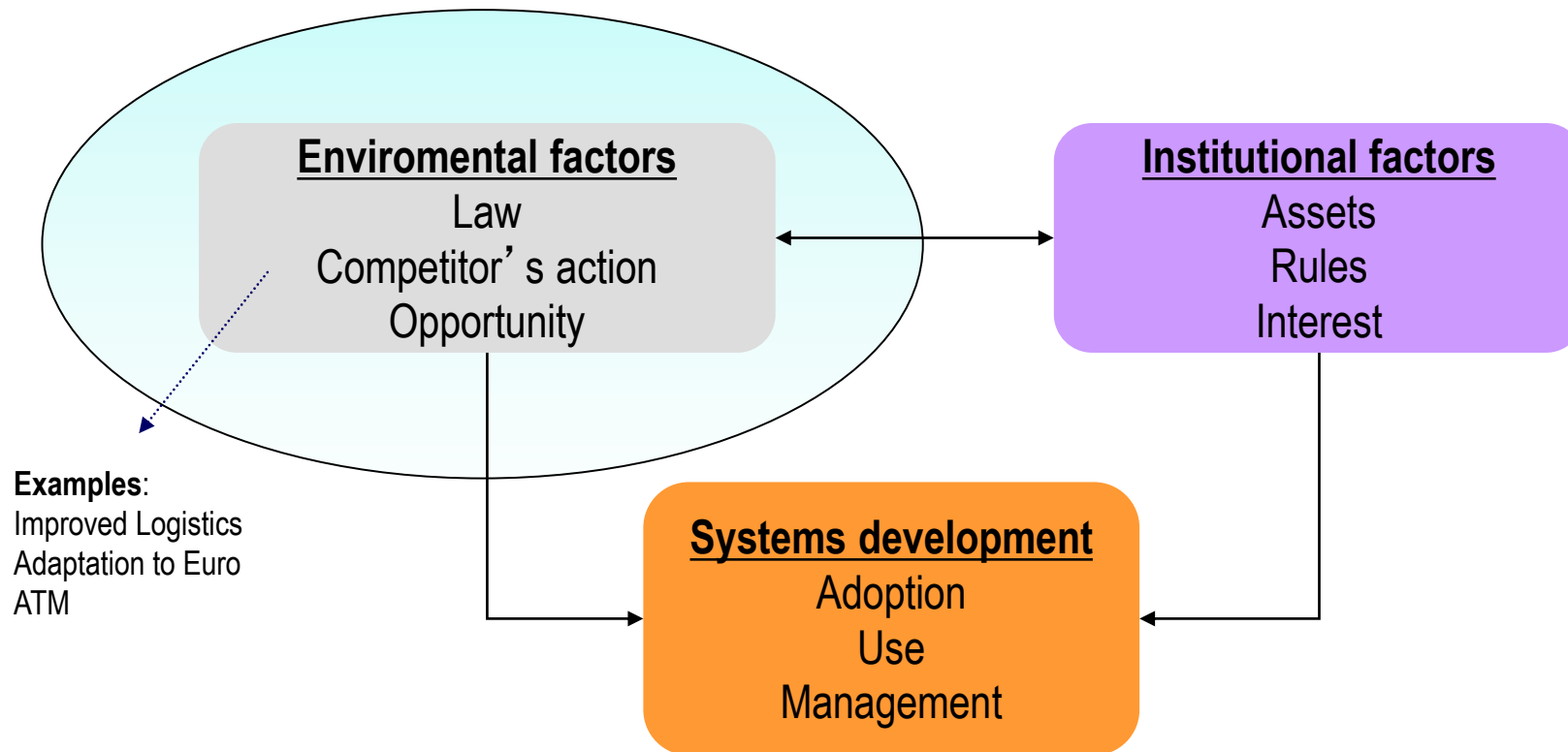


- The system **environment**: his surroundings, within which it is located.
- The system **boundaries**: the boundary between what the system is and what constitutes the environment.
- Feedback**: In many systems the output influences the input of the system.

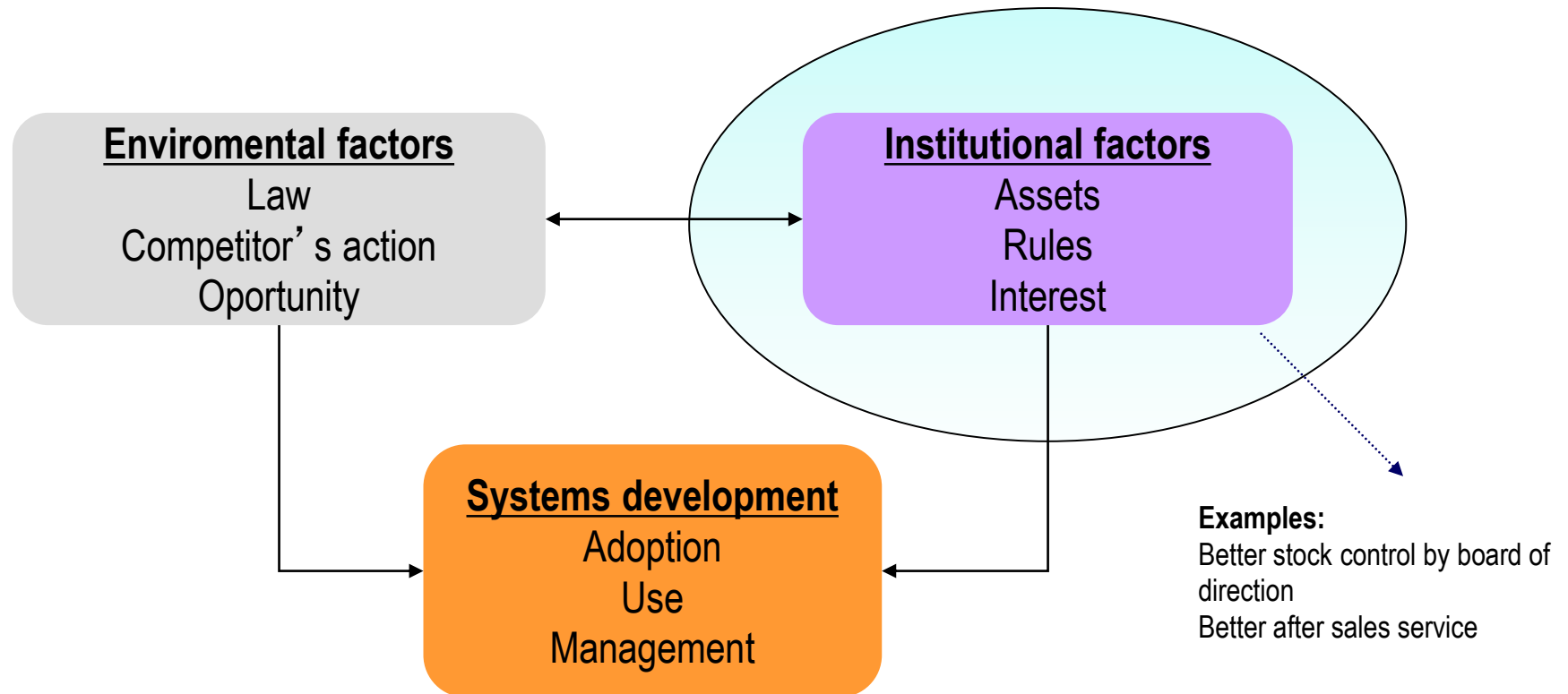
- Definition:
  - A formal set of processes that, operating on a collection of data structured according to the needs of the company, collect, process and distribute the information (or any part of it) necessary for the daily operations of the company and for the direction and control (decisions) to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





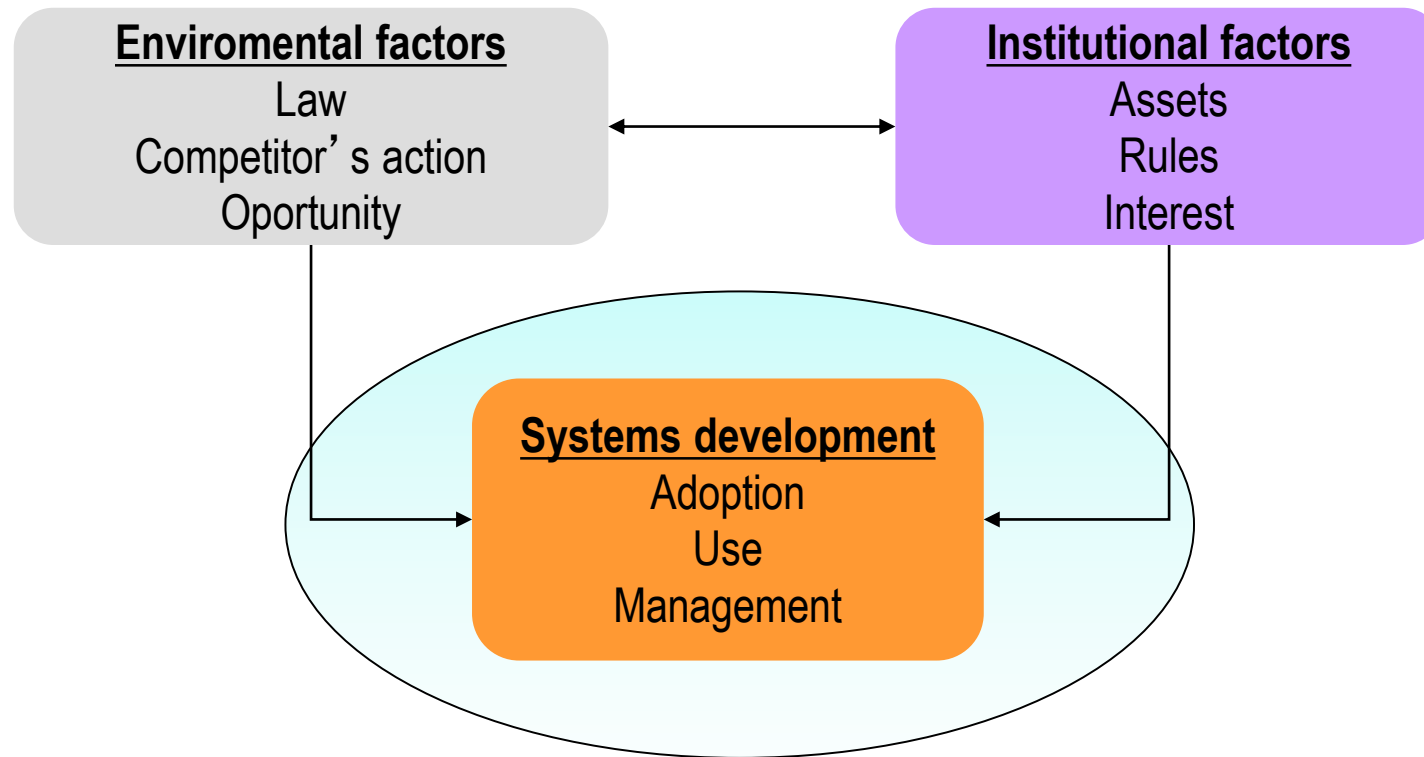


- From a business perspective, an information system is a solution for business organization and management based on information technology whose aim is to deal with an emerging challenge in the business context

