

# Data Mining & Data Warehouse

Associate Professor Dr. Raed Ibraheem Hamed

University of Human Development,  
College of Science and Technology

2016 – 2017

Department of IT- DMDW - UHD



# Points to Cover

---

- ❖ What is DATA
- ❖ What is Information
- ❖ Information Example
- ❖ Data vs. Information
- ❖ What is Knowledge
- ❖ Knowledge Example
- ❖ Knowledge Workers
- ❖ The Main Idea

# What is DATA

---

**Data:** a collection of facts used for reference or analysis; a series of observations, measurements, or facts.

## Example of Data

The data shown below are Mark's scores on five Math tests conducted in 10 weeks.

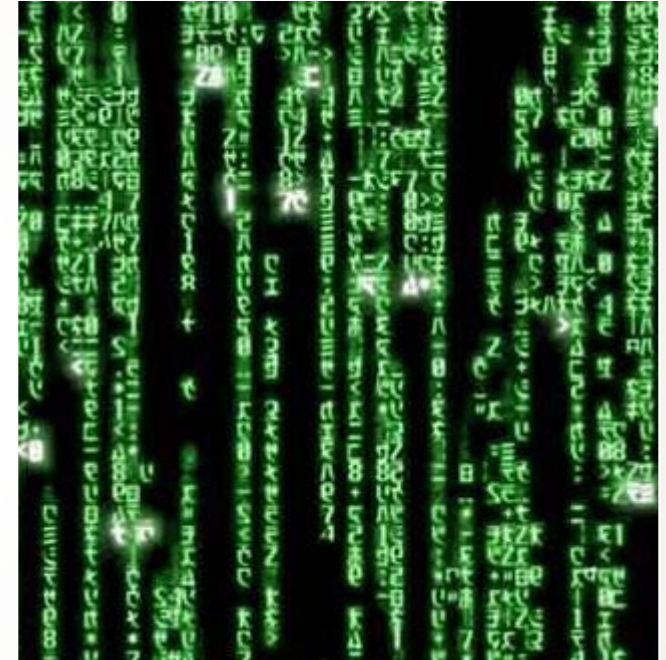
- 45, 23, 67, 82, 71
- Yes, No, No, Yes

# What is DATA

---

Data are raw facts and figures that on their own have no meaning

These can be any alphanumeric characters i.e. text, numbers, symbols





# What is DATA

Data: a collection of facts used for reference or analysis

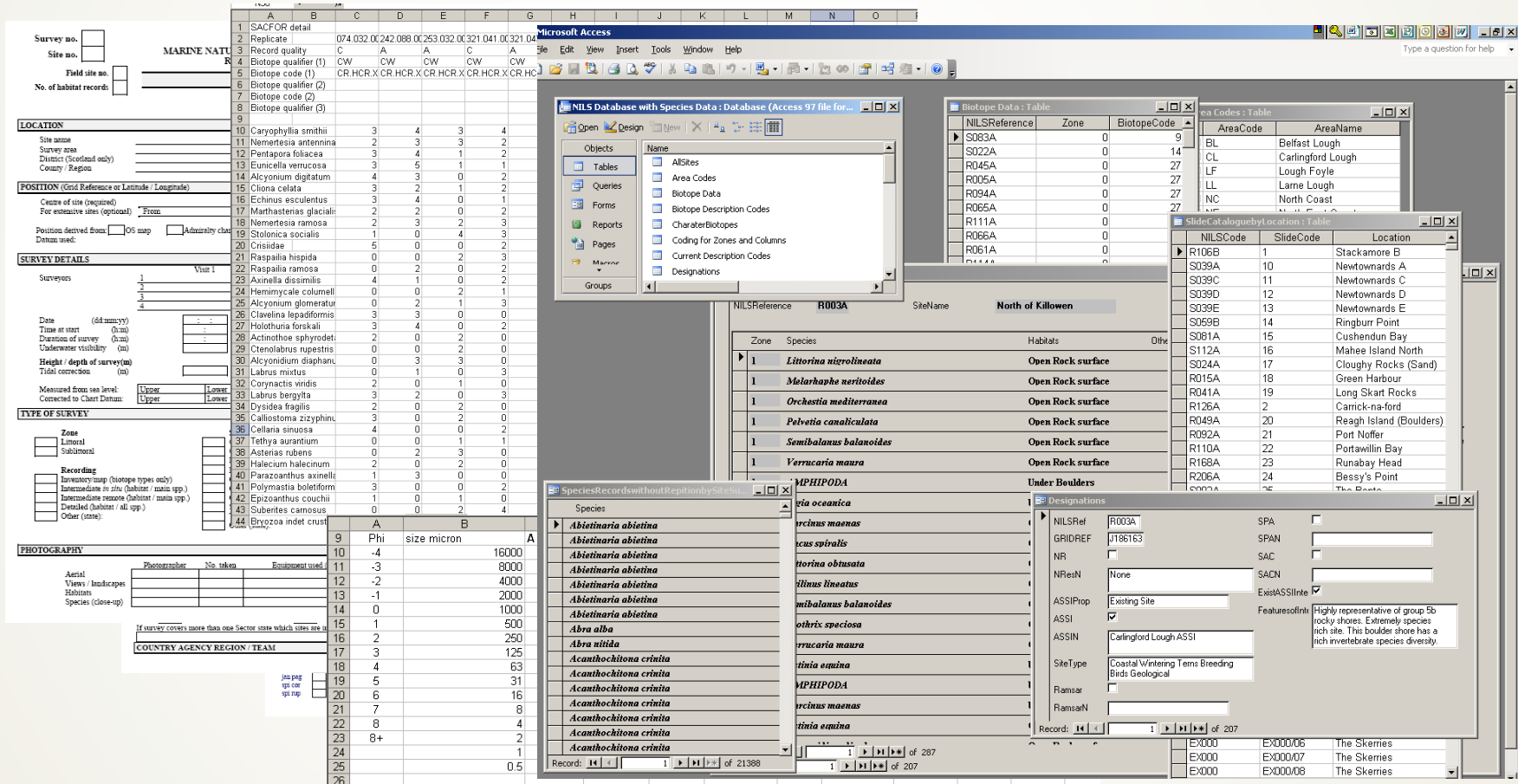
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O						
1 SACFOR detail																					
2 Replicate	074	032	00242	088	00253	032	00321	041	00321	047	00465	017	00639	009	00701	030/0	708	006	002	1	
3 Record quality	C	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
4 Biotope qualifier (1)	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	CW	
5 Biotope code (1)	CR	HCR	X	CR	HCR	X	CR	HCR	X	CR	HCR	X	CR	HCR	X	CR	HCR	X	CR	HCR	X
6 Biotope qualifier (2)																					
7 Biotope code (2)																					
8 Biotope qualifier (3)																					
9																					
10 Caryophyllia smithii	3	4	3	4	4	4	4	4	4	1	4										
11 Nematostella vectensis	2	3	3	2	2	2	2	4	5	4											
12 Pentapora foliacea	3	4	1	2	1	4	3	5	4												
13 Eunicella verrucosa	3	5	1	1	1	5	3	3	3												
14 Alcyonium digitatum	4	3	0	2	2	5	3	5	4												
15 Cliona celata	3	2	1	2	2	3	4	2	2												
16 Echinus esculentus	3	4	0	1	2	2	3	0	3												
17 Mathastieria glacialis	2	2	0	2	2	3	3	0	3												
18 Nematostella ramosa	2	3	2	3	2	0	0	1	3												
19 Stolonica socialis	1	0	4	3	2	4	0	0	4												
20 Oncididae	5	0	0	2	5	0	5	4	0												
21 Raspailia hispida	0	0	2	3	3	2	2	4	0												
22 Raspailia ramosa	0	2	0	2	2	2	4	0	3												
23 Axinella dissimilis	4	1	0	2	2	1	0	4	0												
