Data, information and lifecycle

This morning you probably unlocked your phone, checked the bus time and paid for a coffee with a card. Each of those moments left a small trace somewhere: a timestamp, a card number, a bus stop code. None of those traces means much on its own. So why do companies care so much about keeping them?

Think of a hotel in Calvià in July. It receives thousands of bookings: names, dates, room numbers, prices. Stored in a database, they are just rows. But the moment somebody asks "is this July better or worse than last July?", those rows turn into something the manager can act on: hire more staff, change prices, open the second restaurant. Rows become decisions. This lesson is about that jump, and about the path every piece of data follows from the day it is captured to the day it should disappear.

What you will learn

- Tell **data** from **information**, and say what turns one into the other
- Place data, information, knowledge and decisions on the **DIKW pyramid**
- Describe the six stages of the **data lifecycle** and recognise them in a real company
- Find the data a real company says it collects by reading its **privacy policy**
- Avoid the trap: a full database is not the same as an informed company

Data and information

> A **datum** (plural **data**) is a raw fact, value or symbol with no context. **Information** is data that have been processed and put in context so that they mean something to somebody.

Write the number **38.5** on the board. What does it mean? A fever? A shoe size? A price? Nobody knows. Now add context, one piece at a time: "38.5 °C", "temperature of the server room", "today at 3 a.m.", "the manufacturer's limit is 35 °C". Same number. Now it is an alarm.

	Data	Information
What it is	A raw fact, value or symbol	Data processed and put in context
Answers	Nothing on its own	Who, what, when, where, how much
Hotel example	"214", "2027-07-14", "2"	"Room 214 is booked for 2 people from 14 July"
Good for	Being stored and processed	Making decisions

In the city-builder game, data are every single reading coming from every sensor in the city. Information is the **info-view overlay**: the colour map that tells you, at a glance, that the north district is jammed. The game does the processing for you. In a real company, somebody has to program it. That somebody is often a DAM developer.

Careful: information is not a different kind of file

A spreadsheet full of numbers is not information yet. Information appears when somebody, or some program, asks the data a question and gets a useful answer. The same datum can give different information to different people: 38.5 means one thing to a doctor and another to a system administrator.

The DIKW pyramid

A classic way to draw the climb is the **DIKW pyramid**: data → information → knowledge → wisdom. It is usually credited to the systems thinker Russell Ackoff (1989). It is a teaching model, not a law: different authors define the levels differently. Use it as a ladder, not as a definition to memorise.

Level	Hotel example	What was added
Data	"214 booked", "215 free"	Nothing: raw records
Information	"Occupancy this week is lower than the same week last year"	Data organised and compared
Knowledge	"When there is a congress in Palma, weekday occupancy goes up"	A pattern learned from a lot of information
Wisdom (decision)	"On congress weeks, raise prices and add staff"	Somebody acting on the knowledge

The data lifecycle

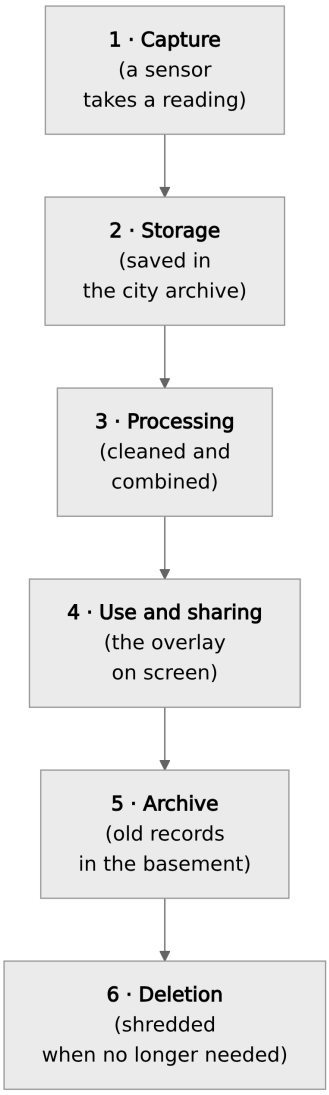
Data are not eternal. They are born, they are used, and at some point they should die.

The **data lifecycle** is the sequence of stages a piece of data goes through, from the moment it is captured or created until it is archived or deleted.

Stage	What happens	At the hotel in Calvià
1 · Capture or creation	Data enter the system: forms, sensors, logs, imports	The guest fills in the booking form; the electronic lock records every opening
2 · Storage	Data are saved somewhere: a database, files, the cloud, with backups	The booking is saved in the hotel management system
3 · Processing	Data are validated, cleaned, transformed and combined	Availability is checked, the price is calculated, a room is assigned
4 · Use and sharing	Data are analysed, shown and sent to whoever needs them	Reception sees today's arrivals; management reads the occupancy report
5 · Archive	Data are no longer used every day but are kept, for legal reasons or history	Bookings from past seasons go to an archive with restricted access
6 · Deletion	Data are erased or anonymised safely when there is no reason to keep them	At the end of the retention period, personal data are deleted or anonymised

There is no single official model: depending on the source you will find five, six or eight stages with slightly different names. What matters is the **order** and being able to give an example of each stage.

Careful: the stage everybody forgets



> The lifecycle does not end when data are stored. Stage 6 is the one companies forget most, and it matters: data you no longer need still cost money to store and can still be stolen. Each stage also has its own security risk (a form that leaks, a backup next to the original, an archive anybody can open). You will come back to this in Cybersecurity and regulation.

