Welcome all...

... to this course whose lectures are on an introduction to one of the most important applications for Data Science and Business Informatics:

Information storage and management
to support
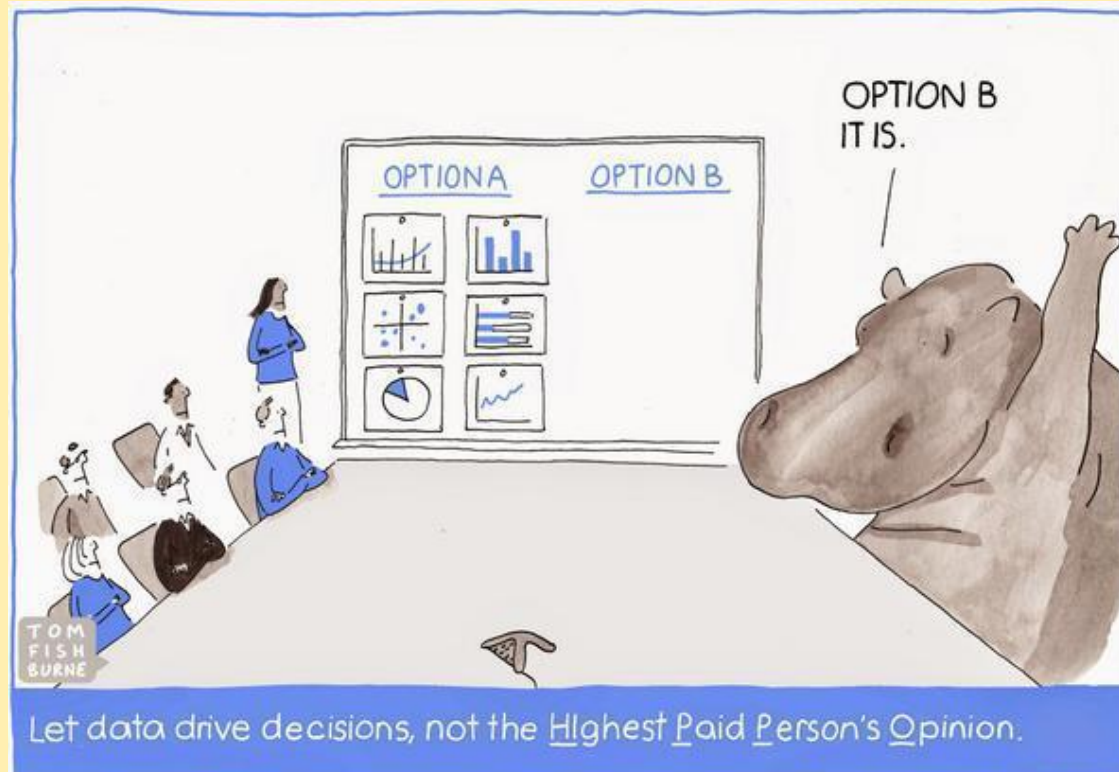**business decisions** of organizations.

Lecturer
Salvatore Ruggieri
salvatore.ruggieri@unipi.it
Office hours: Tuesdays 14:00-16:00 or by appointment in presence/on Teams

FACT In organizations, often the most important decisions are not based on fact (**informed decisions**), but, unfortunately, on intuition and experience of managers.

Let data drive decisions, not the Highest Paid Person's Opinion.

FROM DECISIONS TO INFORMED DECISIONS

FACT   In organizations, often the most important decisions are not based on fact (**informed decisions**), but, unfortunately, on intuition and experience of managers.

FACT   Organizations (companies) accumulate large quantity of data, that are often a resource scarcely used.

FACT   **Decision support information systems professionals** with a **business perspective** are needed to create company success and are rewarded by the job market.

FACT   Companies to compete **today** must use **data-intensive Business Intelligence techniques** to make better and timely informed decisions.

