

UNIT 4 • SESSION 4

Data and Information Processing

Data is what you collect. Information is what it tells you. They are not the same thing.

2ND SEMESTER • SCHOOL YEAR 2026-2027

Introduction to Computing • Prof. Ryan V. Labrador

Saturday, 5 September 2026 • 2-hour lecture + 3-hour lab

Two hours, four parts

Part 1	Where data comes from	Sources, primary versus secondary, and the four kinds of measurement
Part 2	Three words	Data, information, knowledge — and the data types that carry them
Part 3	The cycle	Collect, verify, process, store, retrieve, present
Part 4	Structure and quality	Files and databases, tables and keys, and garbage in / garbage out
After	Laboratory	You survey this room, clean what you get, and publish the results

By the end of today you will be able to...

1

Name where data comes from, and tell primary data apart from secondary data.

2

Distinguish data, information and knowledge, using an example from your own life.

3

Classify a value into the correct data type, and say why the choice matters.

4

Walk through the six stages of the information processing cycle for a real system.

5

Explain when a spreadsheet is enough and when a database is required.

6

Identify records, fields and keys in a table, and spot six common data defects.

7

Choose the right measurement scale for a survey question, and say what you may do with it.

PART 1

Where data comes from

Before anything can be processed, somebody has to collect it. This afternoon, that somebody is you.

Six places data actually comes from



Transactions

Every GCash send, every jeepney tap, every receipt at the palengke stall that uses a POS. The biggest data source in the country by volume.



Forms and surveys

Somebody is asked and somebody answers. Enrolment forms, census, barangay records — and your lab this afternoon.



Sensors and meters

Rain gauges, weighbridges at the rice mill, electricity meters, the GPS in a phone. Nobody types these; they are read.



Logs

Machines writing down what they did. Web server logs, CCTV timestamps, attendance scans. Collected whether anyone looks or not.



Open and public data

PSA, PAGASA, DOST, LGU open-data portals. Already collected, already published, free to reuse — if you cite it.



People, directly

Interviews and observation. Slow, expensive, and the only way to find out why a number looks the way it does.

Card 4 is the uncomfortable one: most data about you was collected without anyone asking you.

Data you collected, and data you were handed

Primary data

- You collected it yourself, for your question
- You know exactly how, when and from whom
- You can go back and ask again
- Slow, and it costs you time
- This afternoon's survey is primary data

Secondary data

- Somebody else collected it, for their question
- You inherit their definitions and their mistakes
- You cannot go back and ask again
- Fast, cheap, often much larger
- A PSA table you download is secondary data

Neither is better. But you must know which one you are holding before you draw a conclusion from it.

What you may legally do with a number

Nominal

Names, no order

Course, barangay, phone brand. You may count them. You may not average them.

Ordinal

Order, uneven gaps

Strongly agree to strongly disagree. Rank them, yes. The gap between 4 and 5 is not the gap between 1 and 2.

Interval

Even gaps, no true zero

Temperature, calendar years. Differences are meaningful; ratios are not. 2026 is not twice 1013.

Ratio

Even gaps, true zero

Pesos, minutes, kilograms, headcount. Everything is allowed, including 'twice as much'.

Design your survey question knowing which of these four you will get back. Choose wrong and no amount of maths saves it.

PART 2

Three words

Data. Information. Knowledge. People use them interchangeably. They should not.

Data, information, knowledge



Data

Raw facts, unprocessed. "8". "Ilagan". "2026-08-15". Each one is true, and on its own each one answers nothing at all.



Information

Data placed in context and processed. "8 students in Group Hopper attended the session in Ilagan on 15 August." Now it answers a question somebody asked.



Knowledge

Information combined with experience and pattern. "Attendance drops the Saturday after payday." Now it changes what you decide to do.

Data is collected. Information is produced. Knowledge is earned.

The same facts, three rungs up



Nothing new was collected between the bars. Every step up is work somebody did on what was already there.

Six numbers, and what they become

Data

3, 5, 2, 4, 5, 1

Six numbers. Nobody in this room can act on them.

Context

Hours studied last week, per group

Now they measure something. This is where data starts to mean anything.

Information

Average 3.3 hours. Range 1 to 5.

A summary a person can hold in their head and repeat to somebody else.

Knowledge

The spread is the problem, not the average

Some groups are doing five times the work of others. That changes what you do next.

The average hid the problem. Most bad decisions come from stopping one row too early.

Data types



Text

Names, addresses, comments. Anything you will never do arithmetic on.



Whole number

Counts you can add up: attendance, quantity, age in completed years.



