Knowledge Extraction

Paulo Trigo Silva @ ISEL

The data, the formats and the questions placed on them

DataBase

DataWareHouse

DataMart (OLAP)

When?

Where?

By Whom?

DataSet (DataMining)

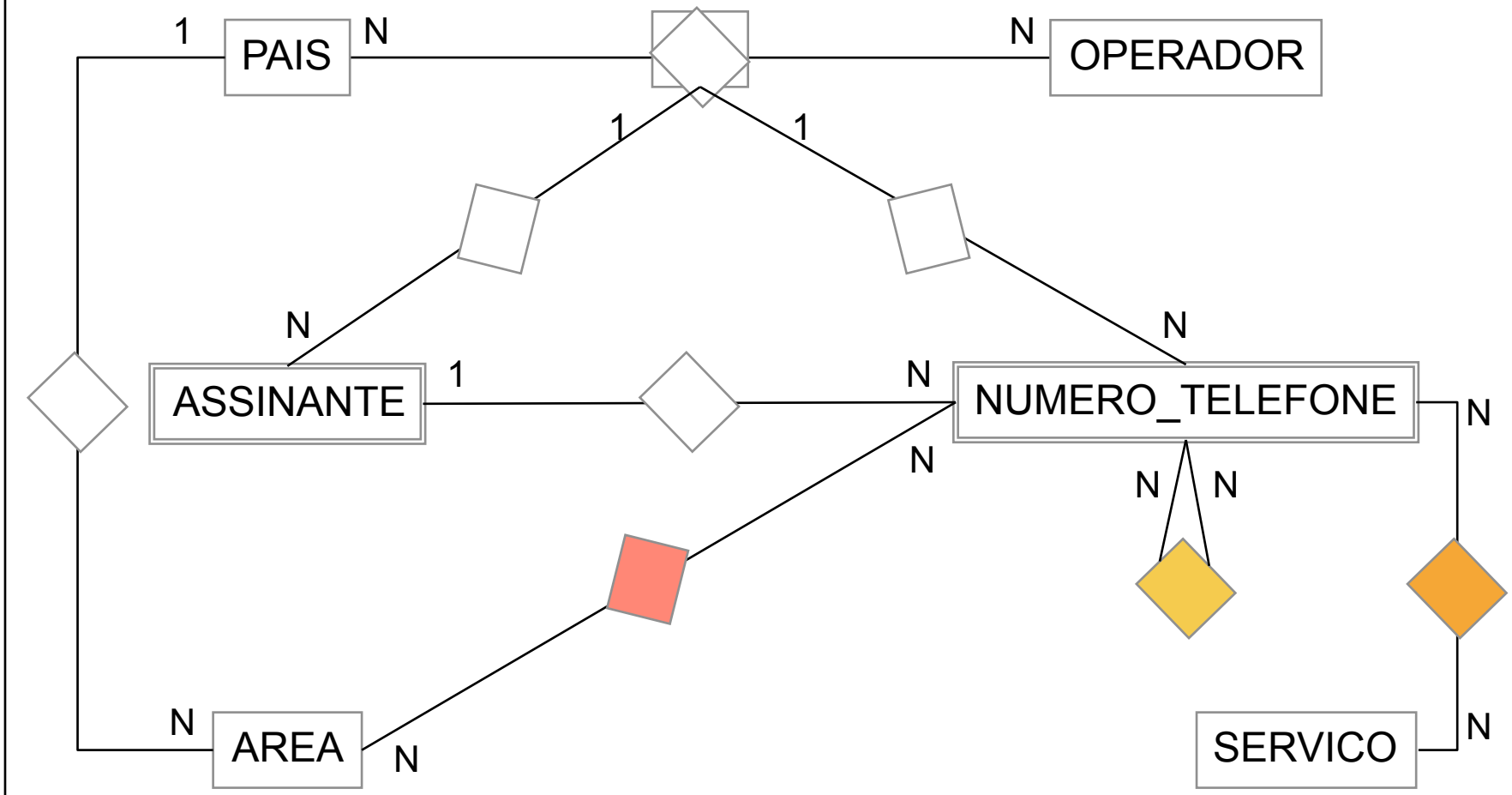
Why!

