BIL3203 – DATABASE MANAGEMENT

Alper VAHAPLAR

2021 – 2022 ©

Data

- ## What is Data?
  - Datum (singular)
  - Unprocessed (raw) form of information... ☺
  - the result of a measurement, event or fact.
  - groups of information that represent the qualitative or quantitative attributes of a variable.
  - a collection of facts from which conclusions may be drawn; "statistical data".
  - known facts, worth to record.
  - Ex: age, eye color, price, date,
- ## Why is it necessary?
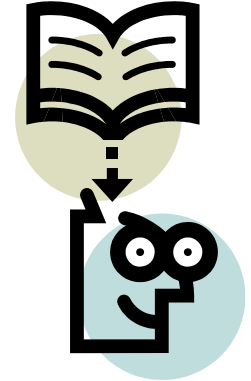
Information

- ## What is Information?

  - Processed form of Data … ☺

  - the data that has been processed to be meaningful to the person who receives it.

  - knowledge acquired through study or experience or instruction.

  - Collection of facts that decisions are made on.

  - Statistically analysed data.

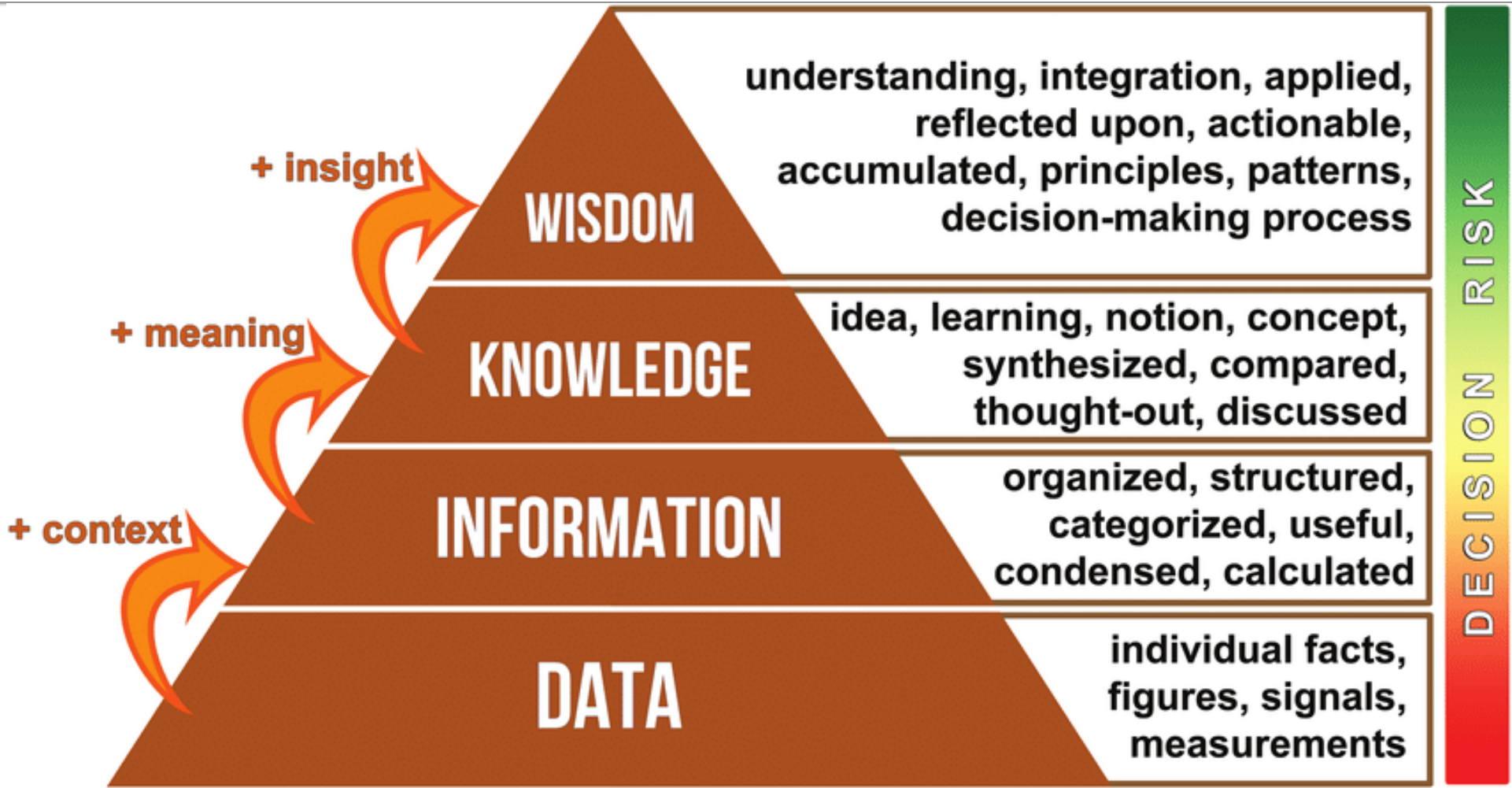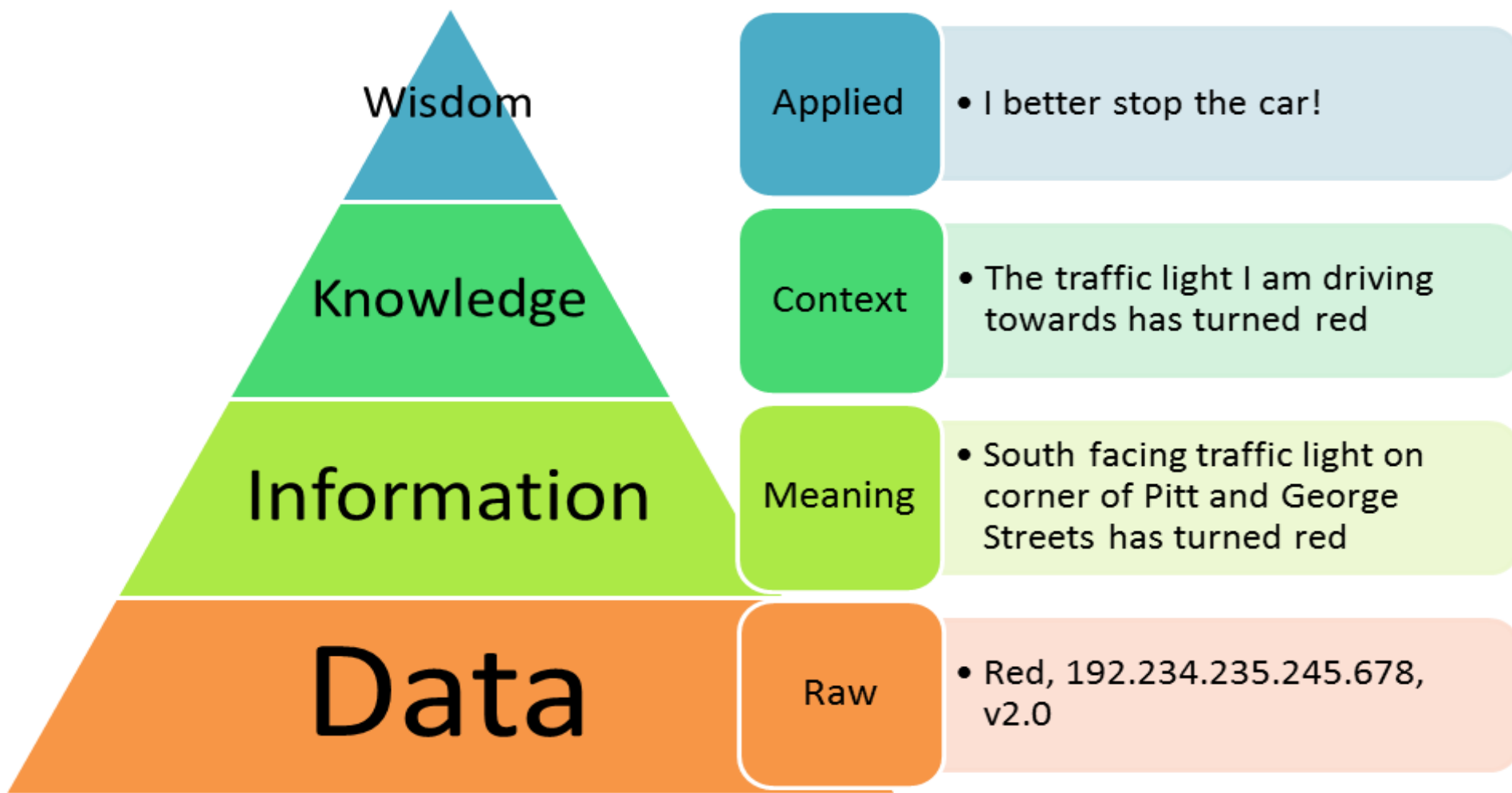 - Ex: increase in the amount of erytrocyte, decrease of sales, etc

Knowledge?