A set of methods and tools for interactive data analysis used to understand and analyze business performance in order to obtain useful information to **support (unstructured) decision making**.
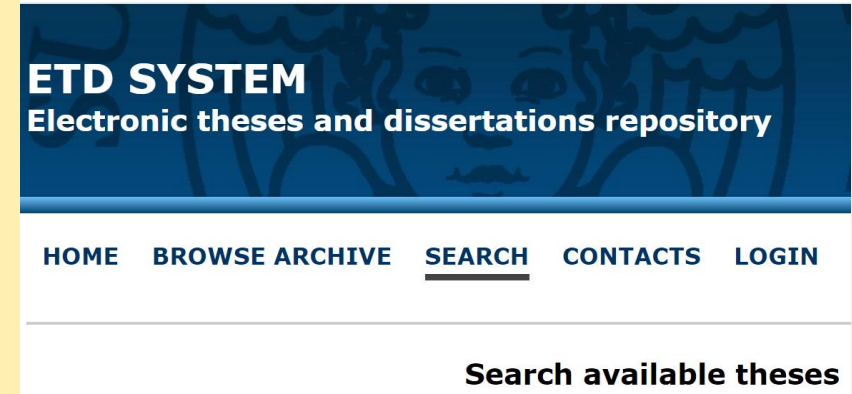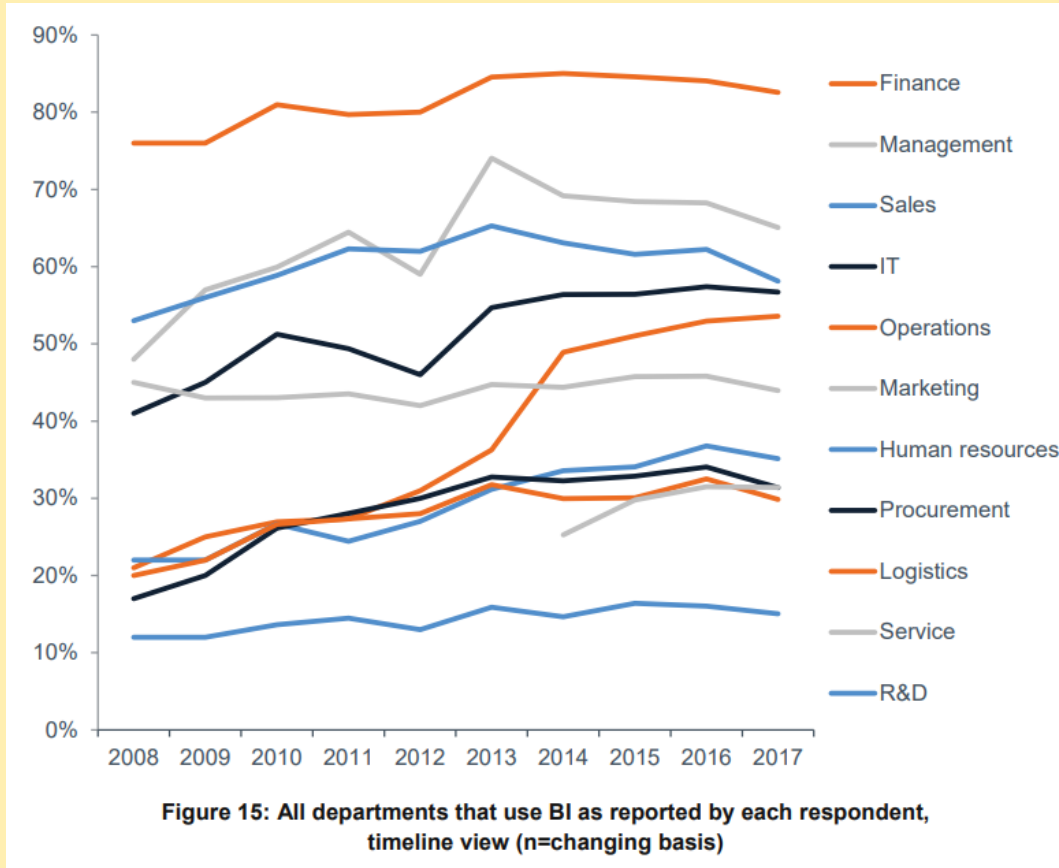
The term **intelligence**...

... is used to mean search for something interesting, as in the **Intelligence Service**.

PROFILE OF COMPANIES THAT USE BI

BI is used by medium to large companies that operate in competitive/global markets

Figure 15: All departments that use BI as reported by each respondent, timeline view (n=changing basis)

Source: BI Survey 2017

The design, implementation and use of a specific database, called **Data Warehouse (DW)**, to produce useful information to support decision-making with **Business Intelligence** applications

Topics

– Data-Driven Decision Support Systems .
– Data Warehouse: Data Models, DW Design and Implementation
– Multidimensional Analysis: Analytic SQL.
– Data Warehouse Systems and Physical Design:

       Storage, Indexing, Materialized Views and Query Evaluation

DECISION SUPPORT SYSTEMS Module II: Laboratory of Data Science covers **tools and technology** of Data Warehousing and Data Mining.