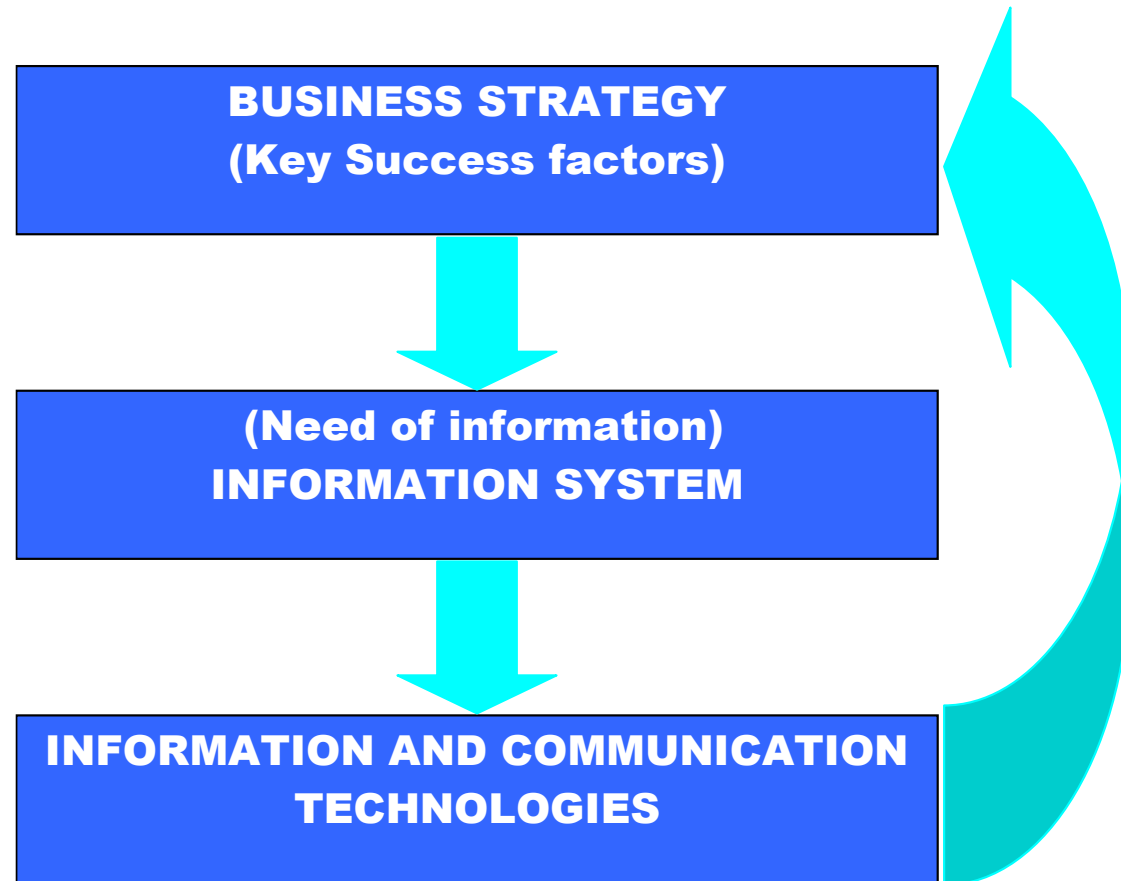


- Strategic information systems: at any organizational level there are changes in goals, operations, products, procedures or relationships with the environment to gain a competitive advantage





- 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..



- “"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

Period	Information	Information system	Objective
1950-1960	Necessary bureaucratic requirement	IS for Accounting	Speed in accounting and document processing
1960-1970	General purpose information	IS for administration	Speed to the general requirements for information to be presented, reports
1970-1980	Administrative control to user needs	Decision support systems (DSS) Management support	Improve and adapt specific decisions.
1985-2000	Strategic Resource Competitive Advantage Strategic Weapon	Strategic systems	Promote the survival and prosperity of the organization

- Systems **operational level**: They monitor the activities, operations and basic transactions of the organization.
- Systems **decision support**: They support the monitoring, control and decision-making and administrative activities of the directive board.
- **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.

- **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, ...**

- **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 semistructured and structured 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.

- Such systems may include production scheduling, stock orders, change of suppliers, simulation models, inventory models, etc.
- Examples: geographic sales analysis, temporal sales analysis, production management, cost analysis, cost/utilities analysis, dashboards, ...
  - Goods purchase system indicating when to do an order to the supplier.
  - Simulation system to support the decision to introduce a new product to market.
  - Financial planning system requires little input data, generates little information as a result, but can perform many calculations during processing.



- **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.
  - Their 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 Systems are creators of entry barriers to the business.
    - For example, ATM or Internet banking
  - They used to be developed ad hoc within the organization.

## ◆ Examples:

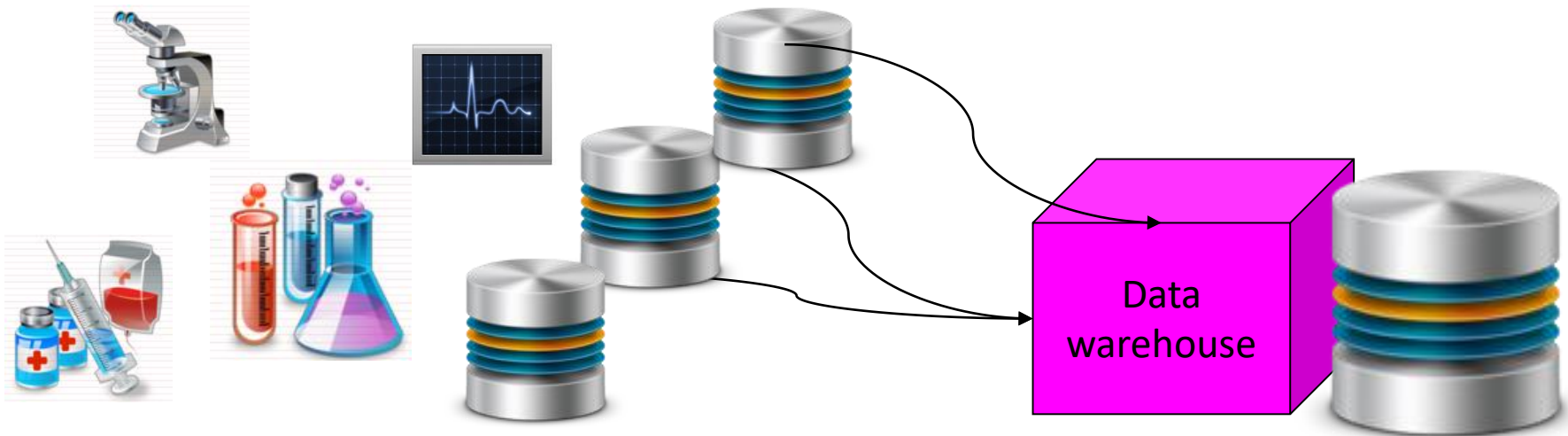
- Sales forecasts
  - Planning marketing campaign
- MRP (Manufacturing Planning Resource) focused on increasing productivity in a manufacturing process
- Product discovery and launching in banking: types of mortgages, types of accounts, ... with the purpose of achieving business goals:
  - Attracting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

System	Input	Processes	Output	User
Operational	Transactions,	Store, report, merge,	Detailed reports,	Operative staff, supervisor
DSS, tactical	Summarized operational information; high volumen of data; Simple models	Interactive; simulations; analysis	Summarized reports, Models, graphics	Technique staff
Strategic systems	Agregated information; External information;	Graphics; simulations; interactive (or not)	Predictions; proposals, patterns	Directive

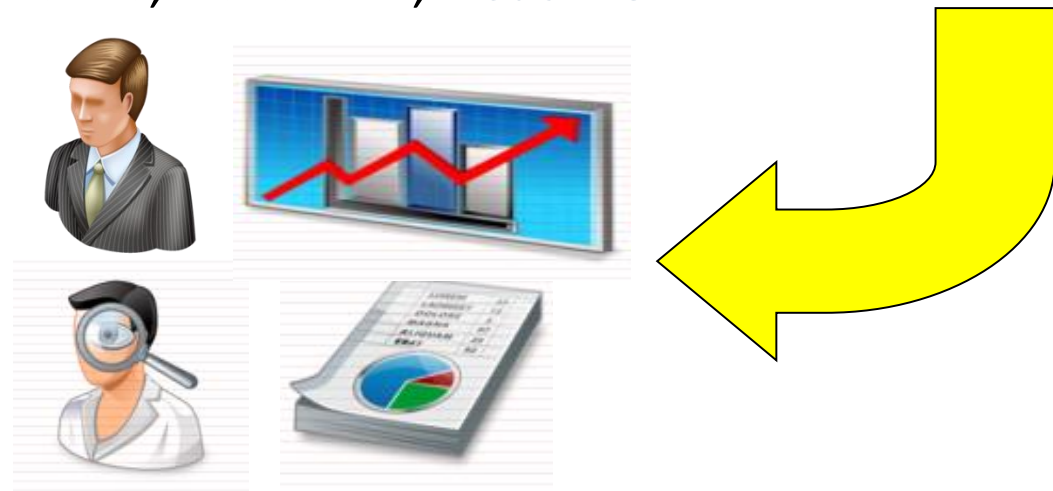
- This term refers to the management of information in a **specific** business or business area.
- It consists of a set of strategies and tools focused on knowledge management by **analyzing existing data** in an organization or company.
- In the business intelligence we focus on:
  - Setting business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



