

Data Management

Version: 1.0



Learning intentions

We will be learning about how organisations care for their data, specifically

- **What data management is** and the various aspects of it
- Why it is **important to manage data** within an organisation
- What **data management activities** organisations need to undertake
- What the consequences of ***not* managing data** can be

Background

In the **Caring for Data** lesson you learned that data is a resource that needs to be cared for in the same way as any other valuable resource within an organisation.

You also learned that data that is not cared for can become less valuable, or even cause problems.

The discipline of caring for data is called **data management**. There are a number of important benefits to organisations that manage their data...



Why is it important to manage data?



To help **make better decisions**



To **improve data security**



To **improve productivity and efficiency**



To **increase trust** in the data by **minimising potential errors** in it

Definition

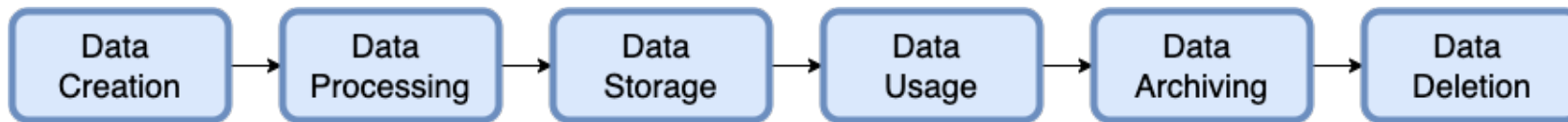


Data Management

The activities involved in treating data as a valuable resource through its entire lifecycle

When is it important to manage data?

Data has a lifecycle. Here is an example.



The NHS is often referred to as caring for the health of individuals from 'cradle to grave' (i.e., throughout their whole lifetime).

In a similar way, data needs to be cared for throughout its lifecycle, from creation to deletion.