24 Hemimycale columella	0	0	2	1	2	0	3	4	2												
25 Alcyonium glomeratum	0	2	1	3	3	3	2	0	0												
26 Clavelina lepadiformis	3	3	0	0	2	0	3	2	1												
27 Haliotuna forskalii	3	4	0	2	3	4	0	0	0												
28 Actinothoe sphyrodeta	2	0	2	0	2	2	0	2	3												
29 Ctenolabrus rupestris	0	0	2	0	0	3	4	4	2												
30 Alcyonium diaphanum	0	3	3	0	0	3	3	2	0												
31 Labrus mixtus	0	1	0	3	3	4	3	0	0												
32 Corynactis viridis	2	0	1	0	2	4	4	0	0												
33 Labrus bergylla	3	2	0	3	3	2	0	0	0												
34 Dysidea fragilis	2	0	2	0	2	0	3	3	0												
35 Calliostoma zephyrinum	3	0	2	0	0	1	2	3													
36 Cellaria sinuosa	4	0	0	2	2	5	0	0	0												
37 Tethya aurantium	0	0	1	1	1	2	3	0	0												
38 Astenias rubens	0	2	3	0	1	4	0	0	0												
39 Haliacium haliacium	2	0	2	0	0	2	0	3	0												
40 Parazoanthus axinellae	1	3	0	0	2	0	3	0	0												
41 Polymastia botleitformis	3	0	0	2	2	1	0	0	0												
42 Epizoanthus couchii	1	0	1	0	0	0	2	4	0												
43 Suberites canosus	0	0	2	4	4	0	0	0	0												
44 Bryozoa indet crust																					
9 Phi																					
10 -4	size	micron																			
11 -3	16000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.20	0.00						
12 -2	8000	25.90	0.00	0.40	16.30	0.60	0.00	0.40	29.80	0.00											
13 -1	4000	39.90	0.40	0.40	21.30	3.60	1.00	26.80	50.40	0.00											
14 0	2000	43.30	2.00	0.80	22.30	5.60	2.20	61.30	58.60	0.40											
15 0	1000	55.71	7.24	1.40	23.70	7.00	2.60	83.53	59.50	0.60											
16 1	500	64.17	13.66	1.40	48.31	7.74	2.60	90.73	60.23	0.94											
17 2	250	80.07	51.25	29.42	91.48	15.50	18.25	93.28	66.80	38.05											
18 3	125	95.53	92.78	91.71	96.13	73.19	86.71	95.98	87.40	92.52											
19 4	63	98.25	97.31	100.00	96.48	99.23	100.00	97.91	99.54	99.40											
20 5	31	98.51	97.60	100.00	97.94	99.27	100.00	98.39	99.79	99.40											
21 6	16	99.09	98.63	100.00	98.56	99.27	100.00	98.90	99.79	99.40											
22 7	8	99.42	99.06	100.00	99.37	99.27	100.00	99.30	99.79	99.44											
23 8	4	99.77	99.71	100.00	99.92	99.64	100.00	99.63	99.87	99.88											
24 8+	2	100.00	100.00	100.00	100.00	100.01	100.00	99.85	100.00	100.01											
25 1	100.00	100.00	100.00	100.00	100.00	100.01	100.00	99.96	100.00	100.01											
26 0.5	100.00	100.00	100.00	100.00	100.01	100.00	100.00	100.00	100.01	100.01											

