Disclaimer

- This presentation is purely for academic purpose and does not carry any commercial value.
- All images and photos used in this presentation are property of respective image holder(s) and due credit is provided to them. Images are used only for indicative purpose and does not carry any other meaning.
- All information and data in this slide are collected from open domain.

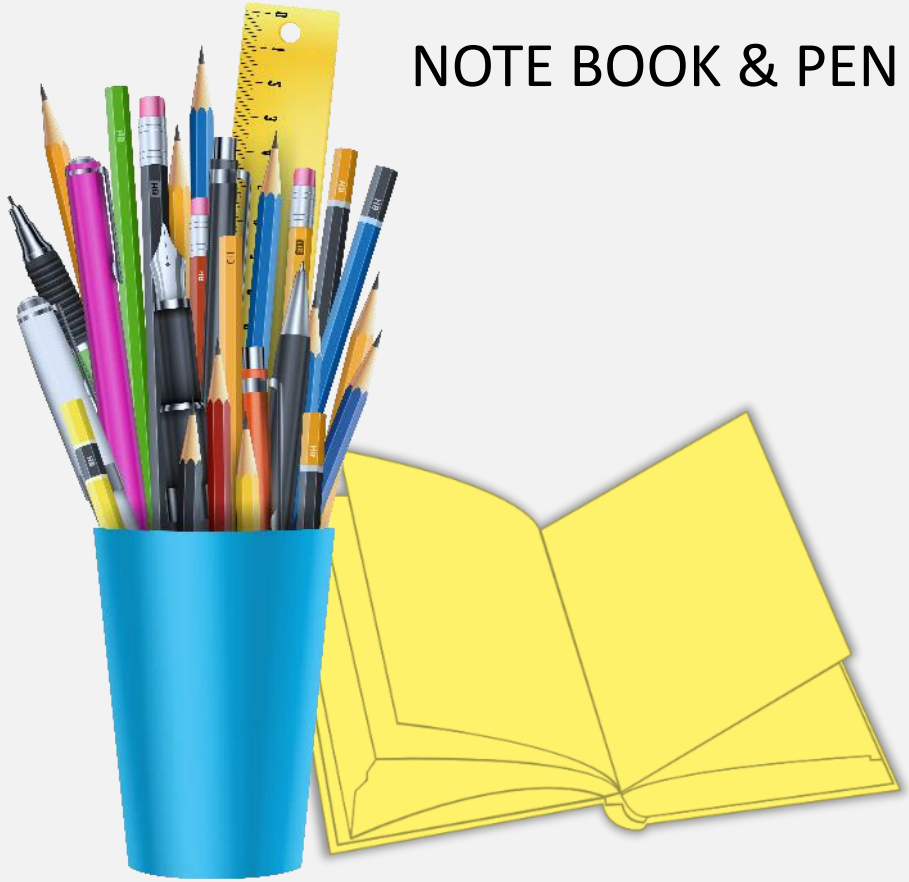
MANISH GODSE, Ph.D.(IIT Bombay)

Request & Instructions

PLEASE OPEN

CALCULATOR

NOTE BOOK & PEN



LAPTOP OR DESKTOP,
IF YOU HAVE.

PLEASE FOLLOW THIS

SILENCE

MUTE MIC

RAISE HAND

NO CHAT

SILENT MODE

DATA TYPES

BOOKS & REFERENCES

- NO

Table of Contents

DATA & DATA TYPES

1

WHAT IS DATA?

DATA is -

- **Facts** and **Statistics** collected together for reference or analysis.
- Facts, **Figures**, Statistics, Details, Particulars, Specifics, Features
- **Quantities**, **Characters**, or **Symbols** on which operations are performed by a computer
- **Things** known or assumed as facts, making the basis of reasoning or calculation.

DATA OBJECT & ATTRIBUTE

- **Data** is the collection of **data objects** and their **attributes.**
- An **attribute** is a characteristic of an object.

  Examples: body temperature, speaking language, etc.

  Attribute is also known as variable, field, characteristic, dimension, or feature.
- An **object** is the collection of attributes. Object is also known as record, case, sample, entity, or instance.
- **Attribute value** is number or symbol assigned to an attribute.