- Operative IS: gather and organize data



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



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



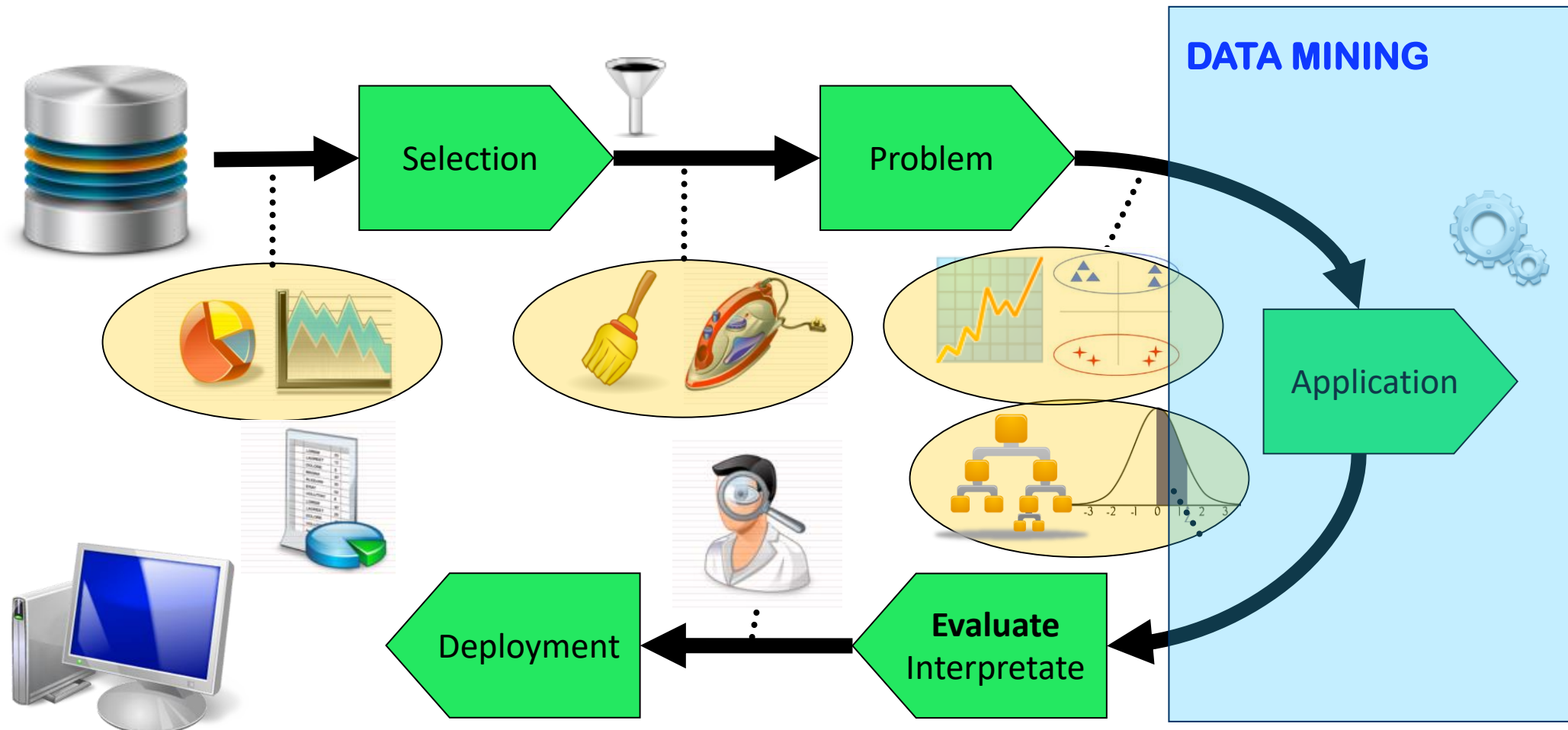
Knowledge acquisition

Knowledge discovery

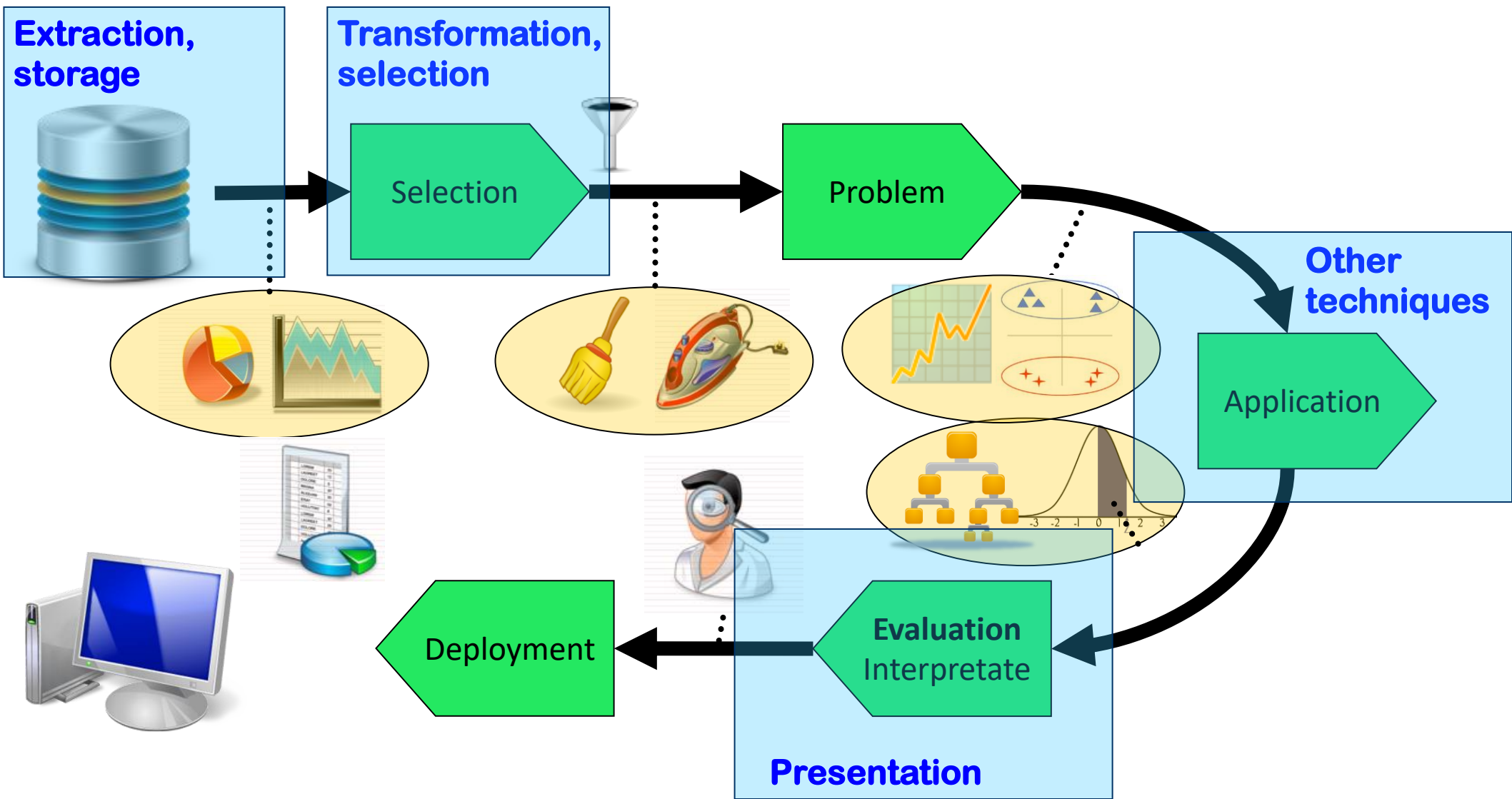


Expert system, dss, knowledge management, data mining, ...

- In the other courses you will focus on:



- Now we will focus on the complete process:





- 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
    - ...
- Each model has at least 5 stages of maturity: complex path
  - 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.

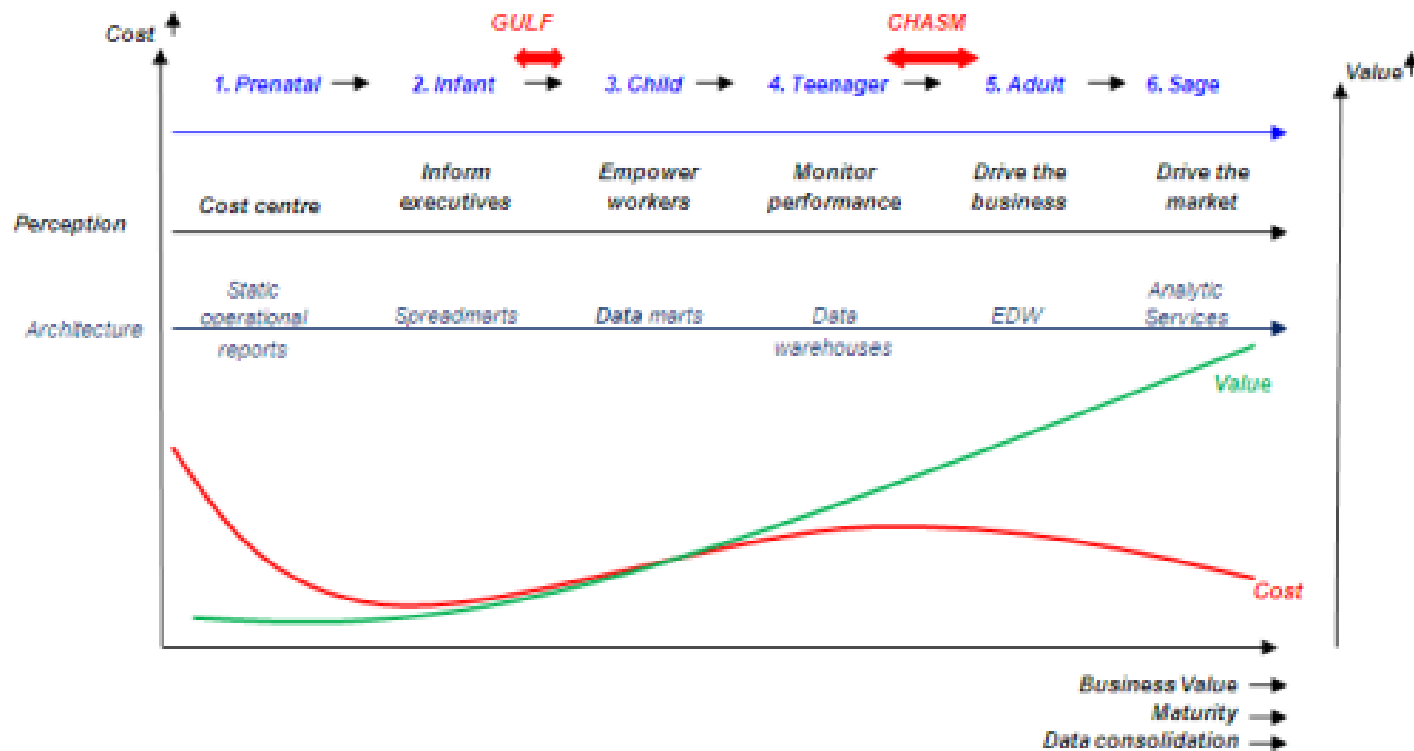
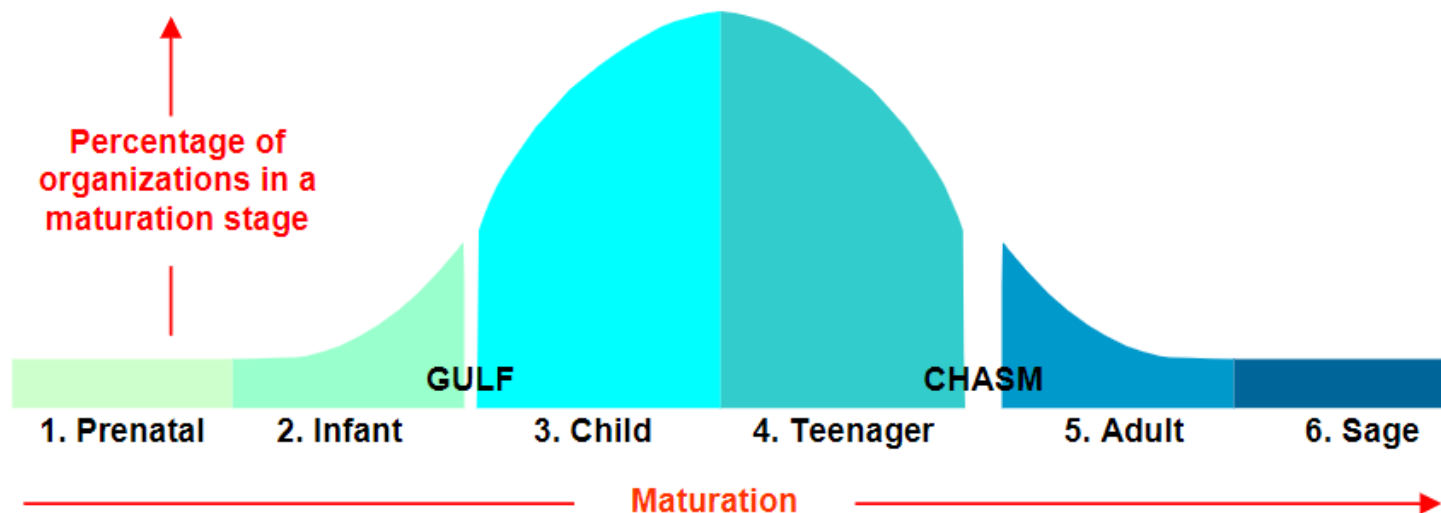
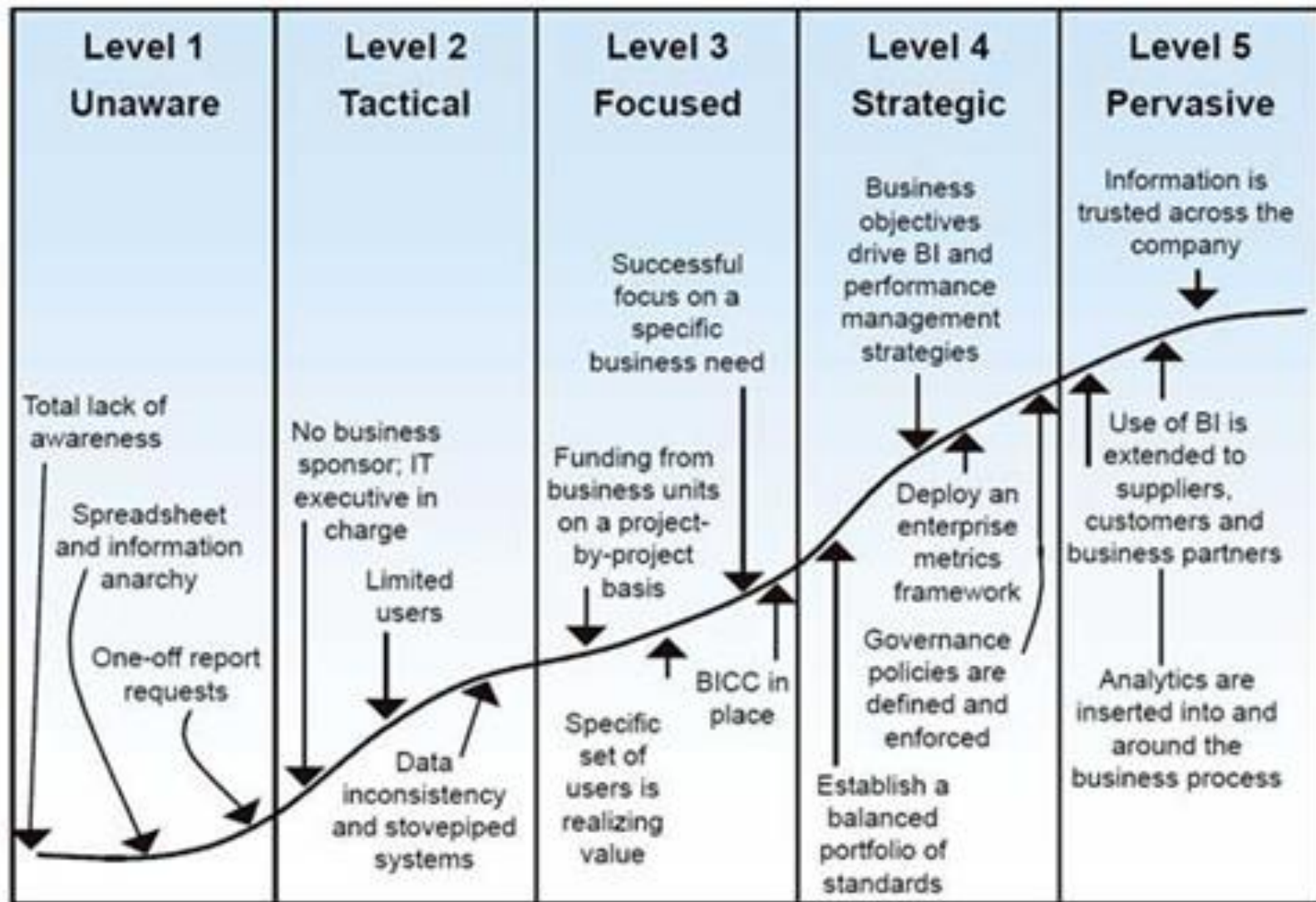