Before taking this module you should be comfortable with...

Fundamentals of relational systems.
Database modeling.
Relational algebra, basics of normalization theory,
SQL querying.

After taking this module you should be able to...

Design and implement data warehouses,
Understand the concepts and techniques in data warehousing,
Use a data warehouse to extract strategic information,
Pursue further studies and research in data warehousing, large data
analysis, business intelligence tools, and data mining.

Lecture Notes:

A.Albano, S. Ruggieri. Decision Support Databases Essentials, 2023

Course home page, teaching material, and recorded lessons (past years!):

http://didawiki.di.unipi.it/doku.php/mds/dsd/

Microsoft Teams channel for:
- Q&A
- Notebook for exercises
- Files for exercises

The teaching is integrated with exercises and discussion of the student solutions.

Some software tools will be used, but the main focus is methodological and problem solving based.

Examination. Oral exam on the basis of a written test (examples on website).
No mid-terms! Final grade of DSS (12 ECTS) = 50% DSD + 50% LDS
LDS project can be discussed only after passing DSD and $\leq$ 1 year since then.

Each organization to pursue specific objectives consistent with its mission uses:

a **structure**

several resources (human, financial, physical, information), and

several processes to transform resources into goods and services

Kind of processes: Anthony Model

Strategic activities
(long-range planning)

Tactical activities
(short-term planning, and control)

Operational activities
(day-to-day operations)

TYPES OF DECISIONS

Structured: It is one for which a well-defined decision-making procedure exists

Unstructured: It is one for which a well-defined decision-making procedure does not exists. The decision depend **only** on the manager experience.

Semistructured: It is one that has some structured aspects but cannot be completely structured. The decision depend **also** on the manager experience.

There is no strict correspondence between the types of decisions and their organizational impact (strategic, tactical, operational). In general, moving down in the Anthony pyramid the decisions tend to be more structured.

ANTHONY MODEL

Strategic activities
(long-range planning)

Tactical activities
(short-term planning, and con

Operational activities
(day-to-day operations)
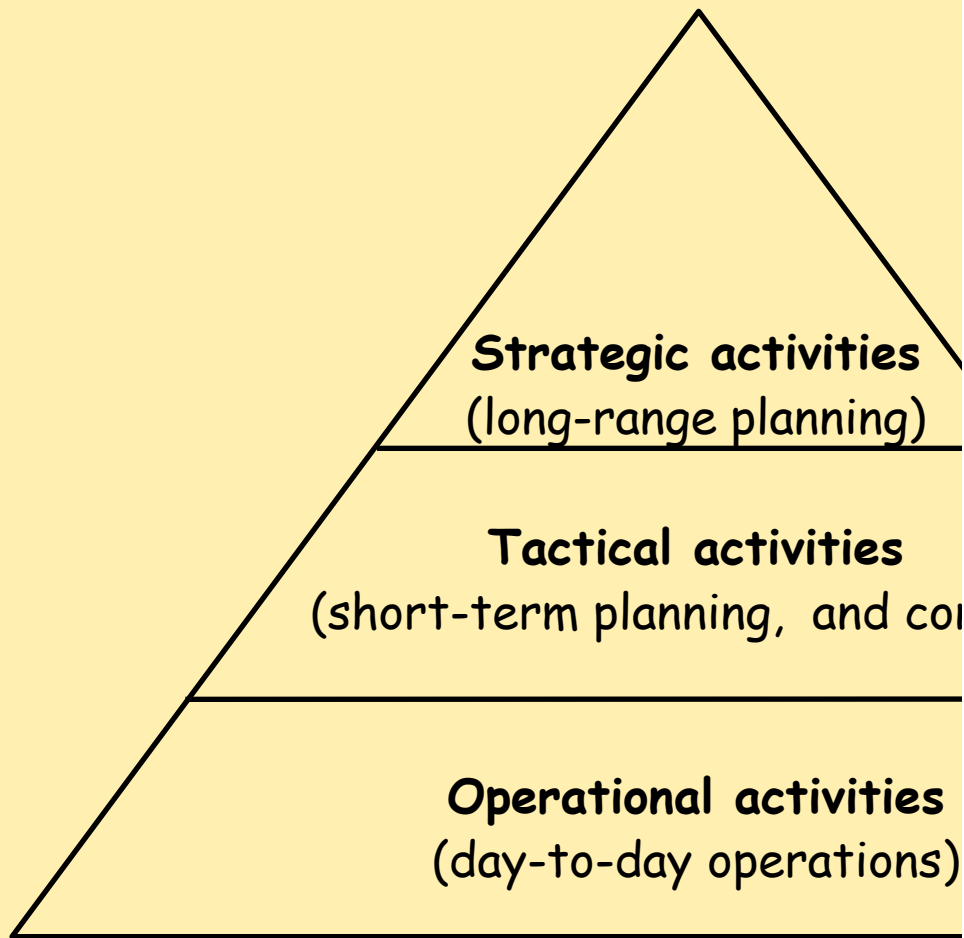
Figure 1   **Information Systems: A Framework**