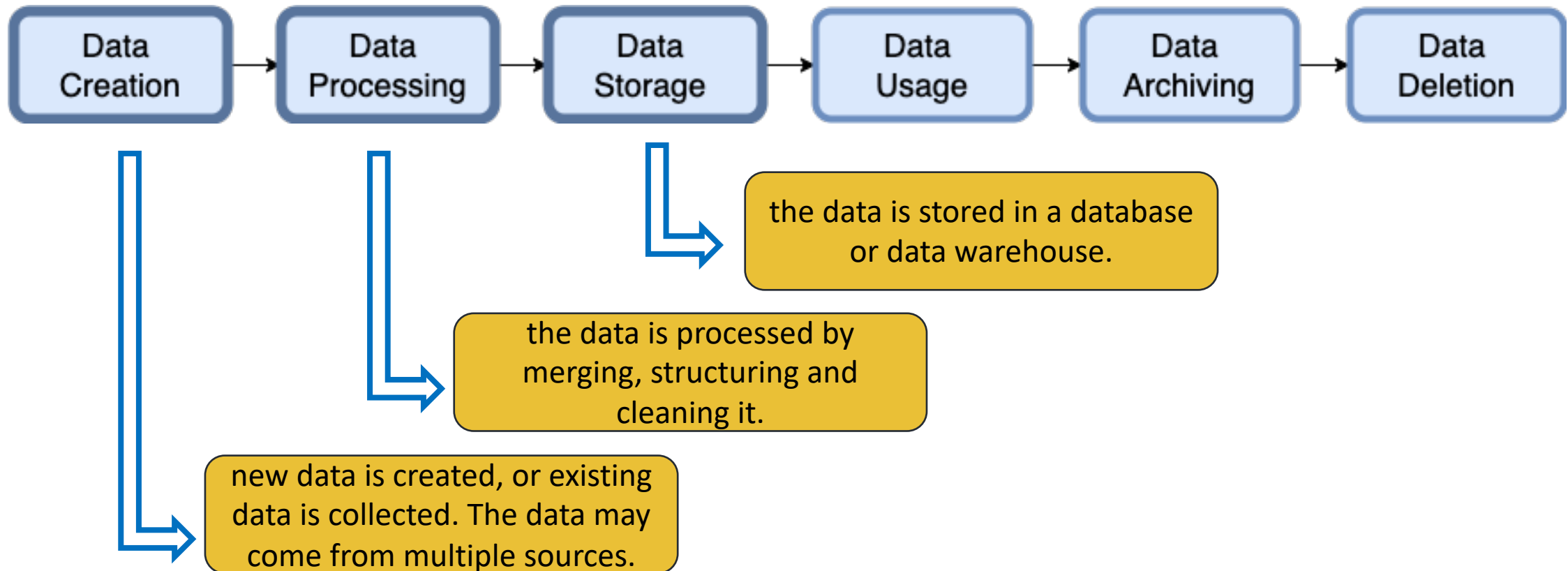
Data Lifecycle

The sequence of steps that data goes through from its creation to its deletion

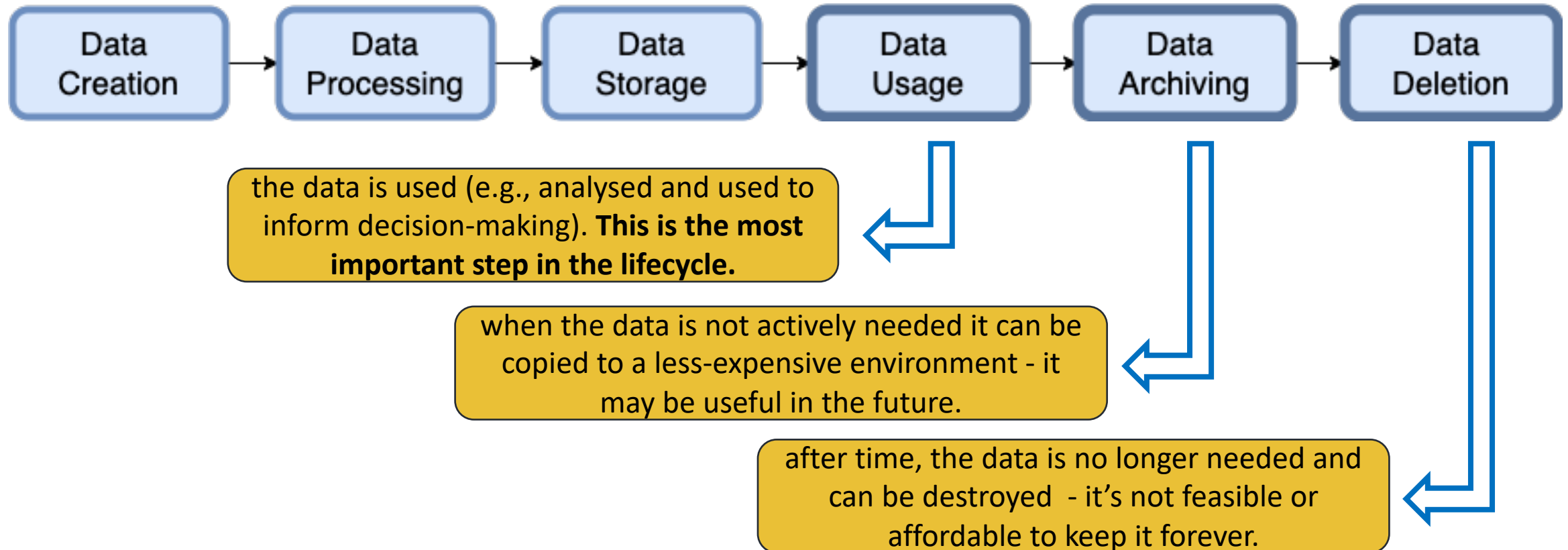
Show me....



Here is what happens to the data at each step.



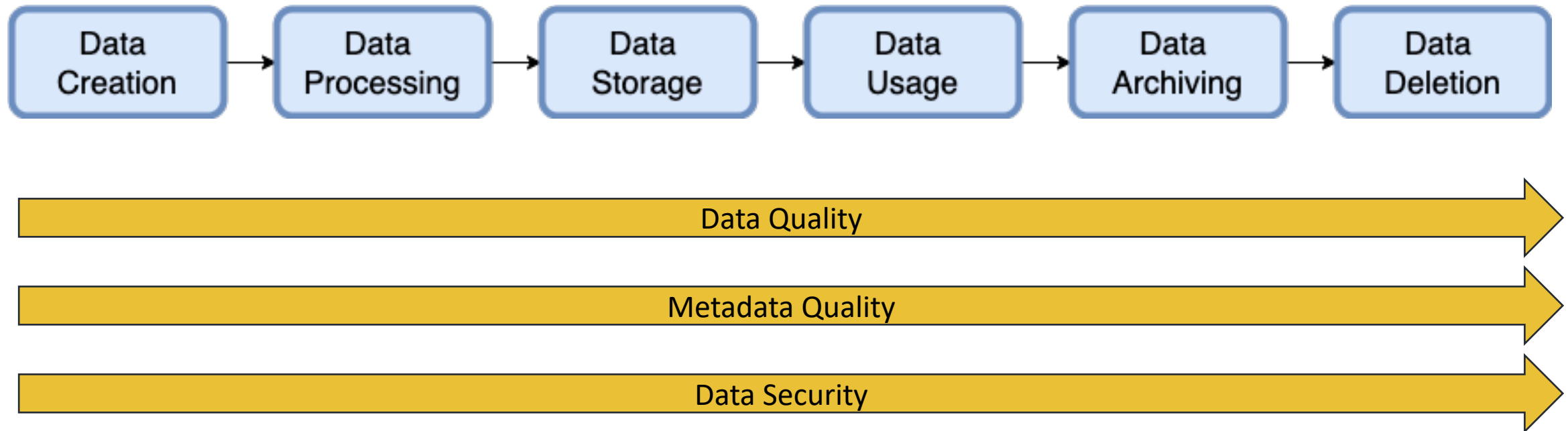
Show me....