Data forms

Spreadsheets

# What is DATA

Data: a collection of facts used for reference or analysis



The screenshot displays a Microsoft Access database interface with several open windows:

- Form:** A data entry form for 'MARINE NATU' with fields for Survey no., Site no., Field site no., No. of habitat records, LOCATION (Site name, Survey area, District, County), POSITION (Grid Reference or Latitude / Longitude), SURVEY DETAILS (Surveyors, Date, Time at start, Duration, Visibility, Height/depth, Tidal correction, Measured from sea level, Connected to Chart Datum), TYPE OF SURVEY (Zone, Litoral, Sublitoral, Recording), and PHOTOGRAPHY (Photographer, No. taken, Equipment used, Views, Inclusions, Habitats, Species).
- Table:** 'Biotope Data : Table' showing columns for NILSReference, Zone, and BiotopeCode.
- Table:** 'Area Codes : Table' showing columns for AreaCode and AreaName.
- Table:** 'Slide Catalogue by Location : Table' showing columns for NILSCode, SlideCode, and Location.
- Table:** 'SpeciesRecords without Reptition by Site' showing columns for Species and other attributes.
- Table:** 'Designations' showing columns for NILSRef, GRIDREF, NR, NRResN, ASSIProp, ASSI, ASSIN, SiteType, Ramsar, and RamsarN.
- Table:** A detailed species list for 'MPHIPODA' including *Abietinaria abietina*, *Abietinaria obtusata*, *Abietinaria abietina*, *Abra nitida*, *Acanthochitona crinita*, *Paraschoenus axonella*, *Polymastix bolleiformis*, *Epizoanthus couchii*, *Suberites carus*, and *Bryozoa indet crust*.