Where to find a company's real data

You cannot see inside your project company's database. But there is a public document that has to say, by law, which personal data the company collects, what for, who receives them and how long they are kept: its **privacy policy** (*política de privacidad*). The European data-protection regulation (RGPD, Articles 13 and 14) requires companies to give people that information when they collect their data.

Usually in a privacy policy	Which lifecycle stage it tells you about
Which data are collected (name, email, booking dates, card…) and how (forms, cookies, app)	1 · Capture
Purposes: what the data are used for	3 · Processing and 4 · Use
Recipients: other companies that receive the data (payment, cloud, marketing)	4 · Sharing
International transfers: whether data leave the European Union	2 · Storage
Retention period: how long the data are kept	5 · Archive and 6 · Deletion
Your rights and how to exercise them	All stages

In class: inventory and journey of one dataset

Work **individually**, on your group's project company. Use the company's own website. **No AI tools** in this activity.

Step	Do this	You should see
1	Open the company's website and find the privacy policy (usually a link in the footer: *Privacidad*, *Aviso legal*, *Privacy*). Copy its URL.	A long legal page, often in Spanish
2	Write a table with **at least five kinds of data** the company handles. For each one: what it is, where it comes from, whether it is **personal** data, and which information or decision it allows. Mark each row as *found* (with the source) or *hypothesis*.	Some rows found in the policy, some hypotheses about the OT side (sensor readings, machine logs)
3	Pick the **most important dataset** (bookings, orders, water-meter readings…) and follow it through the six stages: who captures it, where it is stored, how it is processed, who uses it, when it is archived, when it should be deleted.	A six-row table
4	Mark the stage where the data become **information** and for which decision. Mark the stage with the **highest risk** and say why.	Two short sentences
5	Hand in the document on Classroom with the privacy-policy URL and the date you read it.	

If the company has no privacy policy, or it does not list retention periods, write that down: it is a finding too.

For your project

This activity is the first draft of **section A (data inventory and lifecycle)** of Deliverable 5. Whoever owns section A in your group will build on everybody's tables, so keep yours in the group's shared folder. Remember the rule: what you found, you cite; what you could not find, you label *hypothesis*.

Key takeaways

- A **datum** is a raw fact; **information** is data processed and put in context to answer a question or support a decision.
- The **DIKW pyramid** (data → information → knowledge → wisdom) is a useful teaching ladder, not an exact definition.
- The **data lifecycle** has six stages: **capture**, **storage**, **processing**, **use and sharing**, **archive** and **deletion**.

- The last stage, deletion, is the one companies forget most, and every stage carries its own security risk.
- A company's **privacy policy** is a public, legally required source that tells you which personal data it collects, why, who receives them and for how long.

Quick check

1 Which of these is information, not just data?

A `07014`
B `2027-03-01T08:15:00Z`
C "60% of this month's incidents were opened on a Monday"
D `404`

2 What turns data into information?

A Saving them in a database
B Processing them and putting them in context so they answer a question
C Collecting a very large amount of them
D Converting them to a PDF

3 Which is the correct order of the data lifecycle?

A Storage → capture → use → processing → deletion → archive
B Capture → storage → processing → use and sharing → archive → deletion
C Capture → processing → deletion → storage → use → archive
D Use → capture → storage → archive → processing → deletion

4 A hotel keeps the bookings of past seasons in a restricted archive because the law requires it. Which stage is that?

A Processing
B Use and sharing
C Archive
D Capture

5 Where can you find, legally and publicly, which personal data a company collects and how long it keeps them?

A In its internal database, by asking an employee for the password
B In its privacy policy, which the RGPD requires it to provide
C Nowhere: companies never publish that
D Only in its annual accounts

6 In the city-builder analogy, what is the info-view overlay that shows the jammed north district?

A Raw data
B Information
C A backup
D The deletion stage

7 A classmate says: “Our company has a huge database, so it is already well informed.” Pick the two answers that are true.

- **A** A full database is not information until somebody asks it a question and gets a useful answer
- **B** Correct: the more rows, the better the decisions
- **C** Data that are kept with no purpose still cost money to store and can still be stolen
- **D** Data can never be deleted once they are in a database

You can now follow one piece of data from birth to deletion. Next: what happens when the data are so many, so fast and so varied that a normal database is not enough? Big data.

Solutions

1. C · The other three are raw values with no context: a code, a timestamp, a number. The sentence about incidents is data that have been **grouped, counted and put in context**, so it can support a decision.

2. B · Storing or accumulating data does not make them meaningful. **Information** appears when the data are processed and given context for somebody who needs to decide.

3. B · Data are first **captured**, then **stored**, **processed**, **used and shared**, then **archived** when they are no longer used every day, and finally **deleted** or anonymised.

4. C · The data are not used day to day any more, but there is a reason to keep them: that is the **archive** stage. Deletion comes when that reason ends.

5. B · The **privacy policy** must tell people which data are collected, for which purposes, who receives them and the retention period. It is the right public source for your data inventory.

6. B · The overlay is what you get after the game **processes** thousands of sensor readings (data) and draws them in context. That is **information**: ready for a decision.

7. A, C · Same lesson in a new situation. **Accumulating data** is not the same as being informed, and data with no purpose are a cost and a **risk**. That is why the lifecycle ends with deletion.