Show me....



There are some data management activities that are important **throughout the data lifecycle**.



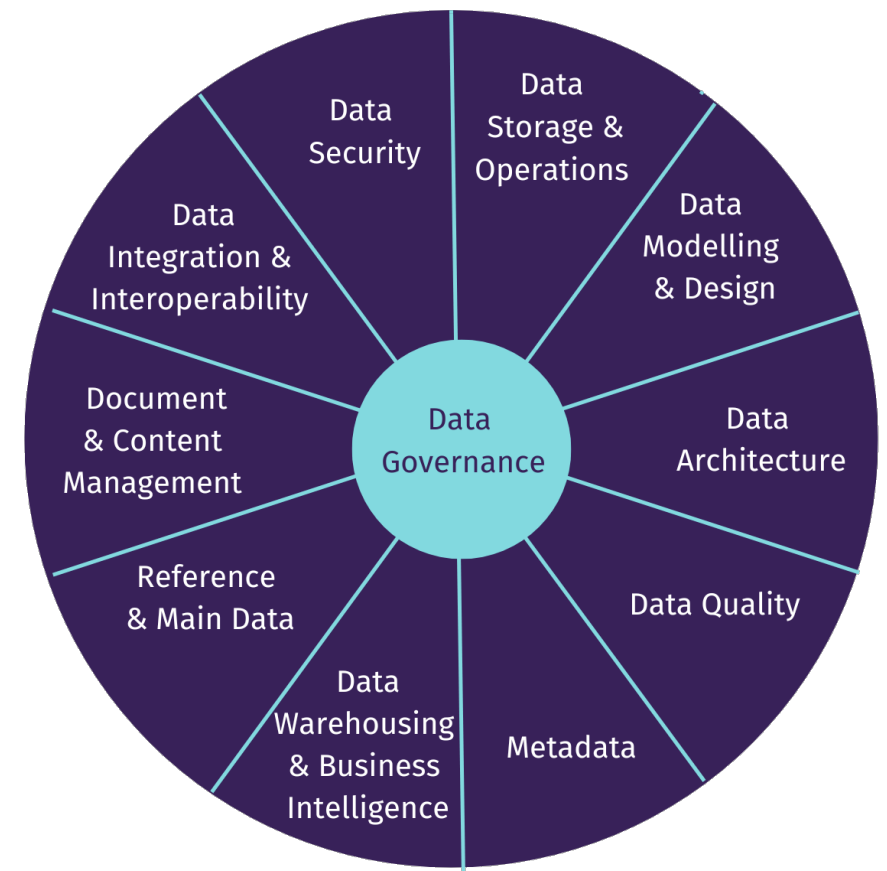
Data Management Knowledge Areas

www.dama.org

There are frameworks which help people use data management in an organisation. One of the most popular is the DAMA (International Data Management Association) framework.

It divides data management into the knowledge areas shown in the diagram, known as the DAMA Wheel.

Another popular framework is DCAM (The Data Management Capability Assessment Model).