Data forms

Spreadsheets

Databases

# What is Information

---

**Information:** Information is data which has been processed within a context in order to give it meaning.

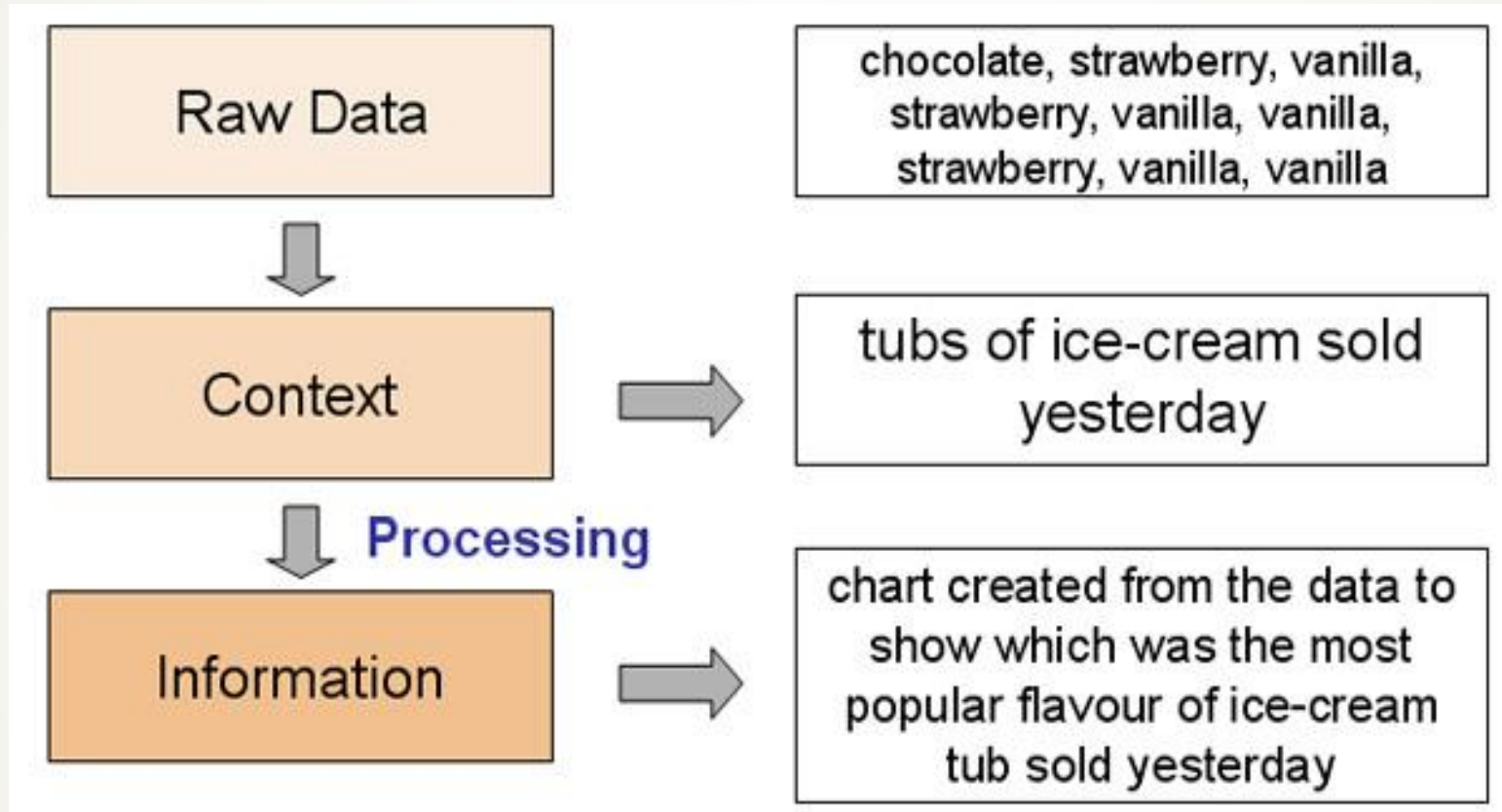


**Data** on its own has no meaning, only when interpreted by some kind of data processing system does it take on meaning and become **information**.

People or computers can find patterns in **data** to create **information**, and this information can be used to reason or make decisions.



# Information Example



The **processing** could take the form of creating a bar chart. This would instantly tell us which flavour sold the most yesterday.

# Data vs. Information

---

## Data

- raw facts
- no context
- just numbers and text

## Information

- data with context