Fig. TDWI BI Maturity Model



- Stage 1: Prenatal – Executive perception is that of a cost-center, which primarily churns out static reports for management operational reporting. It is also the stage which costs the most.
- Stage 2: Infant – The BI function's role is to inform executives, with several reports leading to "spreadmarts".
- Stage 3: Child – The BI function's role is perceived to empower workers, and this is the first evolution into an analytical system where OLAP and ad-hoc reports are used off data marts.
- Stage 4: Teenager – The BI function has evolved into a performance monitoring system by now, using Dashboards and Scorecards, supported by data warehouses.
- Stage 5: Adult – This is where the ROI from the BI function shoots up, with predictive analytics answering what-if questions making the BI a strategic utility. The TDWI thinks that organizations' BI architecture has evolved to have enterprise DW by now, with BI becoming a 'Drive the Business' function.
- Stage 6: Sage – The BI function at this stage has the highest ROI and decreasing costs based off Analytic Services (SOA) with pervasive BI (e.g. embedded BI) making it 'Drive the market'.

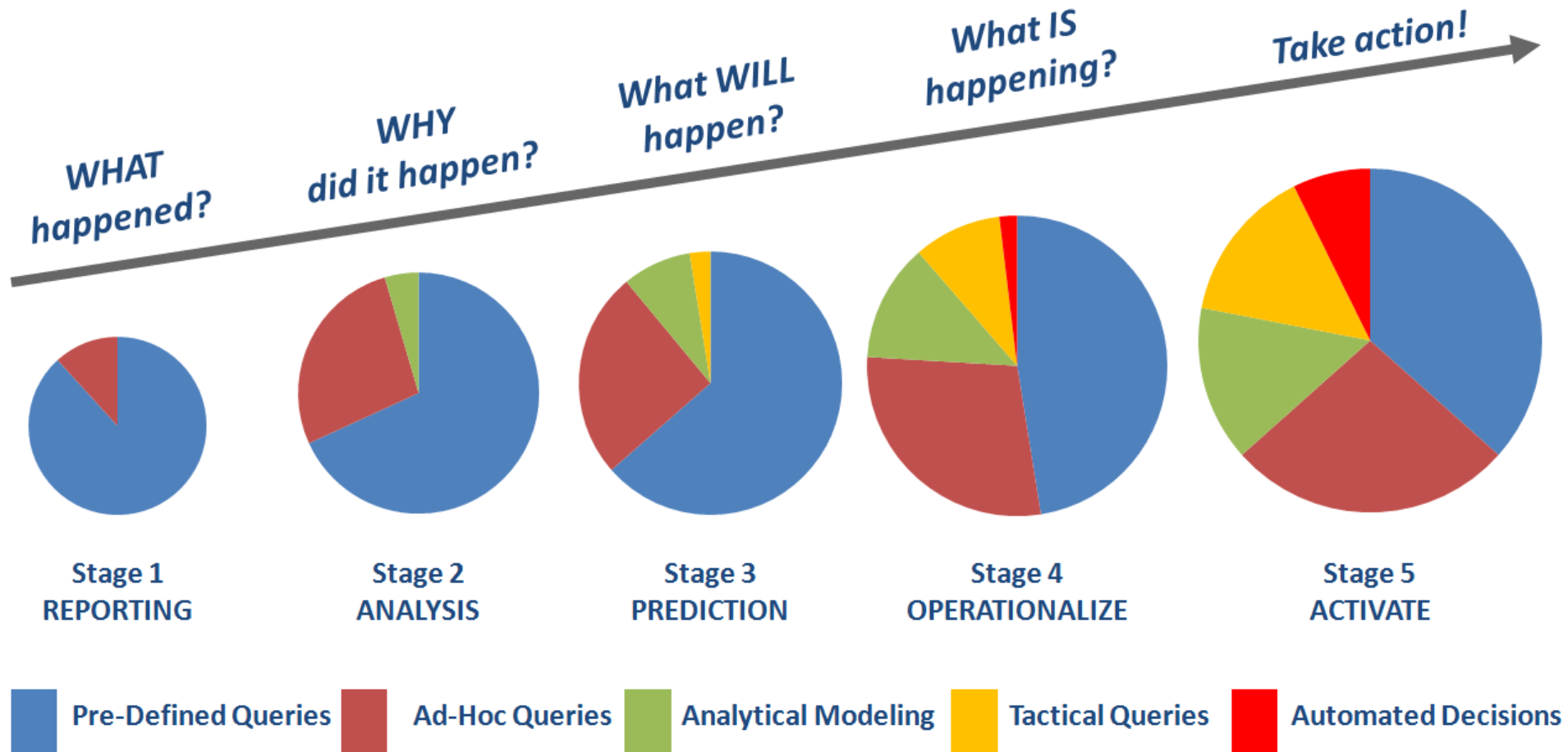
- Evolution of Business enablement, Information technology, and strategy and program management.
- Stage 1 – Operation (Running the business) – involves ad-hoc solutions focused at project activities alone.
- Stage 2 – Improvement (Measuring and monitoring the business)– involved localized solutions with project management.
- Stage 3 – Alignment – includes shared resources with program management and governance integrating performance management and BI programs.
- Stage 4 – Empowerment – includes enterprise operationalization with portfolio management focusing on organization innovation and people productivity through knowledge management.
- Stage 5 – Transformation (Change the business) – involves enterprise services tracked by service management creating strategic agility and differentiation.



BI = business intelligence

BICC = BI competency center

Source: Gartner (December 2008)



- **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 concrete staff

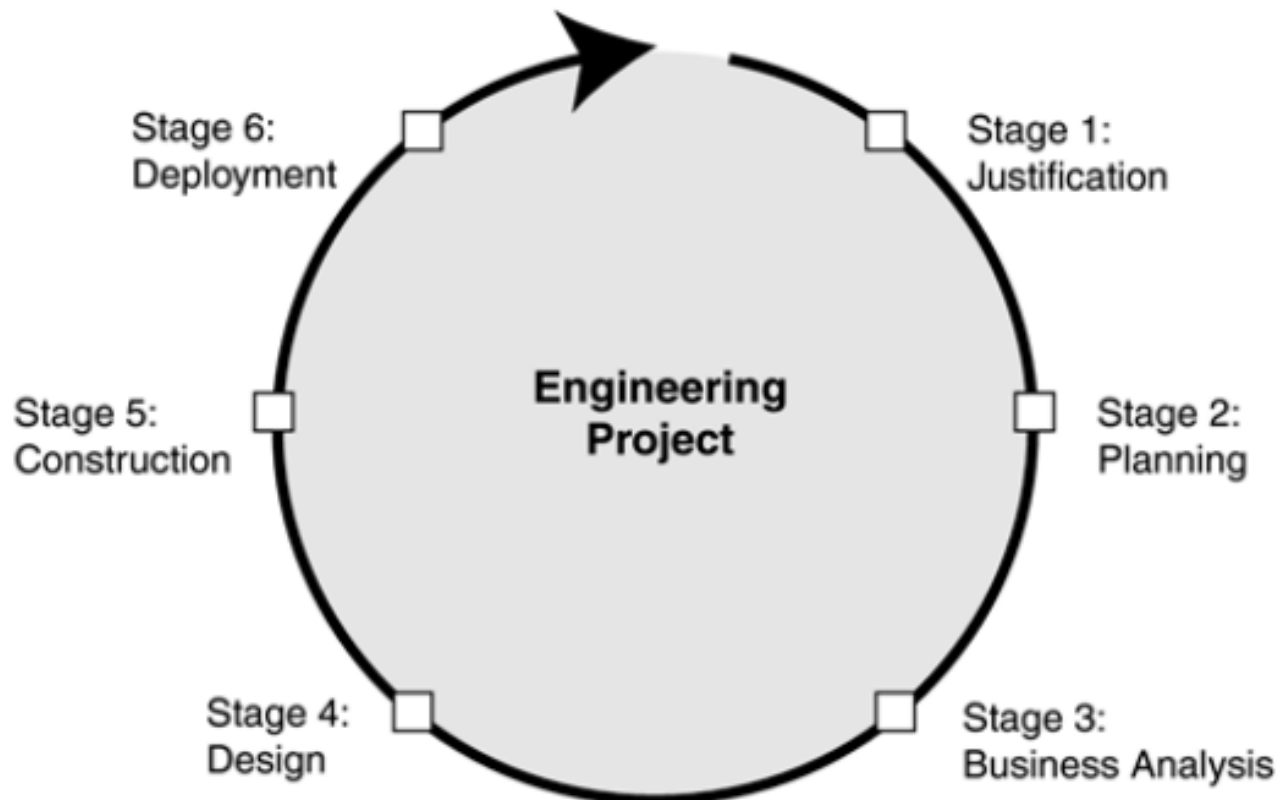
- **Focus on management and development**
  - Management: PMBook, Agile, ...
  - Management+Development: Larissa, Kimball, Inmon, SAFE, QPM (QlikView), ASAP (SAP), ... other tools
- **CRISP-DM: Generic tasks for data mining**
  - Cross Industry Standard Process for DM
  - Proposed by a consortium (SPSS,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