Attributes

Objects

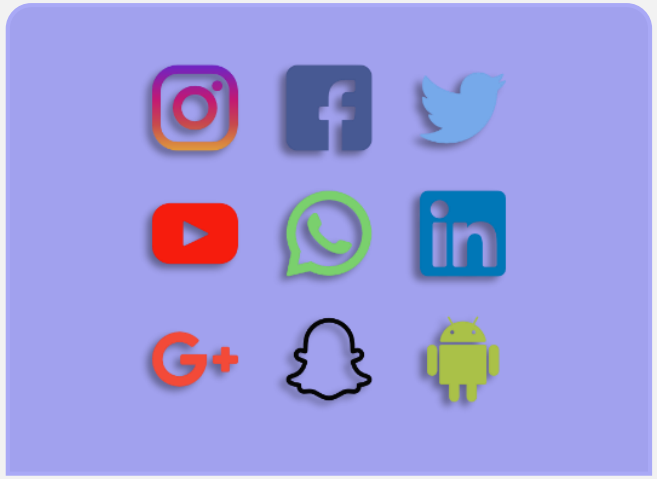
Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

DIFFERENT FORMS OF DATA

Numeric & Category

Text

Speech/Voice



Social Media

Geospatial

Video & Images

DIFFERENT TYPE OF DATA

NUMERIC & CATEGORY

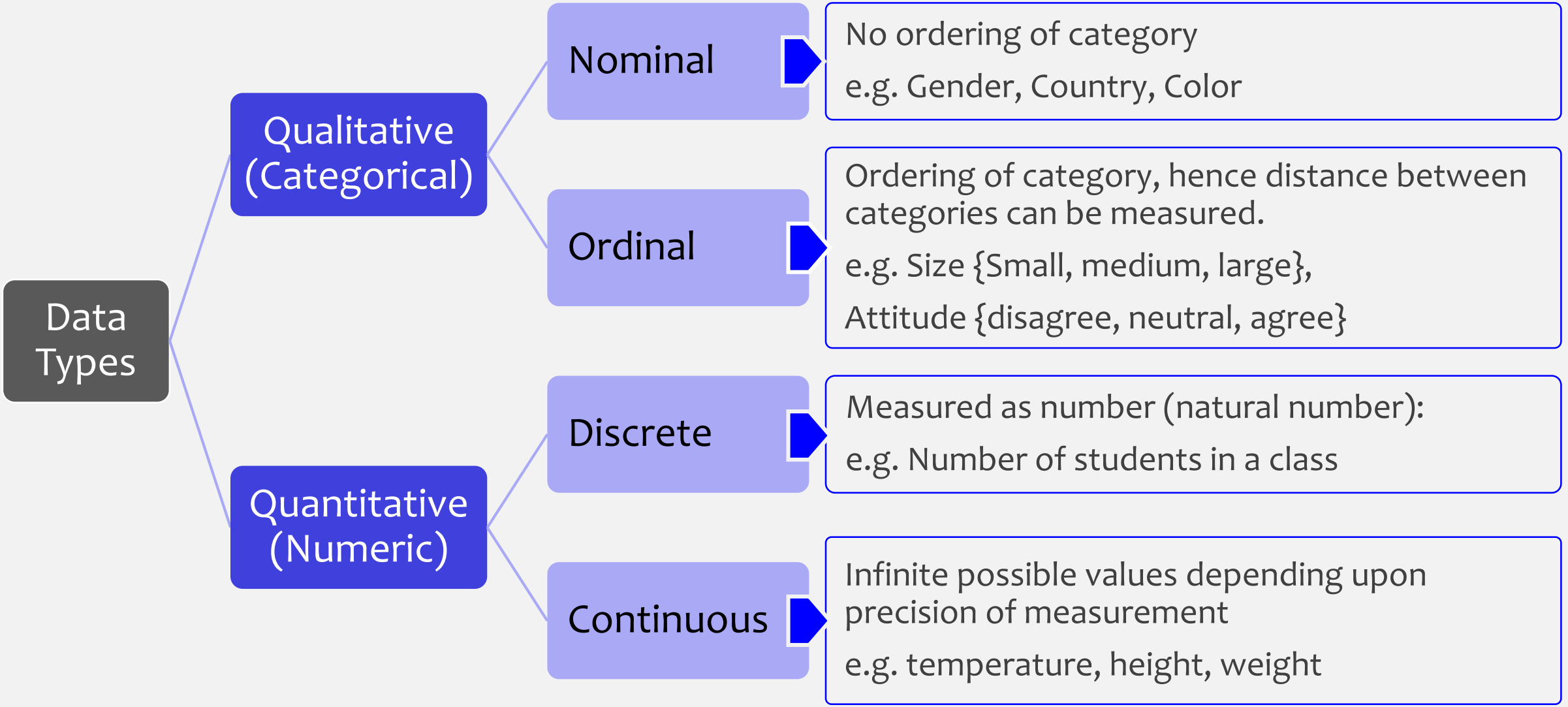
TEXT

IMAGE & VIDEO

AUDIO

GEOSPATIAL

STRUCTURED DATA TYPES



NUMERIC MEASUREMENT SCALE

Interval

Variables is measured on an interval scale with uniform difference in values.

e.g. Temperature 5, 10, 15, 20, 25, ……

Ratio

While comparing values, this scale is used.

e.g. % increase in height of students- it may be ratio of previous year height Vs. current year height

Circular

While measuring annual dates, clock times, etc. circular scale is used.

LABELLED & UNLABELLED DATA

2

LABELED DATA

- If data has a label or tag , then it is labelled data.
- Label depends on the context of the problem.

LABELLED DATA FOR REGRESSION

Height	Weight
150	53
155	58
160	62
170	68
185	77

When you predict the weight, then **Weight is a label.**

LABELLED DATA FOR CLASSIFICATION

Dog
15 kg
11 year old

Dog
6 kg
3 year old

- If you predict type (e.g. Dog, Cat), then type is label.
- If you predict weight (e.g. 15 kg, 6 kg), then weight is label.
- If you predict age (e.g. 11 year, 3 year), then age is label.