- processed data
- value-added to data
  - summarized
  - organized
  - analyzed

# What is Knowledge

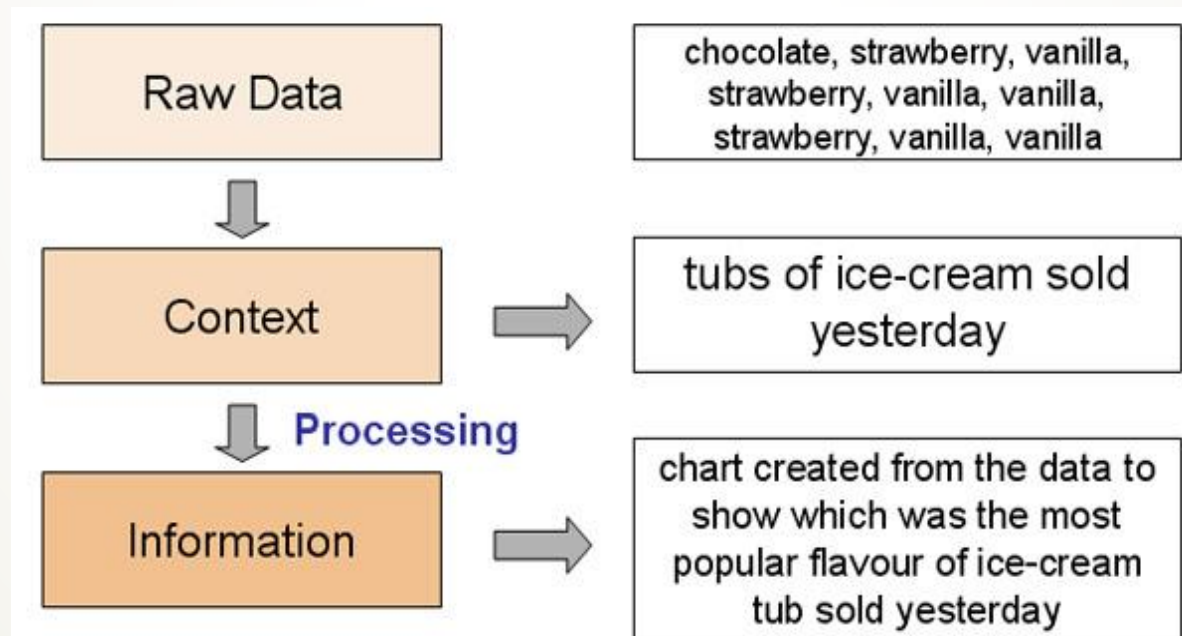
---

**Knowledge** is the understanding of rules needed to interpret information

*“...the capability of understanding the relationship between pieces of information and what to actually do with the information”*

# Knowledge Examples

A Marketing Manager could use this information to decide whether or not to raise or lower price ice-cream



# Knowledge Workers

---

- Knowledge workers have specialist knowledge that makes them “experts”
  - Based on formal and informal rules they have learned through training and experience
- Examples include **doctors, managers, librarians, scientists...**

# The Main Idea