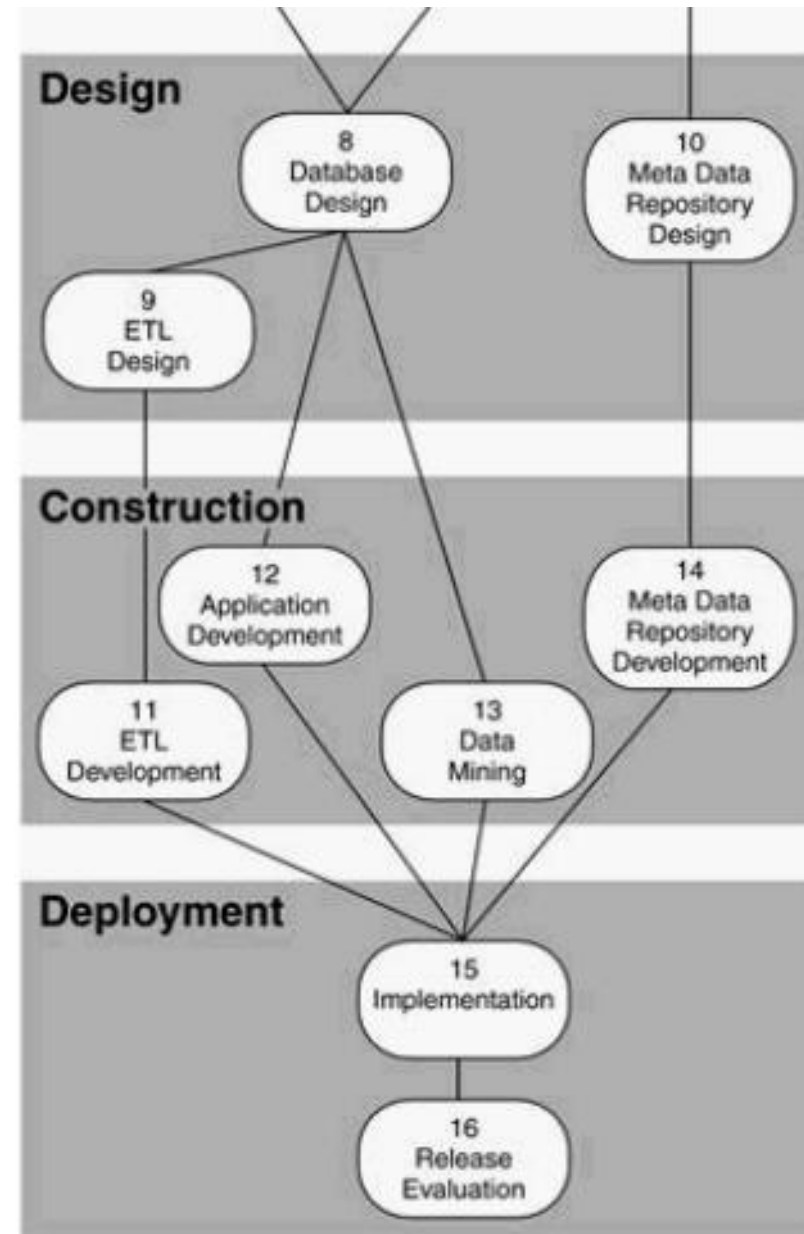
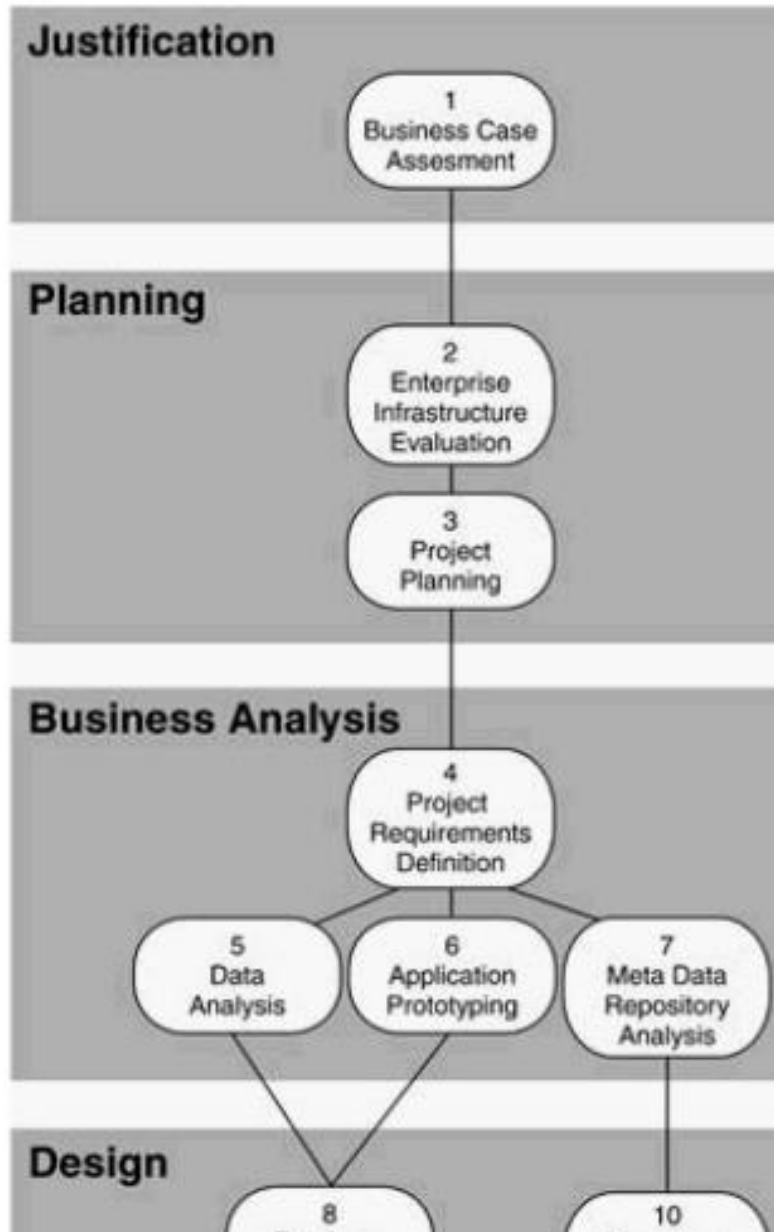
# Larissa and Moss: Roadmap to BI with 6 stages and 16 steps