Decimal

Measurements and money: 3.25 hours, 149.50 pesos, 12.7 hectares.



Date and time

15 August 2026, 09:30. Its own type, so a machine can sort and subtract them.



Boolean

Two values only. Yes or no. Paid or unpaid. Present or absent.



Media

Images, audio, video. Usually stored as files, with only the location in the table.

A mobile number is TEXT, not a number. Store 09171234567 as a number and the leading zero disappears.

PART 3

The processing cycle

The same six stages, in every system, in every industry.

The information processing cycle

1 Collect

Get it in. Forms, sensors, scans, imports. Most errors are born right here.

2 Verify

Check before you trust. Complete? In range? Already recorded once?

3 Process

Sort, count, compute, compare, combine. Turning records into an answer.

4 Store

Keep it safely and be able to find it again. Structure, and a backup.

5 Retrieve

Ask for exactly the part you need — not everything you happen to hold.

6 Present

Report, chart, page. If nobody can read it, it is not information yet.

Stage 2 is the one organisations skip. Every disaster in this course starts there.

It does not stop at 'present'



Stage 2 is the one organisations skip. Every data disaster in this course starts there.

Today's lab is stages 1, 2, 3 and 6. Your group will feel exactly where the time goes.

A barangay rice farmer registry

Collect

A form at the barangay hall

Name, farm size, crop, contact number. Handwritten first, typed later.

Verify

Two Juan Dela Cruzes

Same name, different farms. Somebody must decide: one person, or two?

Process

Total hectares per barangay

Now the municipal office can plan how much seed to distribute.

Store

One file, and a backup

Without the second copy, the whole registry lives on one laptop.

Present

A one-page summary

The mayor reads one page, not four hundred handwritten forms.

PART 4

Structure

Where data lives, and how a machine finds one row among millions.

Files and databases

Files

- One document, one purpose, one owner
- Easy to start — anyone can make a spreadsheet
- Two people editing means two versions
- Finding one record means opening everything
- Fine up to a few hundred rows

Databases

- Structured tables with declared data types
- Rules that stop bad data getting in
- Many people at once, safely, on one copy
- Ask for exactly the rows you want
- Built for millions of rows

A spreadsheet is not a mistake. It is a starting point that eventually stops fitting.

Tables, records and fields

TABLE: students

STUDENT_NO	LAST_NAME	FIRST_NAME	COURSE	YEAR
2026-0141	Bumanglag	Andrea	BSIS	1
2026-0142	Dela Cruz	Juan	BSIS	1
2026-0143	Sarmiento	Marilou	BSIS	1

A RECORD is one row — one student.

A FIELD is one column — one kind of fact.

A database is many tables that know about each other. That is the rest of your degree.

Keys

TABLE students

student_no	last_name	first_name
2026-0141	Bumanglag	Andrea
2026-0142	Dela Cruz	Juan
2026-0143	Sarmiento	Marilou

the same value links them

TABLE enrolments

enrol_id	student_no	subject
E-001	2026-0142	IT 101
E-002	2026-0141	IT 101

PRIMARY KEY

One field that is different for every single record in the table.

NOT THE NAME

Names repeat and names change. There are many Juan Dela Cruzes.

FOREIGN KEY

A field that points at another table's primary key.

Choosing the primary key is the first real design decision in any information system.

DATA QUALITY

Garbage in, garbage out

No amount of processing repairs data that was wrong when it was collected.

Six defects you will meet this afternoon



Missing

Blank fields. The most common defect, and the easiest to prevent at the form.



Duplicate

The same person recorded twice, slightly differently. Now every count is wrong.



Inconsistent

"Ilagan", "ilagan", "Ilagan City", "Ilagan, Isabela". One place, four spellings.



Wrong type

A date typed into a text box. It looks fine, and it will not sort correctly.



Out of range

An age of 213. Nobody checked, so it is now inside your average.



Outdated

Correct in 2019, wrong today, and nothing on the record says which.

You will find at least three of these in your own survey data today. Every group does.

Spot what is wrong here

NAME	TOWN	BIRTHDATE	MOBILE
Dela Cruz, Juan	Ilagan	2007-03-04	09171234567
DELA CRUZ, JUAN	ilagan city	04/03/2007	09171234567
Bumanglag, Andrea	Ilagan		9171234568
Sarmiento, Marilou	Cauayan	1907-11-02	09181234569

Duplicate. Inconsistent spelling. Two date formats. A missing value. A lost leading zero. A birth year of 1907.

Four rows. Six defects. Now imagine four hundred rows collected by forty people.

Five questions to ask of any dataset

Accuracy

Is it right?