|	Operational Control	Management Control	Strategic Planning
Structured	Accounts Receivable	Budget Analysis — Engineered Costs	Tanker Fleet Mix
	Order Entry	Short-Term Forecasting	Warehouse and Factory Location
	Inventory Control		
Semi-Structured	Production Scheduling	Variance Analysis — Overall Budget	Mergers and Acquisitions
	Cash Management	Budget Preparation	New Product Planning
Unstructured	PERT/COST Systems	Sales and Production	R&D Planning

FACT For all processes **information is a key resource**

FACT An **Information System** is a system whose purpose is to collect, store, process, and communicate information relevant to an organization.

FACT Organizations have used information systems for centuries and they have used a variety of technologies to process information (Ebla clay tablets, 2500 BC).
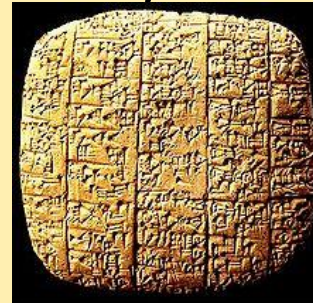
FACT Nowadays organizations have a **computerized information system.**

A **Computerized Information System** is a subset of the **Information System.**

In the following, for brevity, we use the term

Information System
for
Computerized Information System.

DBMS

DW

DSS

ESS

MIS

OLTP

DM

GIS

BI

OLAP

MRP

SRM

CRM

KMS

EDP

SCM

TPS

ERP

SI

A TAXONOMY OF INFORMATION SYSTEMS

INFORMATION SYSTEM

Bringing the business to the WEB

Analyzing the business

E-COMMERCE — **OPERATIONAL** — **DECISION SUPPORT**

Performing the business

MODEL-DRIVEN

DATA-DRIVEN
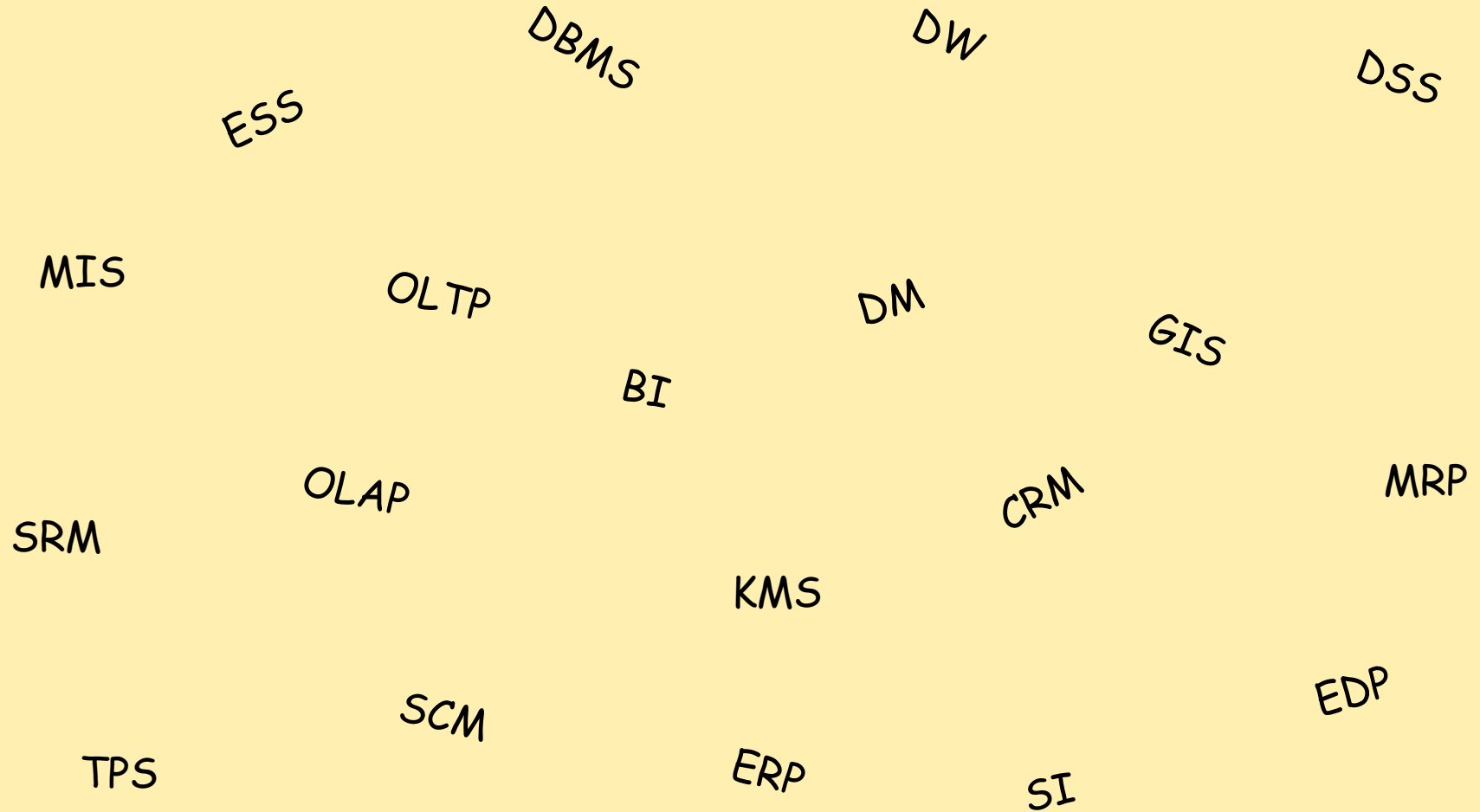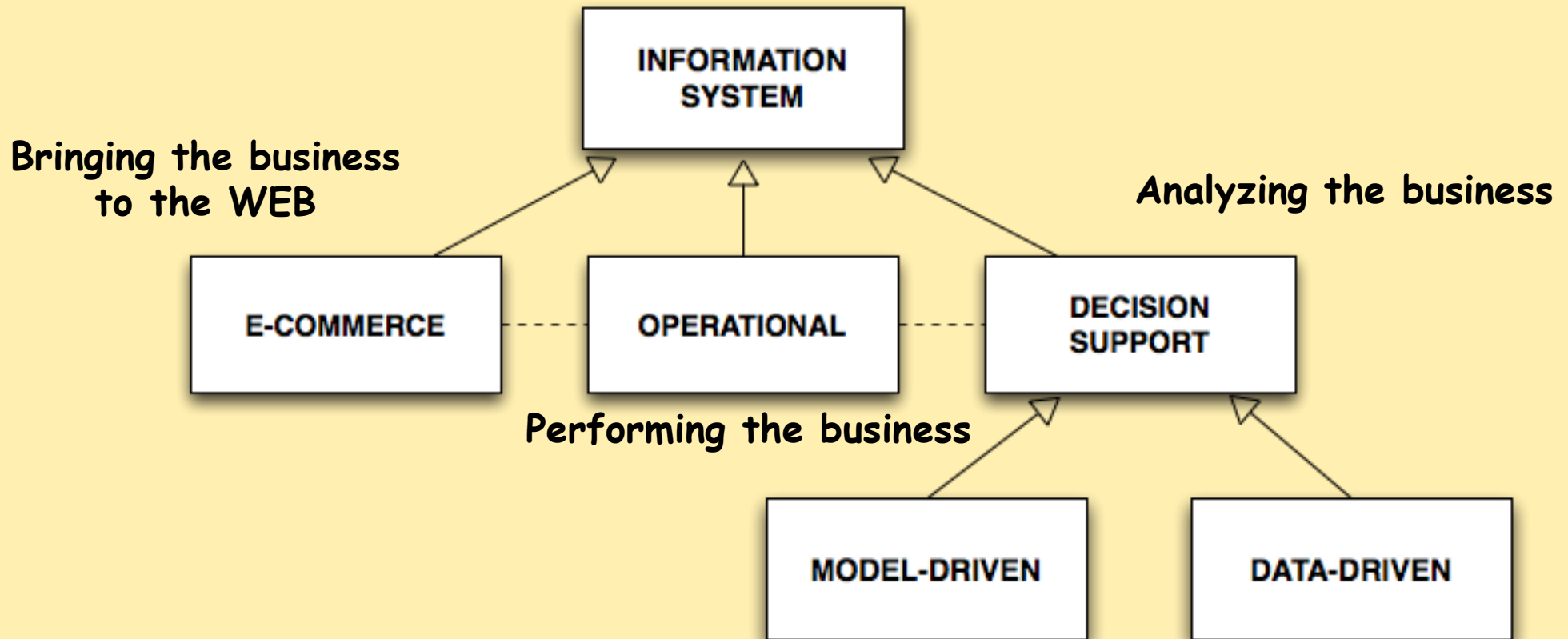
Decision Support System (DSS)
to provide information on which decisions can be made