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

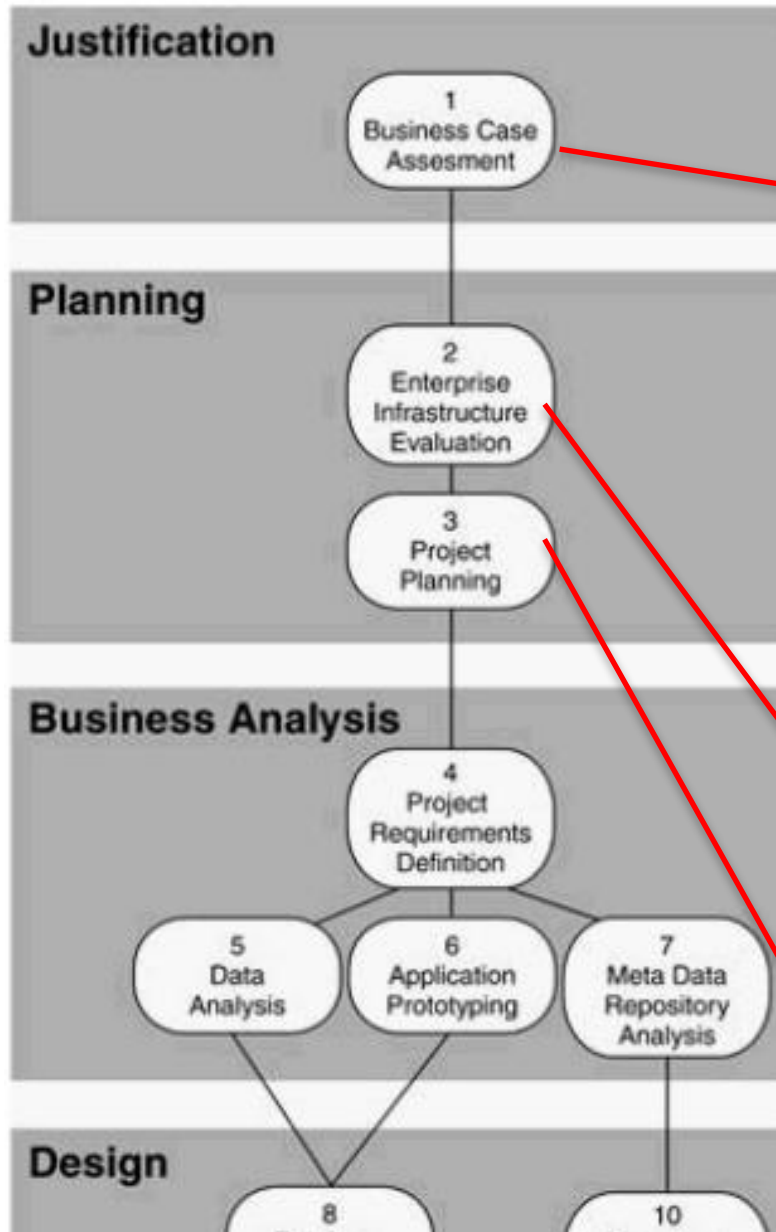
**Figure 0.1. Engineering Stages**



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

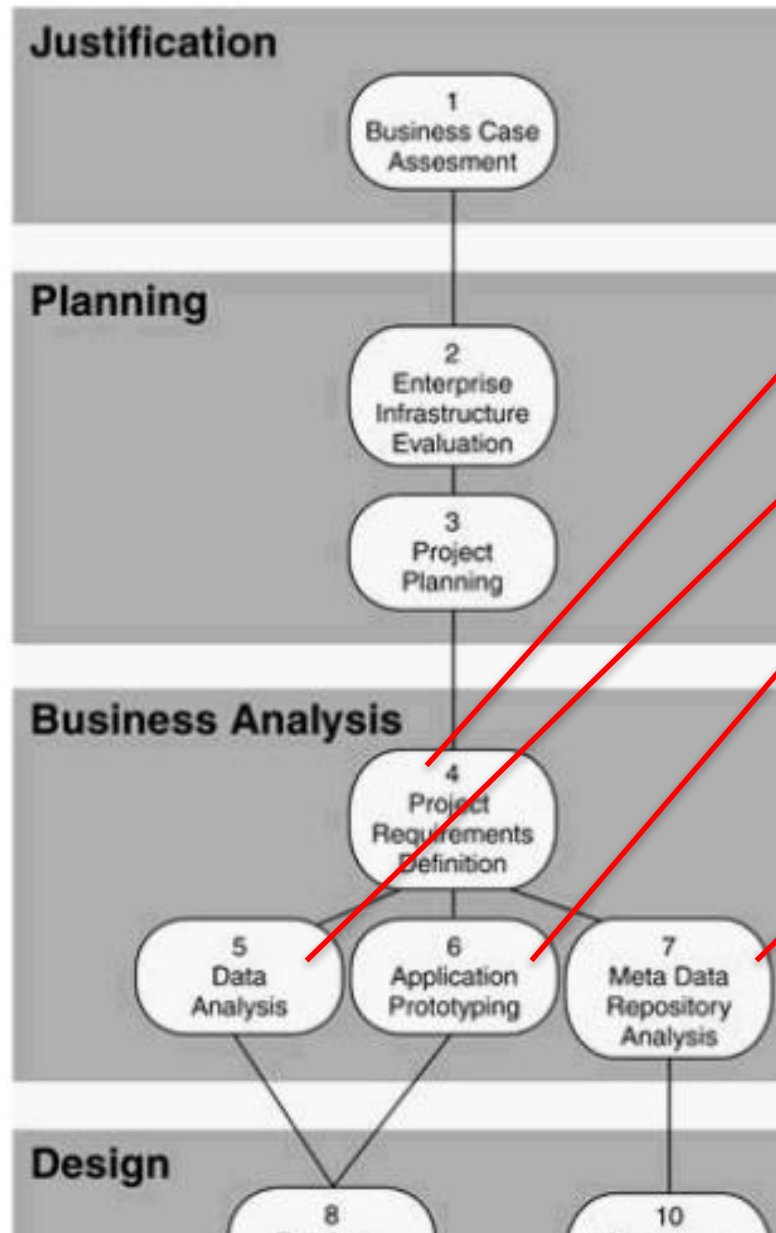


# Larissa and Moss: Roadmap to BI with 6 stages and 16 steps



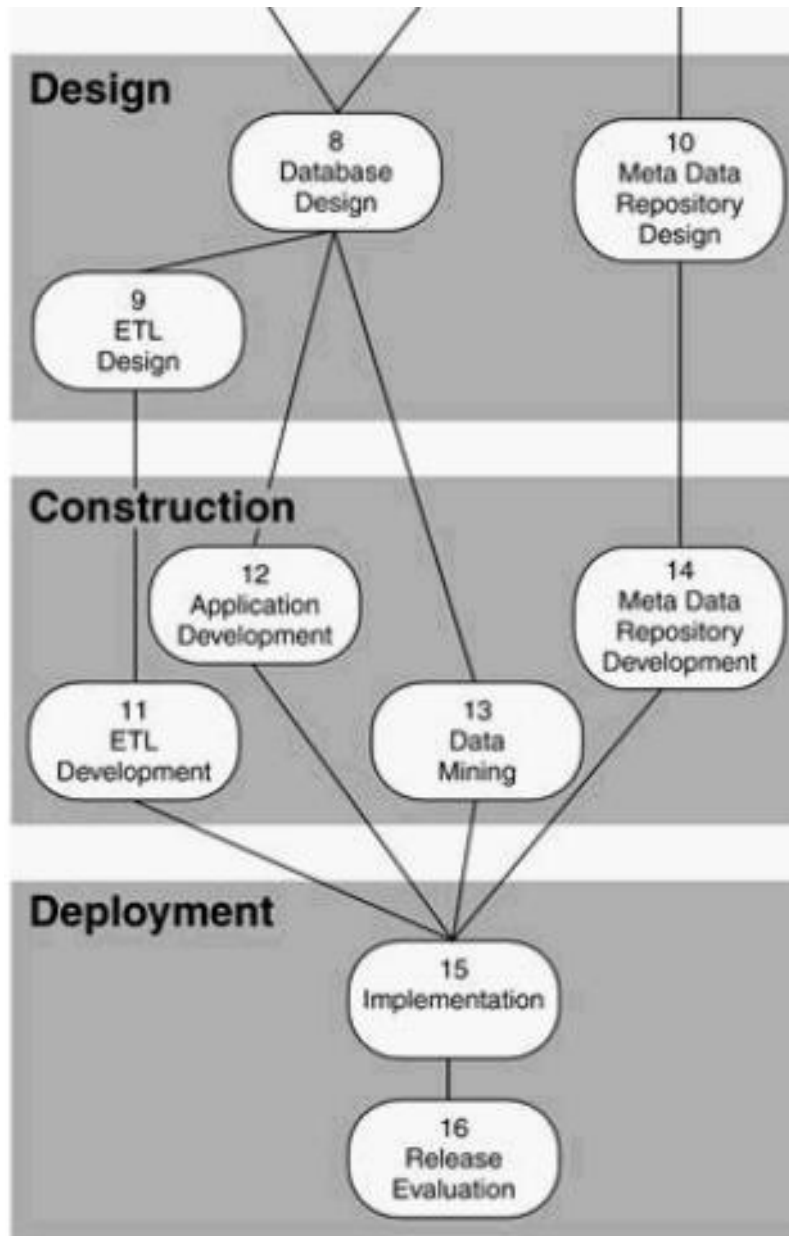
- **1. Justification:** Assess the business need that gives rise to the new engineering project.
- Step 1: Business Case Assessment
  - Defines business problem or business opportunity and proposes a BI solution.
  - Each BI application release should be cost-justified and should clearly define the benefits.
- **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 and 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 Requirements Definition
  - Managing and specifying scope, user needs, ...
- Step 5: Data analysis, availability, 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 and 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.

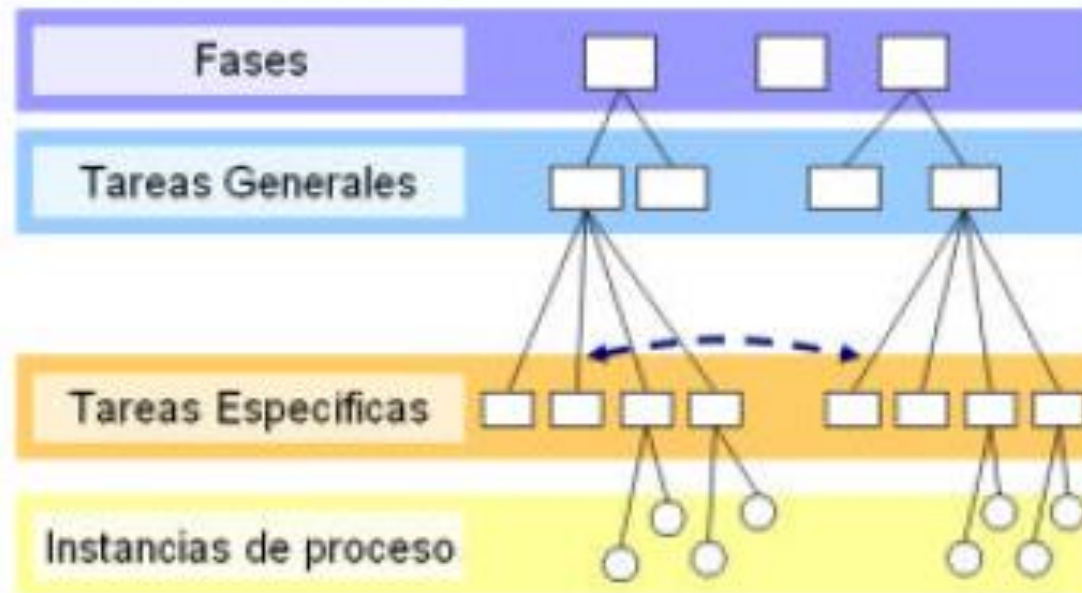
- Process without owner
- Application-independent and
- Context-independent
- Tool-independent
- It is a guide that consider the problem and techniques
- Based on experience (developed in several workshops)

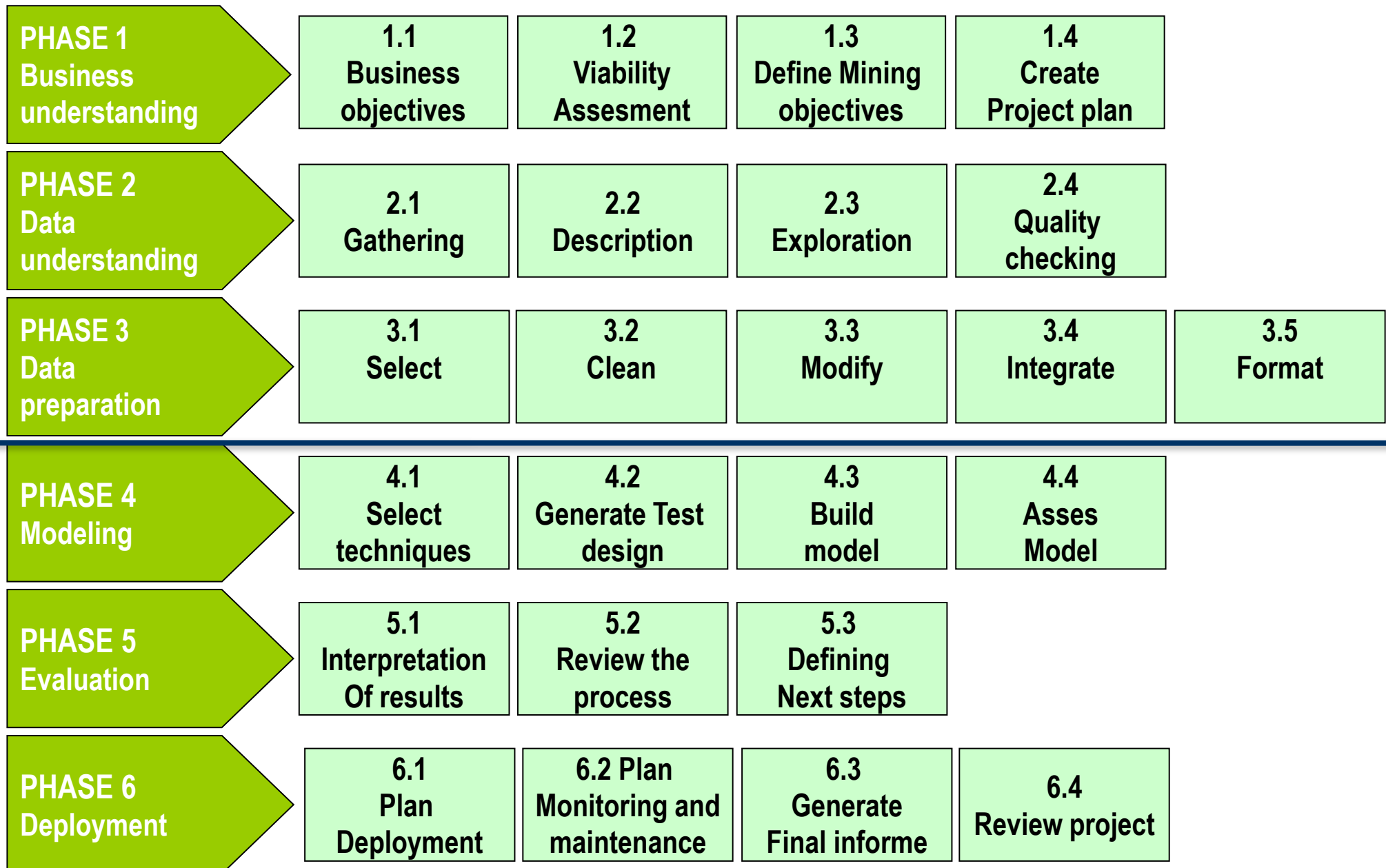


- 2 documents:
  - Reference model: describes phases, tasks and outputs.
  - User guide: practical application tips, check list