Paulo Trigo Silva @ ISEL

... the formats – different for different goals

DataBase	DataWareHouse	**DataMining**
SEVERAL-TABLES	SEVERAL-CUBES	**ONE-TABLE**
Relacional-Schema	*Data-Mart*	***DataSet***
query and manipulation of data via SQL	exploitation of aggregated data via MDX	extraction of "tendencies" via algorithms

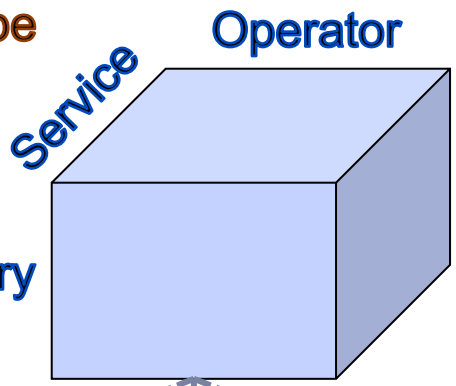
Relational-Schema (via Entity-Relationship model)

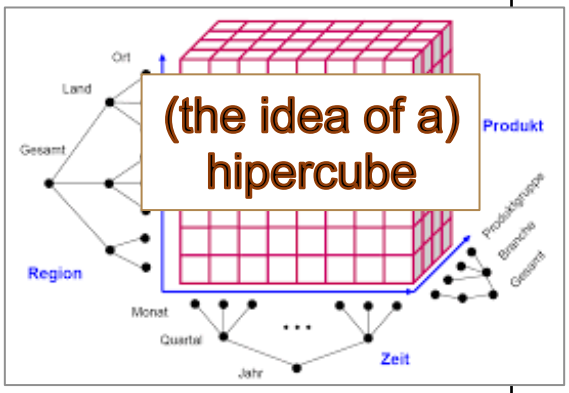


Paulo Trigo Silva @ ISEL

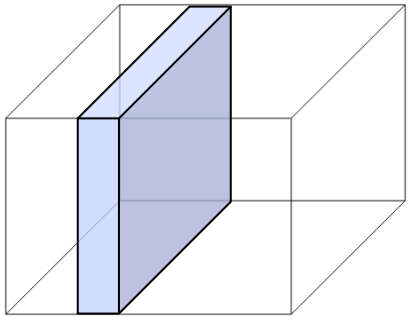
DataMart

hipercube

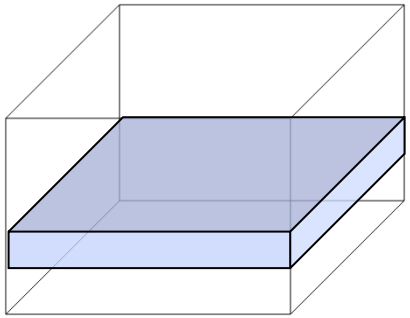




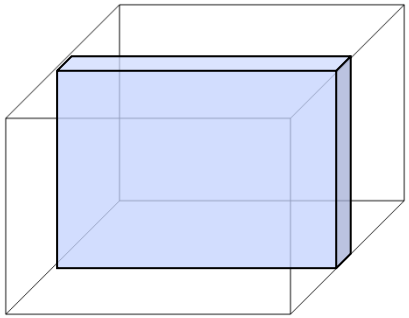
selection by the value of a dimension (*slice*)



all countries and services of an **operator**



all services and operatos of a **country**



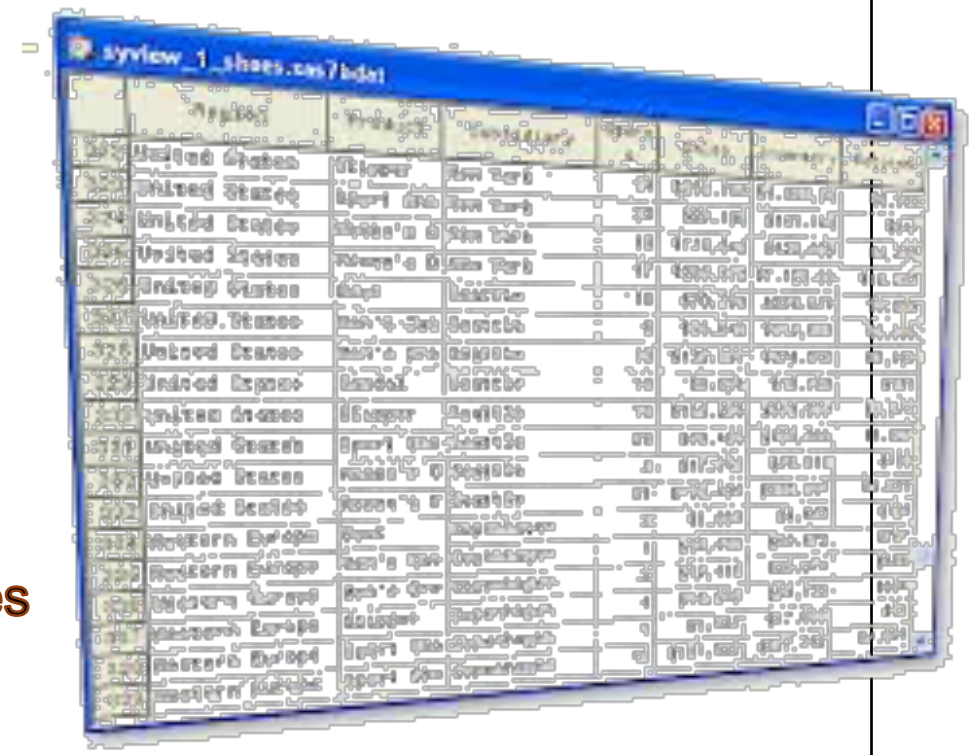
all countries and operators with a **serviço**