UNLABELED DATA

- If data doesn't have a label or tag , then it is unlabeled data.

UNLABELED DATA CLUSTERING

Age	Monthly purchase ($)	Top buying category	Gender
25	2000	Cosmetics	Female
35	850	Books	Male
36	800	Electronics	Male
45	1500	Fashion	Female
50	1200	Grocery	Female

- *Discover groups of customers, which can be used for targeted marketing program. There is target variable to solve this problem.*

UNLABELED DATA FOR CLUSTERING

- *There is no label or tag to above images.*

WHEN TO USE LABELLED DATA?

It is good to have labeled data than unlabeled data. In set sometime data may be partially labeled.

CLASSIFICATION & CLUSTERING

- With a labeled dataset, problem of **classification** can be solved.
- With an unlabeled dataset, problem of **clustering** can be solved.

SUPERVISED & UNSUPERVISED

- The set of algorithms where labeled dataset is used are called **supervised** learning.
- The set of algorithms where unlabeled dataset, is called **unsupervised** learning

DATA SOURCES & ACQUISITION

3

HOW DATA IS GENERATED?

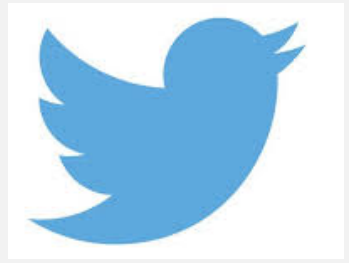
Data is generated by various **FUNCTIONS**.

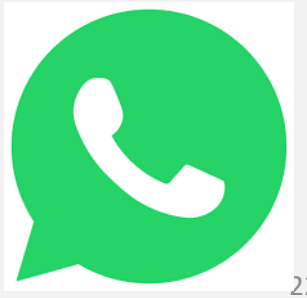
Data is generated by various **PROCESSES**.

Data is generated by various **ACTIVITIES**.

CAN YOU RECOGNIZE THESE LOGOS?

...and many more...