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







**FASE 1**  
**Business**  
**understanding**

**1.1**  
**Business**  
**objectives**

**1.2**  
**Viability**  
**Assesment**

**1.3**  
**Define Mining**  
**objectives**

**1.4**  
**Create**  
**Project plan**

## •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

- 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

**FASE 1**  
**Business**  
**understanding**

**1.1**  
**Business**  
**objectives**

**1.2**  
**Viability**  
**Assesment**

**1.3**  
**Define Mining**  
**objectives**

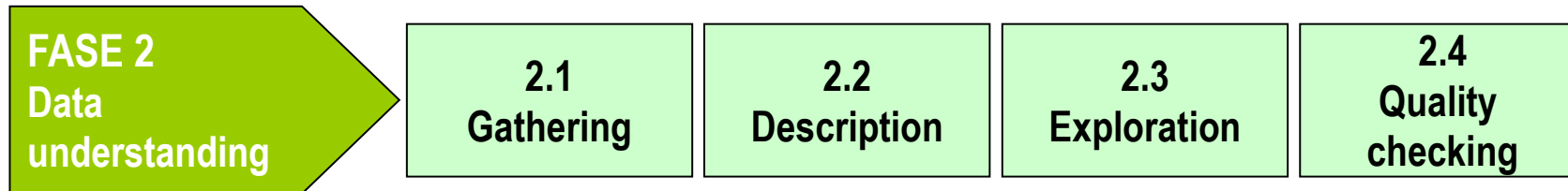
**1.4**  
**Create**  
**Project plan**

- **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



## •2.1 Gather

- Define sources, integration, localization, problems and solutions, ...

## •2.2 Describe

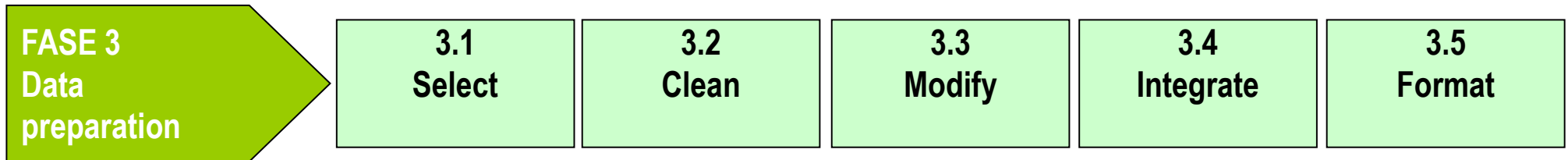
- Formats, keys, amount...

## •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?



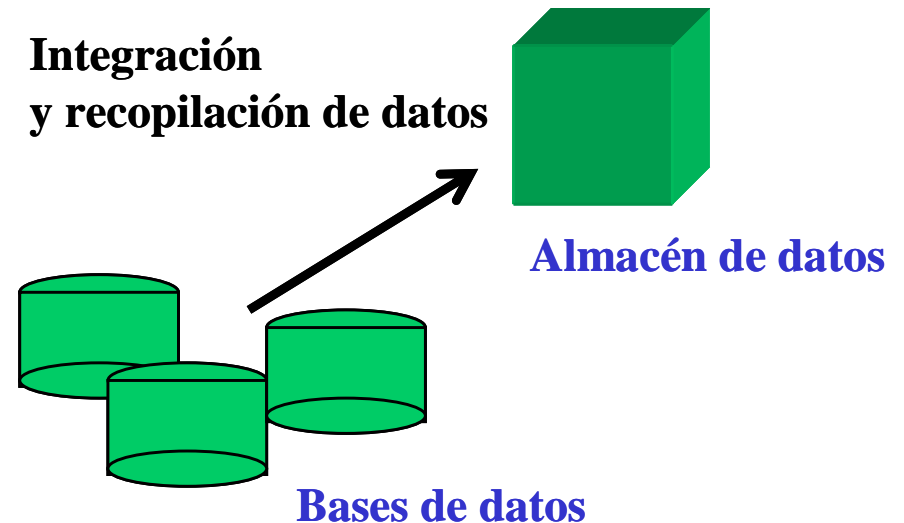
- It involves most of the total time

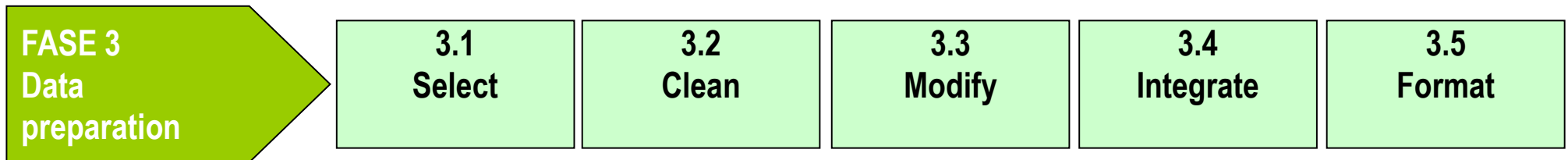
- **3.1 Selecting Data**

- Selecting tables, attributes and rows
- Volumen suitable for tools
- selection:
  - Partitioning the data set:
    - Training data
    - Test Data
    - Validation Data
  - Sampling Data

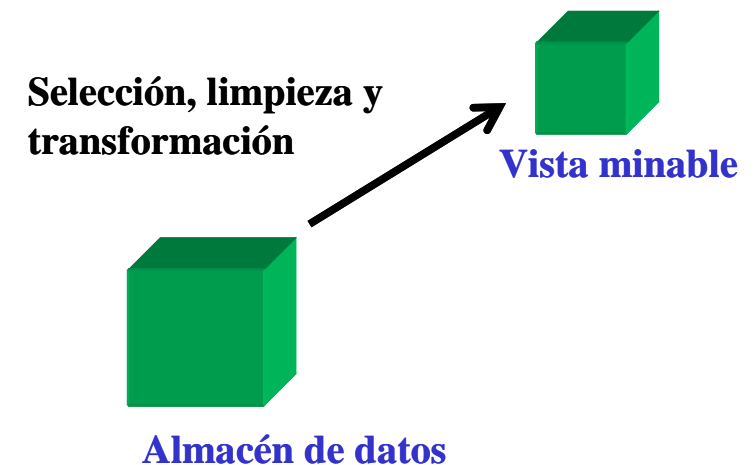
- **3.2 Cleaning: incomplete and erroneous data records:**

- Improving data quality
- Select subset of data, replace null
- Complete, delete or ignore





- 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
  - Frequency of execution
  - It involves:
    - Execution monitoring
    - Recording
    - Exceptions and errors



## FASE 4 Modeling

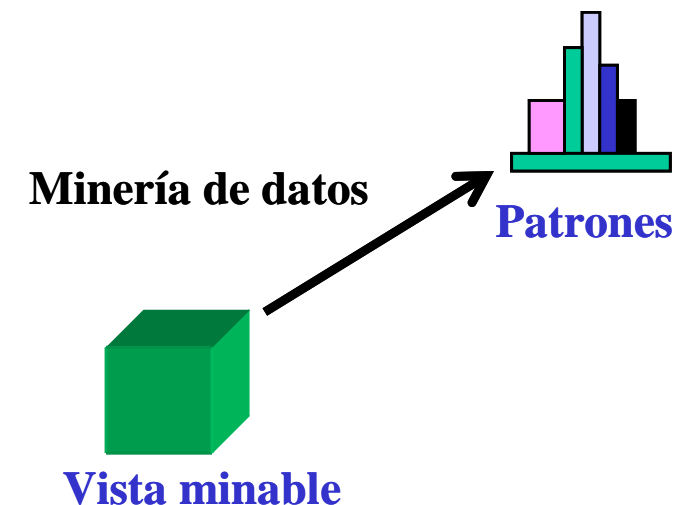
4.1  
Select  
techniques

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



## FASE 4 Modeling

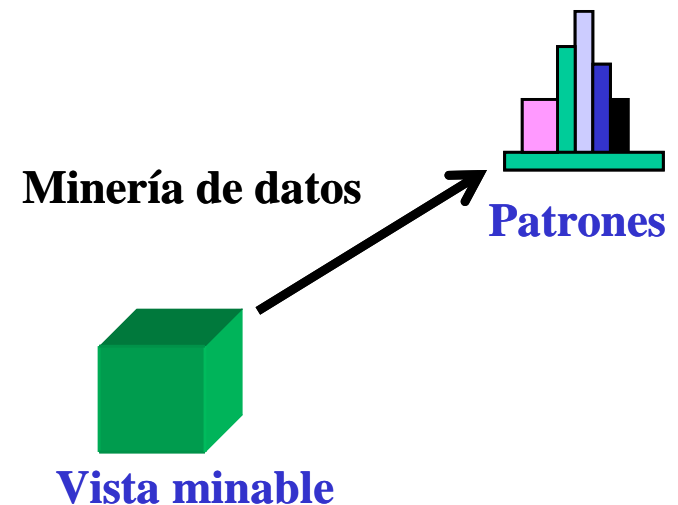
**4.1**  
**Select**  
**techniques**

**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?



## FASE 5 Evaluation

### 5.1 Interpretation Of results

### 5.2 Review the process

### 5.3 Defining Next steps

#### • 5.1 Interpretate the results

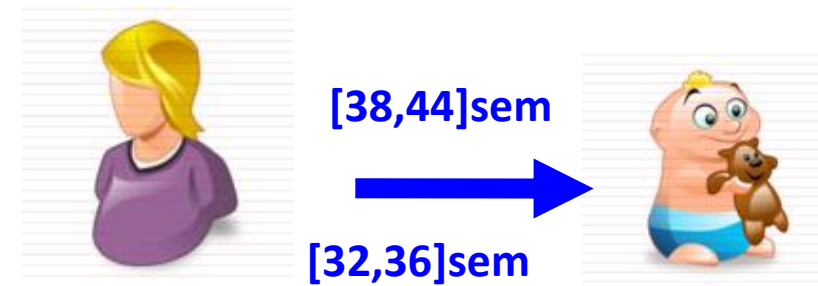
- Is the problem solved?
- Is the answer appropriate?
- Is it valid?
- Is the business objective well defined?
- Is the knowledge new?
- Is the model useful?
- Is it better than we had?
- Are there too many patterns?

#### • 5.2 Review the process

- Any error from the technical point of view?
- Have we overlooked anything?

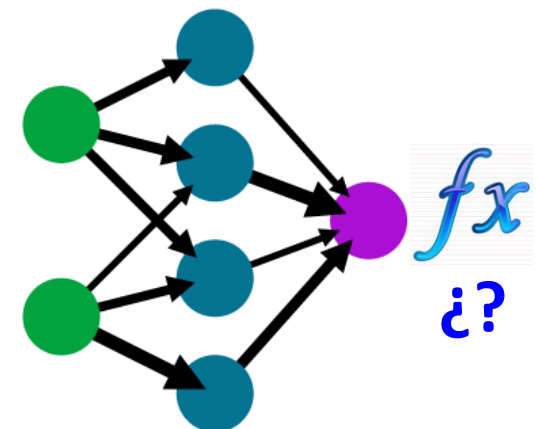
#### • 5.3 Define next steps

- Iterate? Deploy? Rebuild?