Paulo Trigo Silva @ ISEL

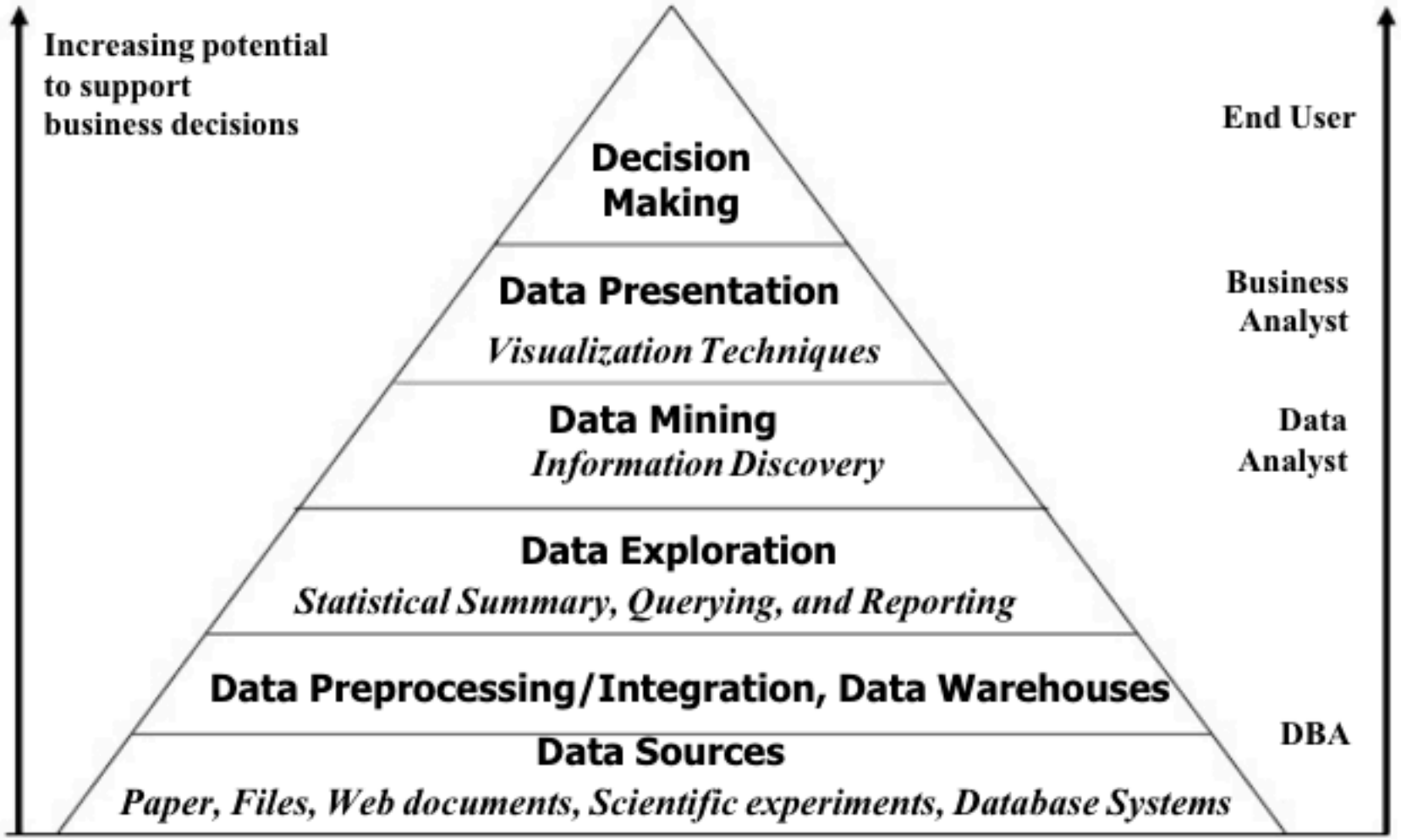
DataSet

The data is represented as:

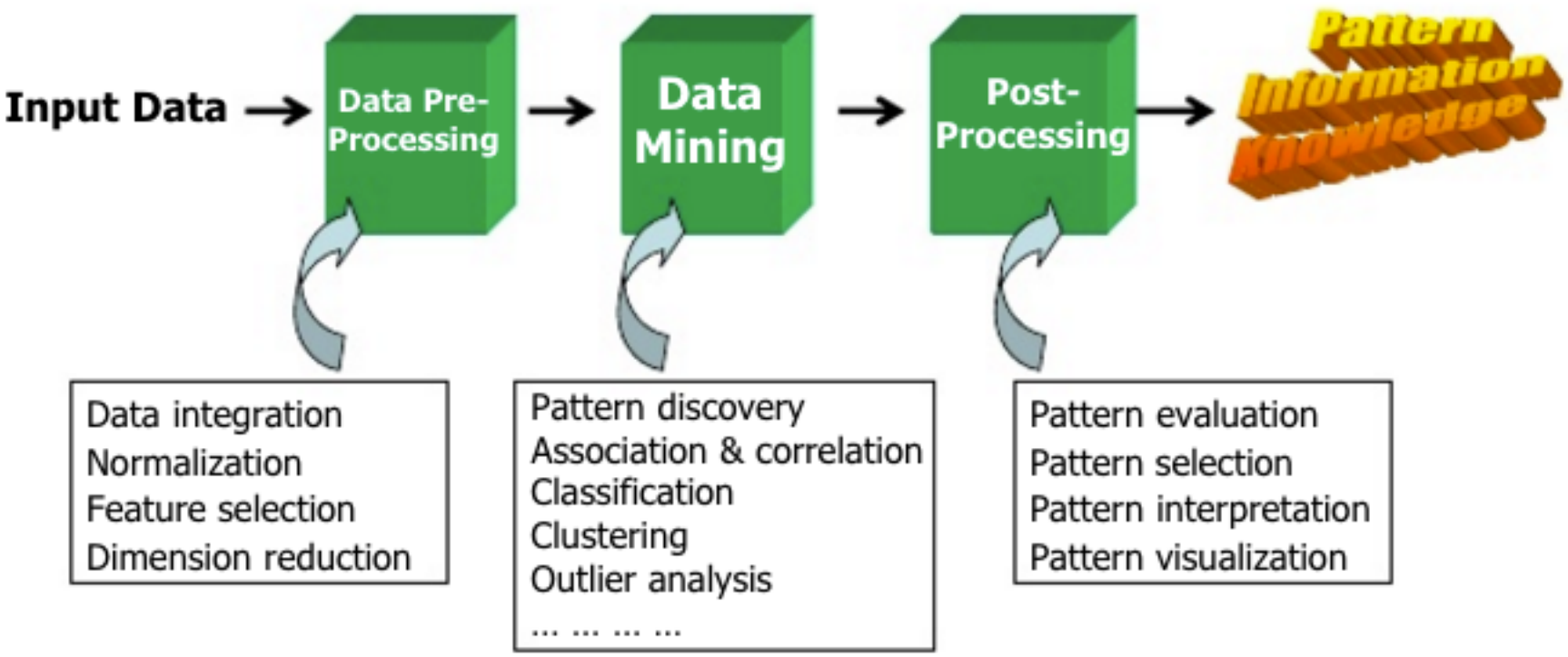
- a “single table”, i.e., a unique and “huge” vector with *D*-dimensions,
- each “column” represents a characteristic (“*feature*”) of what is intended to be analyzed,
- each “column” (“*feature*”) has values that vary according to their type.



Data-Mining and Business-Intelligence \ integrated view



KDD – Knowledge Discovery in Databases \ the process



this is a view from typical machine learning and statistics communities

Data Mining – Confluence of Multiple Disciplines