Operational System

- Data are organized in a **DB**.

- Data are managed by a **traditional DBMS**.

- The applications **are used to perform** structured business operational activities.

What is modeled in a DB?

Decision Support System (DSS)

- Data are organized in a **separate specialized DB (Data Warehouse (DW))**.

What is modeled in a DW?

- Data are managed by a **specialized DBMS**.

- The **Business Intelligence** applications, **are used to analyze data**.

The extraordinary life of Oracle CEO Larry Ellison (Forbes Magazine)

Origin of the word **ORACLE**

In the mid-1970 **Larry Ellison** worked on a **CIA** (Central Intelligence Agency in USA) project for a "**system to give all answers to all questions about their data collections**". The project code-name was "**Oracle**".

ORACLE® is the major vendor of database systems. The company totals $50 billion in annual revenue.

Data

A representation of certain facts without context, which can be processed by computers.

442266	INF	2000	2003	Pisa
442277	TINF	2000	2004	Pisa
461176	IEA	2001	2003	Pisa
460076	TINF	2001	2003	Pisa
482299	INF	2002	2006	Pisa
481188	TINF	2002	2004	Pisa
441155	INF	2000	2002	Pisa
440033	TINF	2000	2002	Roma
498899	IEA	2003	2004	Bari
461178	INF	2001	2004	Bologna
……		…		……
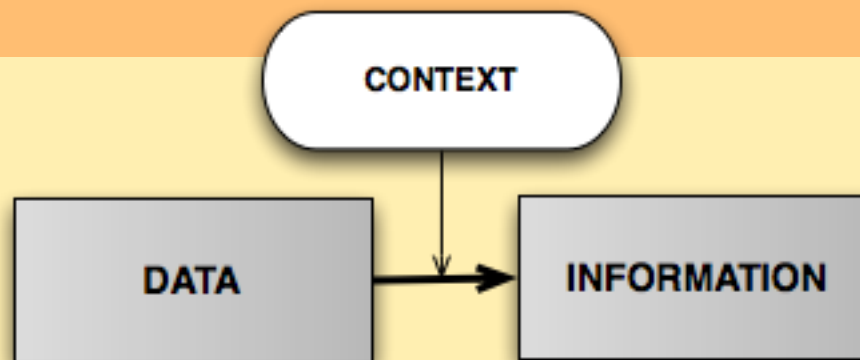
Information

Data, or a condensed form of them, become information when they are interpreted in a certain context.