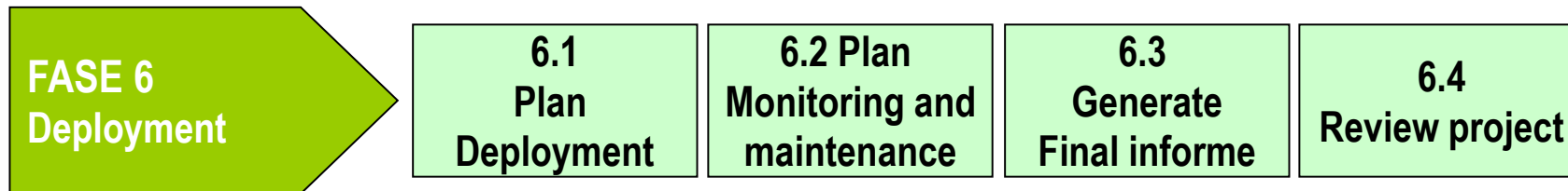


(hair=blonde)  
(profession=actress)

carrier=no







## •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?

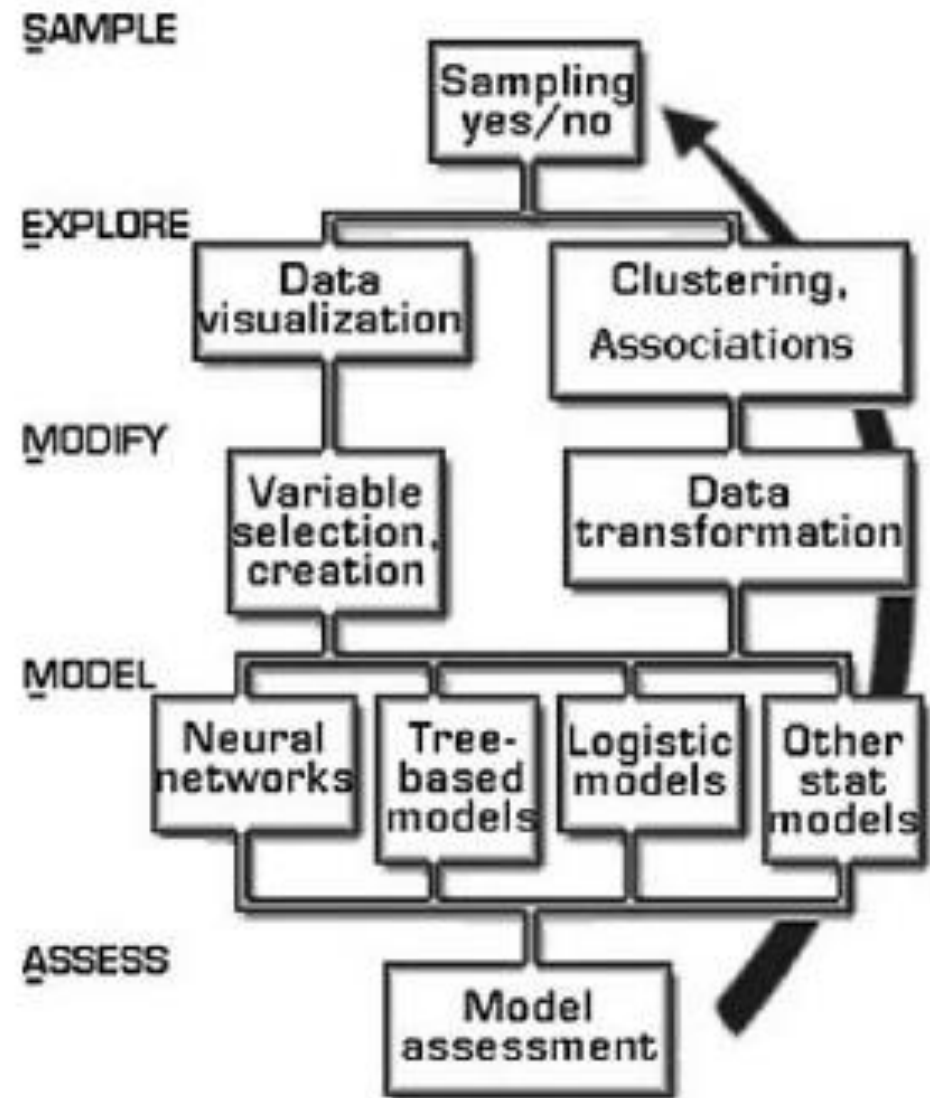
## •6.3 Generate final report

## •6.4 Review the project

- Strong and weak points, any aspect that can be improved?...

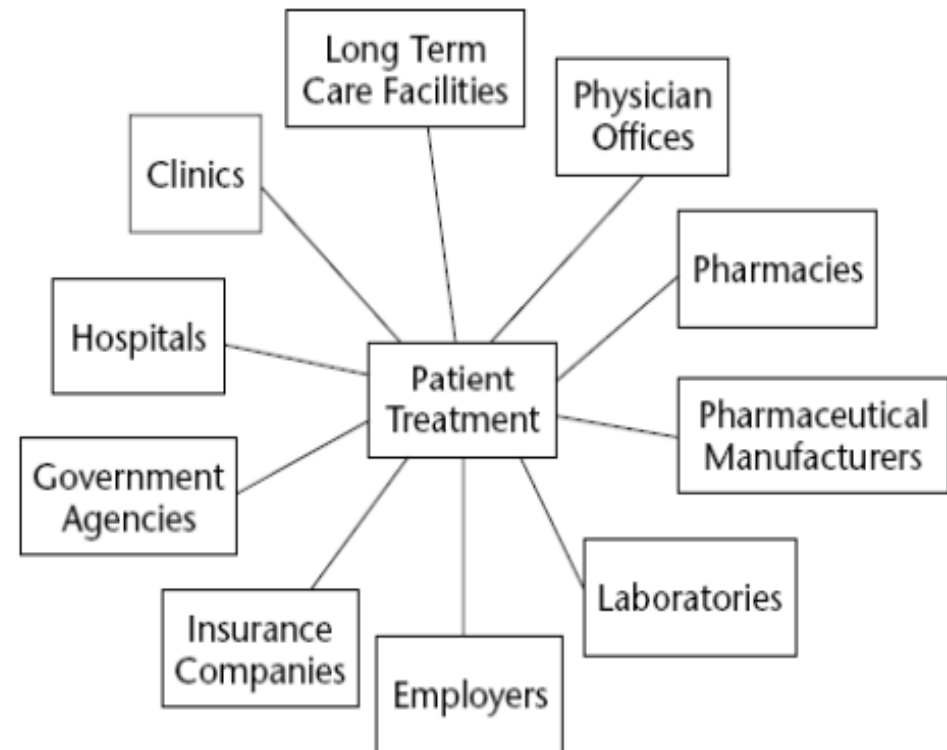
- **CATs: Clementine Application Templates: [CATs]**
  - Specific libraries with best practices for specific applications: mapping to a project type
  - Following the CRISP-DM standard.
  - Each flow CAT is assigned to a phase of CRISP-DM.
- **Templates:**
  - Telco CAT - loyalty and customer retention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

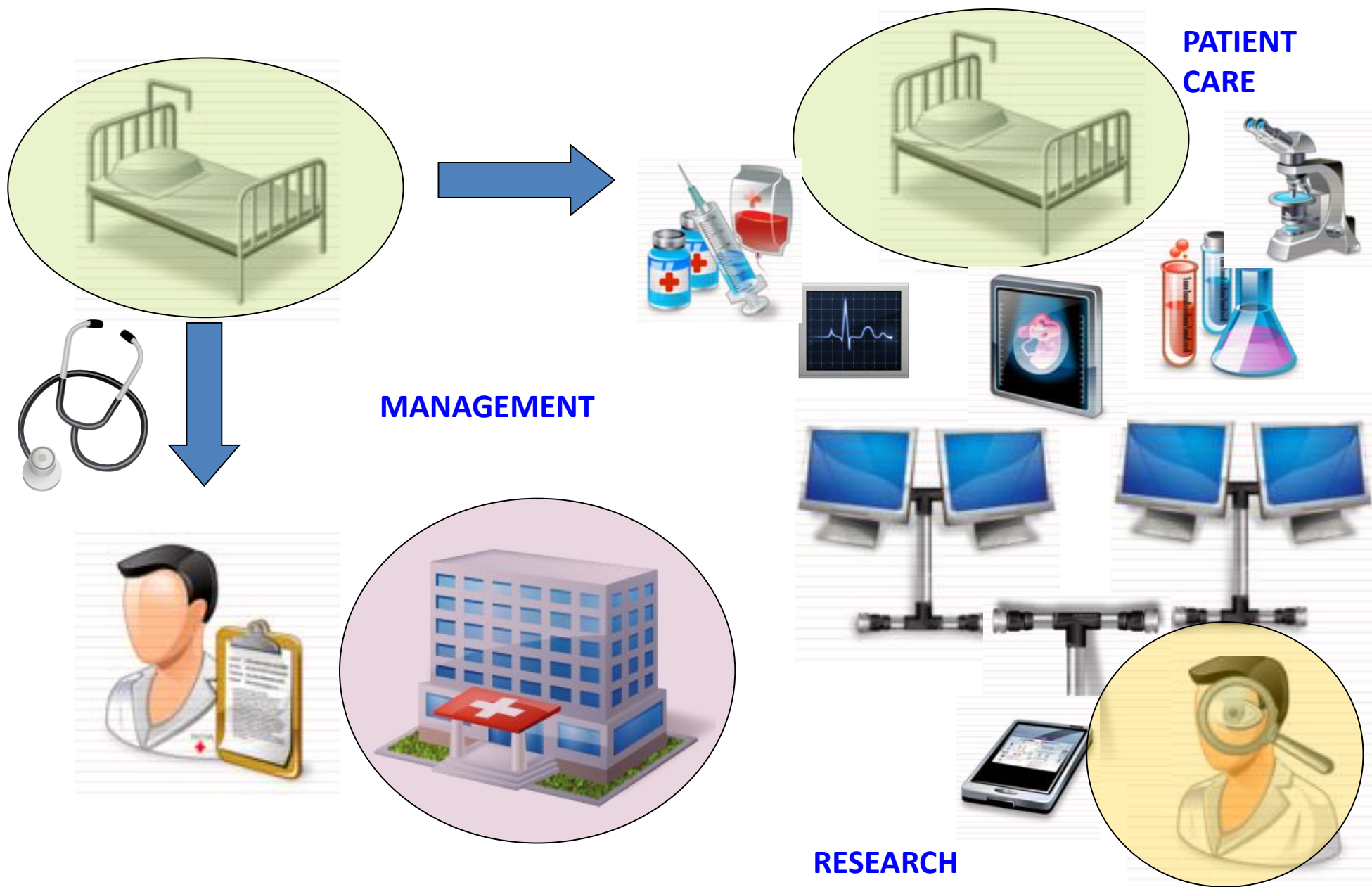
- SEMMA: SAS Enterprise Miner proposal for data mining projects
- Sampling:
  - A good sampling strategy almost guarantees good patterns
  - Rare patterns are statistically discovered
  - Subsets: training, validation, test
- Explore: simplify the problem



## ◆ Application

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





## FASE 3 Data preparation

**3.1  
Select**

**3.2  
Clean**

**3.3  
Modify**

**3.4  
Integrate**

**3.5  
Format**

### • Besides:

- Imprecise data, noise, incomplete data
- Subjective data due to 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

- Proprietary EHR
- Too many standards
- Low interoperability



### • It implies:

- Big difficulty to acquire and consolidate changing data