Expertise and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject.
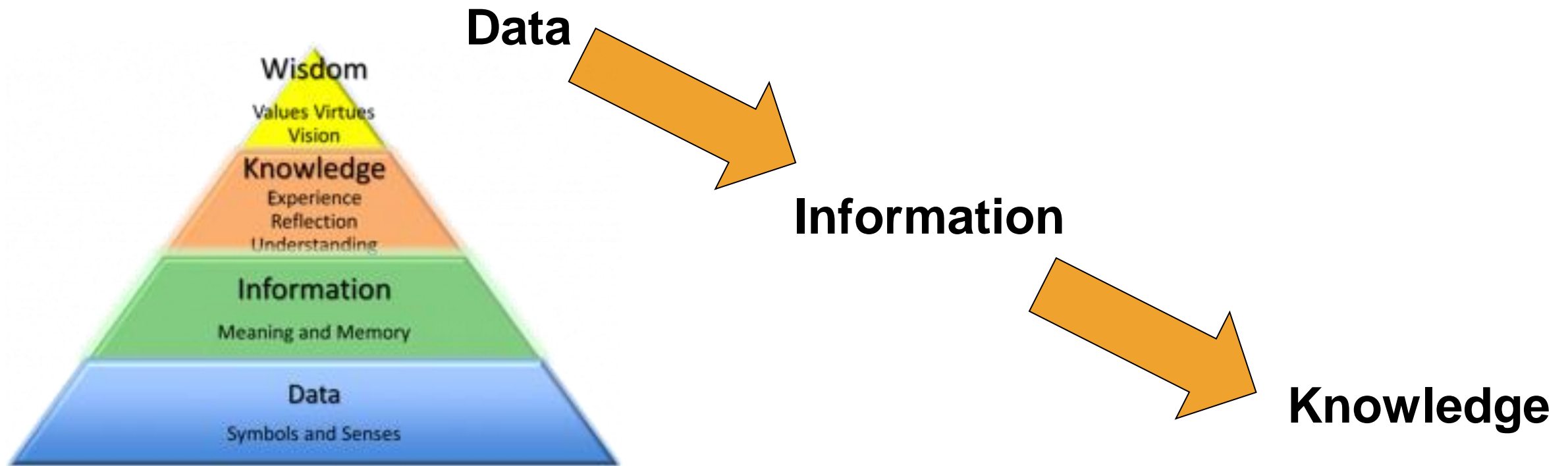
DIKW Pyramid

DIKW Pyramid

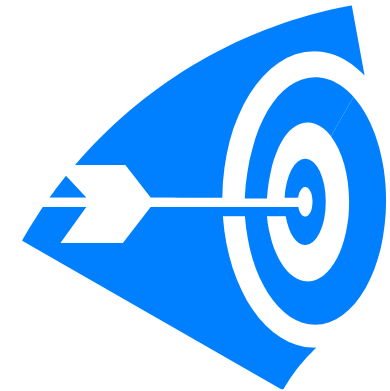# Characteristics of "Information"

- Accurate and Reliable

- Relevant and Timely

- Understandable and Transferable

- "*Expensive*" to collect

- Provide power and/or advantage

Where to store "Data"

- Stones,
- Wall of caves
- Animal Skins
- Papyrus Leaves
- Paper
- Computers
  - Flat Files - Sequential, Random Files
  - Formatted Files (Excel, Minitab,…)
  - Databases
  - Data Marts, Data Warehouses

Types of Variables

- Categorical
  - Qualitative
- Numeric
  - Quantitative

Types of Variables – Scale

- **Nominal Variables**
  - A variable used to put objects into categories,
  - Ex: color of an object, ID number (1, 2, 3, 4..)
- **Ordinal Variables**
  - similar to nominal variables, except that having values which can be arranged in a meaningful order,
  - Ex: small, medium, large.