StudentN	Course	YearEnrol	YearDegree	FromUniv
442266	INF	2000	2003	Pisa
442277	TINF	2000	2004	Pisa
461176	IEA	2001	2003	Pisa
460076	TINF	2001	2003	Pisa
482299	INF	2002	2006	Pisa
481188	TINF	2002	2004	Pisa
441155	INF	2000	2002	Pisa
440033	TINF	2000	2002	Roma
498899	IEA	2003	2004	Bari
461178	INF	2001	2004	Bologna
………		…		………

%ENROLMENTS
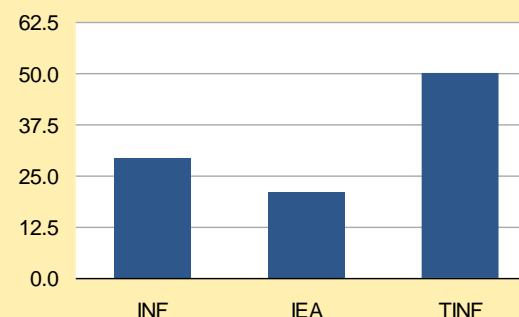
Knowledge

Information become **knowledge** when **expand** the **recipient** capability of understanding the reality, and allow him to make new predictions, informed and effective decisions, and proper actions.
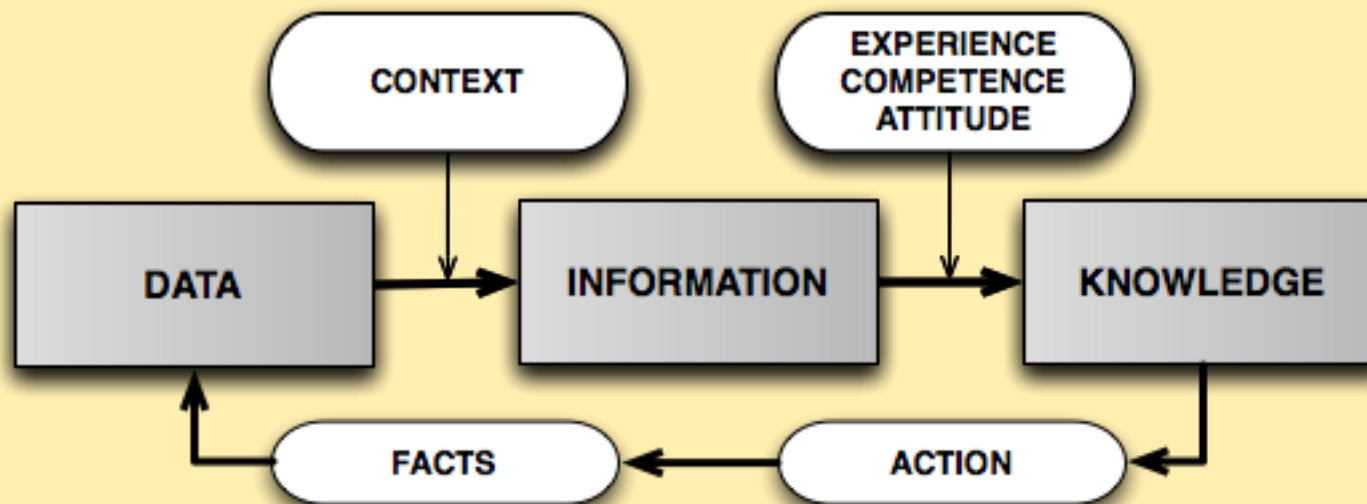
Spread of Covid-19 in Italy in February-March 2020
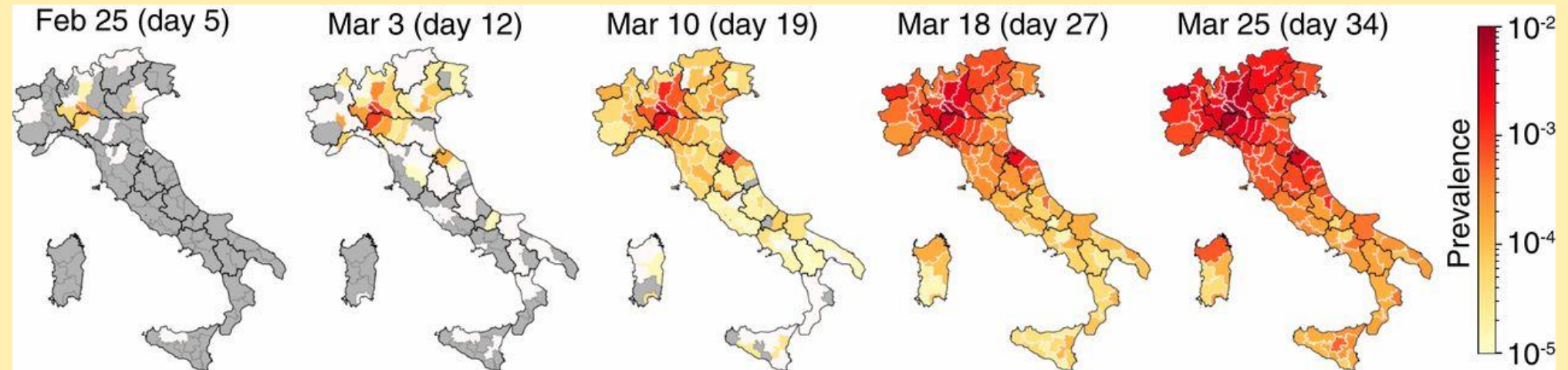
Reports: To find out what happened.