IMPACT OF ERP ON BUSINESS

- Accurate data and information
- Timely access to data and information
- Optimization & standardization of processes

DIGITAL FORM AND STORED ON CLOUD

INFLUENCE OF TECHNOLOGY ON BUSINESS

Time to time decision making and no delay

Reduced capital investment
(Virtual store, Virtual office)

Reduced cost
(Paperless office, Work from home, Video conferencing)

Reduced response time to customer service and employees
(E-mail, Chabot, Video conferencing, Collaboration-Slack)

Increase in performance & efficiency
(Software for Project, Time, and Product Management)

DATA GOVERNANCE

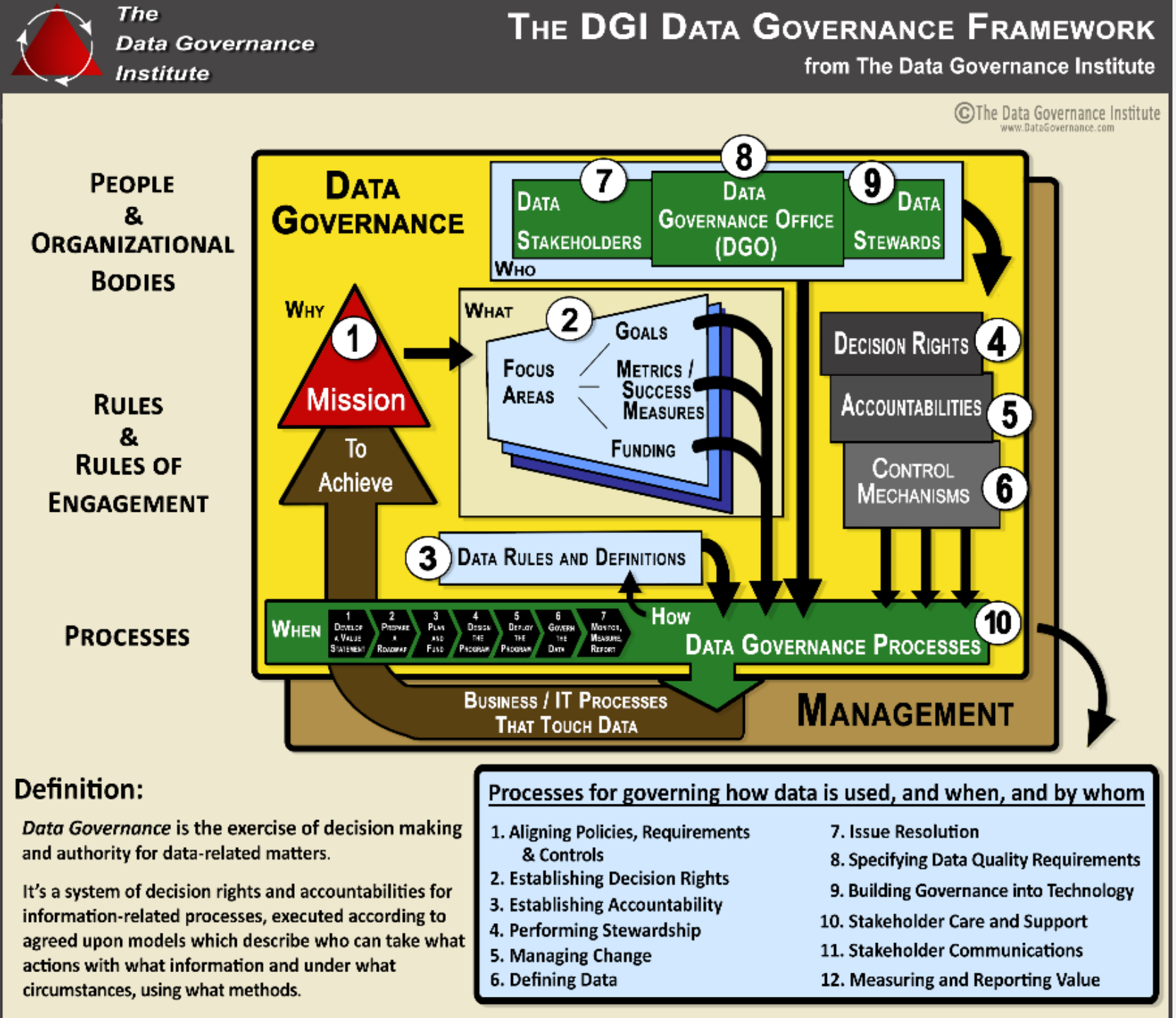
4

DATA OWNERSHIP

Data is owned by **FUNCTION**.

(Marketing, Finance, HR, Supply Chain, Operations, etc.)

Data is managed by **INFORMATION TECHNOLOGY** on the behalf of **FUNCTION**.



Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods.

QUESTION AND ANSWERS