Types of Variables – Scale

- **Interval Scaled Variables**
  - Interval-scaled variables are variables that take numerical values which are measured at equal intervals from a zero point or origin.
  - A unit of measurement exists.
  - However the origin does not imply a true absence of the measured characteristic.
  - Ex: temperature in Celcius,
- **Ratio Scaled Variables**
  - similar to interval-scaled variables except that the zero point does reflect the absence of the measured characteristic.
  - Ex: molecular weight, price in dollars.
  - Differences and ratios is meaningful.

Types of Variables – Scale

Scales of Measurement

Data	Nominal	Ordinal	Interval	Ratio
Labeled	✓	✓	✓	✓
Meaningful Order	✗	✓	✓	✓
Measurable Difference	✗	✗	✓	✓
True Zero Starting Point	✗	✗	✗	✓

Attribute Type		Description	Examples	Operations
Categorical (Qualiatative)	Nominal	The values of a nominal attribute are just different names, i.e., nominal attributes provide only enough information to distinguish one object from another. ($=$, $\neq$)	zip codes, employee ID numbers, eye color, sex: {*male, female*}	mode, entropy, contingency correlation, χ^2 test
	Ordinal	The values of an ordinal attribute provide enough information to order objects. ($<$, $>$)	hardness of minerals, {*good, better, best*}, grades, street numbers	median, percentiles, rank correlation, run tests, sign tests
Numeric (Quantitative)	Interval	For interval attributes, the differences between values are meaningful, i.e., a unit of measurement exists. ($+$, $-$)	calendar dates, temperature in Celsius or Fahrenheit	mean, standard deviation, Pearson's correlation, *t* and *F* tests
	Ratio	For ratio variables, both differences and ratios are meaningful. ($*$, $/$)	temperature in Kelvin, monetary quantities, counts, age, mass, length, electrical current	geometric mean, harmonic mean, percent variation

Data Types

- Characters (alphanumeric)
- Numerical (integers, floating point, real…)
- Date
- Image
- Voice
- Image + Voice (multimedia data)
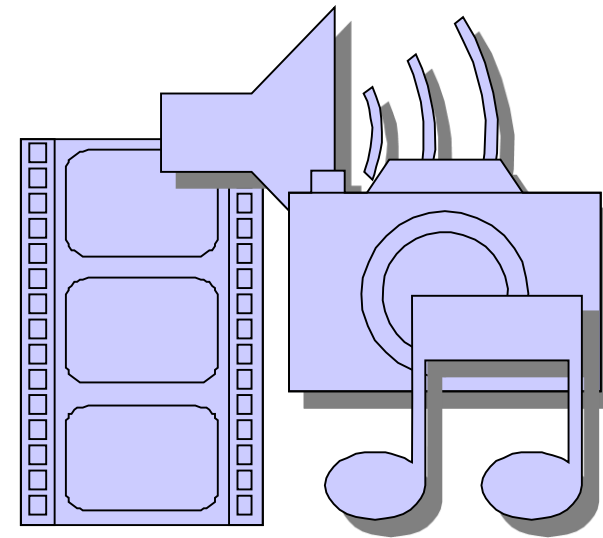
Operations on Data

- Adding new data
- Updating data
- Deleting data
- Reading data

Which Data are kept?

- Migros, BİM, Teknosa, etc
- Pictures, movies
- Facebook, Gmail, Instagram, Twitter, etc.
- Student Information
- Flight and Reservation data
- Patient and disease data
- E-commerce data
- Banking
- Internet

Data Collection

- Ex: A Grocery notebook
  - Who?
  - What?
  - How much/many?
  - When?
  - In which Price?

Data Collection

- Ex: Which data are kept for a "student"?
  - Name, surname, address, telephone, Birth Date, Date of beginning, etc.
  - Courses attended
    - Semester, Course Title, Instructor, discontinuity, etc.
  - Grades (midterm, final, etc.)
  - Grade of Diploma
  - Date of Diploma

Data Collection

- Ex: Which data are kept in Migros?
  - Product name
  - Amount of product
  - Price of product
  - Date and time of sale
  - Cashier name
  - Method of payment (cash or credit card)
  - Cost of purchase, amount of purchase
  - Date of purchase
  - Vendor of the product

- Find and analyse
  - a system that collects and stores data
  - data and data characteristics in this system
  - reason to keep that data
  - the media that keeps the data

and make a prediction about the amout of data to be collected in

  - 1 day,
  - 1 week (distribution of data over the days),
  - 1 month,
  - and 1 year