Multidimensional Data Analysis: To explore summary of data interactively
looking for useful information (knowledge).

Data Mining: To discover useful information (knowledge) by extracting
descriptive/predictive abstractions (models) from data.

In what follows the attention will be on **Multidimensional Data Analysis**

Let us explore the sales data stored in the table

Sales(Product, Market, Date, Revenue)

For 2011, the total revenue, by semester.

Revenue by Semester Year 2011	
Semester	**Revenue**
1	16 000
2	16 000
Total	32 000

Let us see if we can find more information with other business questions.

For 2011, the total revenue, by market

For 2011, the total revenue, by product

Revenue by Market Year 2011	
Market	**Revenue**
M1	8 000
M2	8 000
M3	8 000
M4	8 000
Total	32 000

Revenue by Product Year 2011	
Product	**Revenue**
P1	8 000
P2	8 000
P3	8 000
P4	8 000
Total	32 000

For 2011, the total revenue by semester, by product

Revenue by Semester, by Product
Year 2011

Semester	P1	P2	P3	P4	Total
1	4 000	4 000	4 000	4 000	16 000
2	4 000	4 000	4 000	4 000	16 000
Total	8 000	8 000	8 000	8 000	32 000

For 2011, the total revenue by semester, by market

Revenue by Semester, by Market
Year 2011

Semester	M1	M2	M3	M4	Total
1	4 000	4 000	4 000	4 000	16 000
2	4 000	4 000	4 000	4 000	16 000
Total	8 000	8 000	8 000	8 000	32 000

For 2011, the total revenue
by semester, by product, by Market

OK, now we have got
something interesting !

Revenue by Semester, by Product, by Market Year 2011						
Semester	Product	M1	M2	M3	M4	Total
1	P1			3 000	1 000	4 000
1	P2			1 000	3 000	4 000
1	P3	1 500	2 500			4 000
1	P4	2 500	1 500			4 000
	Total S1	4 000	4 000	4 000	4 000	16 000
2	P1	4 000				4 000
2	P2		4 000			4 000
2	P3			1 500	2 500	4 000
2	P4			2 500	1 500	4 000
	Total S2	4 000	4 000	4 000	4 000	16 000
Total		8 000	8 000	8 000	8 000	32 000

The result must be well visualized...

Category Management Dashboard

Category Sales Analysis by Region

Select Metrics:
- Revenue
- LQ Revenue
- Profit
- LQ Profit

USA Revenue $7,573,463

$1,156,126
$1,380,43
$817,285
$2,768,44
$612,500
$463,556
$375,112

Selection
Southeast

Category Performance For Trailing 6 Months: Revenue vs. Forecast Revenue (USD)

Books [+]
$6,000
$4,000
$2,000
$0
Jul 2006 · Sep 2006 · Nov 2006
Aug 2

Movies [+]
$20,000
$10,000
$0
Jul 2006 · Sep 2006 · Nov 2006
Oct 2006 · Dec 2006

Electronics [+]
$120,000
$60,000
$0
Jul 2006 · Sep 2006 · Nov 2006 · Oct 2006 · Dec 2006

Region: So

Performance Details

Region: Southeast

LQ Revenue	LQ Profit	Category	Revenue	Profit	Sell-through %	Order Count	Units Received	End on hand
$12,216	$911	Music	$13,039	$737	5.9%	886	27,865	14,064
$60,168	$4,644	Movies	$64,021	$3,376	27.2%	4,135	21,685	11,075
$268,185	$49,368	Electronics	$285,029	$47,573	24.8%	832	4,810	2,526
$12,846	$2,972	Books	$13,023	$2,676	48.3%	850	495	911

ANOTHER EXAMPLE

https://marketingplatform.google.com/about/analytics/