Does the value match reality? A neat, well-formatted table can be confidently wrong.

Completeness

Is any of it missing?

Blank fields quietly change every total and every average you compute.

Consistency

Does it agree with itself?

The same fact should not appear in two different forms in the same column.

Timeliness

Is it current enough?

Correct last year is simply a wrong answer today, and nothing flags it.

Uniqueness

Is anything counted twice?

Duplicates inflate every number you report, and they are hard to see.

What an information system is made of



People

The ones who enter it, use it and depend on it. Always the hardest part.



Procedures

The agreed steps. Who enters what, when, and who is allowed to change it.



Data

The records themselves. The part that is expensive to collect and easy to ruin.



Software

The programs that capture, process and present it.



Hardware

The machines it runs on — including the one serving your group's website.

You met these five in Session 2. An information system is those same components, organised around a purpose.

Four kinds of system, four kinds of question

TPS	Transaction processing	Records what happened: an enrolment, a sale, an MRT tap. High volume, must never lose one.
MIS	Management information	Summarises transactions into reports. How many enrolled, by course, by week.
DSS	Decision support	Answers "what if". Models the options before anybody commits money.
ESS	Executive support	A handful of indicators for the person who signs. One screen, no detail.

Each layer feeds the one above it. You cannot report what you never recorded.

Why this is your degree, not a side topic

The name

Information Systems

Not computer science, not IT support. Systems that turn data into decisions.

The position

You sit in between

Between the people who have the problem and the machines that could solve it.

The skill

Asking the right question

Any system can produce numbers. Very few produce the number somebody actually needed.

Today

You already do it

Your lab this afternoon is collect, clean, process, publish. That is the whole cycle.

First-year students are usually told what their degree is. You get to do it in week four.

Four things from today



One

Data is not information

Data becomes information when context and processing make it answer a question.



Two

Six stages

Collect, verify, process, store, retrieve, present. In that order, every time.



Three

Garbage in, garbage out

No amount of processing repairs data that was already wrong when collected.



Four

Structure has names

Tables, records, fields, keys. Learn the four words; you will use them for years.

If you remember one thing: verify at collection. Everything downstream depends on it.

In the laboratory today



Collect

Survey this room

Your group designs one question and collects real answers from real classmates.



Clean

Fix what you collected

You will find blanks and duplicates. Every group will. That is the lesson, not the failure.



Publish

A route of your own

/results live on your group site, showing data your own group gathered today.



AI task

Judge its advice

Describe a messy dataset to an AI and decide whether its cleaning advice is safe to follow.

Group leaders: bring the USB stick, and know your group's web address by heart.

PROGRAMMING LAB ACTIVITY

This afternoon

Collect it, clean it, publish it

What you are building today



Together

The group page

A /results page carrying your group's own survey — the question you asked, the counts after cleaning, each answer as a share, and an honest line about what you threw away.



On your own

Your own version — this is the marked one

Your data card: your name, your town, and one number you measured about yourself yesterday, turned into a sentence.



To match

Your page.html twin

The same thing again, typed out by hand as HTML.

WHERE IT GOES LIVE

`gNN.cic.rmlsystems.net/results`

`gNN.cic.rmlsystems.net/<yourfolder>/`

`gNN.cic.rmlsystems.net/<yourfolder>/
page.html`

Eight identical pages is eight zeros. Your version has to be yours — your data, your words.

The Python you have not used before

```
town = "Ilagan"
```

a variable — a box with a name, holding one thing

```
print(town)
```

shows what is in the box, while you are working

```
f"{walk} of {total}"
```

an f-string; whatever is in braces is worked out first

```
f"{share:.1f}%"
```

the same, to one digit after the point

at the shell

```
walk = 23
```

```
tricycle = 31
```

```
total = walk + tricycle
```

```
share = walk / total * 100
```

```
print(f"{walk} of {total} walk - that is {share:.1f}%")
```

It prints: 23 of 54 walk - that is 42.6%

How this is marked

- | | | |
|---|----------------------------------|---|
| 6 | It runs | Your address loads and there is no traceback on it. |
| 6 | It does this week's thing | Real variables and real f-strings. The percentage is worked out, never typed. |
| 4 | It is yours | Your own number, your own town, your own sentence. |
| 2 | The twin matches | page.html shows the same thing, by hand. |
| 2 | Worksheet complete | Including the AI sentence and the twin count. |

DUE Sunday 6 September, 11:59pm · the full guide is at lab.cic.rmlsystems.net

Anyone can collect numbers.

Being trusted with them is a different job. That is the one you are training for — and this afternoon eighty people in this room hand you their answers.

Questions?