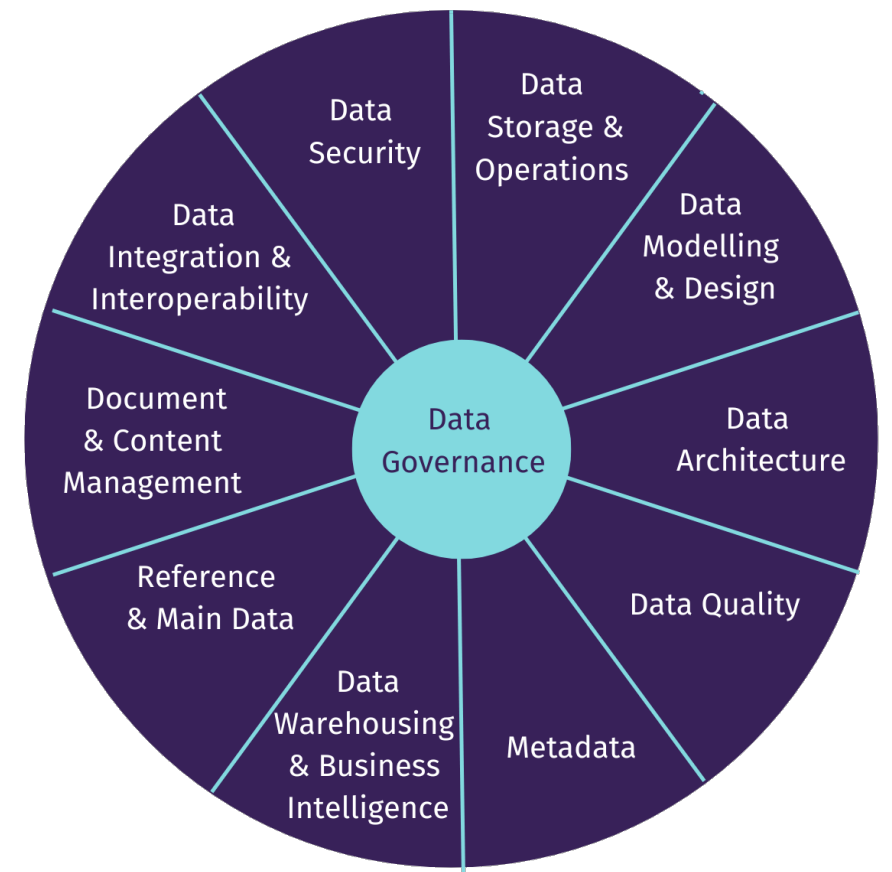


Data Management Knowledge Areas

To help you understand data management, you'll discover more about each of these knowledge areas in the rest of this lesson. You'll find out what they are and why they are important.

Data governance is at the centre of the diagram for a reason. It's there because it provides the foundation for the other areas.

Because it has such an important role to play, it's the perfect place to start.



Data Governance

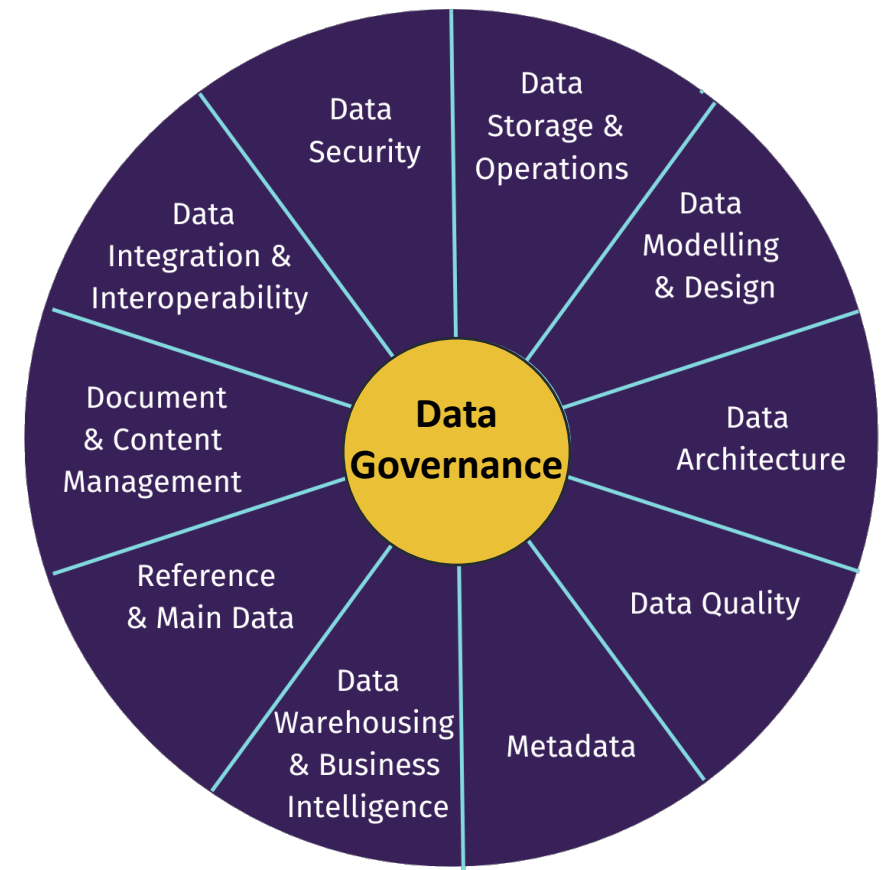


What is it?

Data governance is all about providing **guidance and direction** that helps an organisation take care of its data in way that **helps it succeed**.

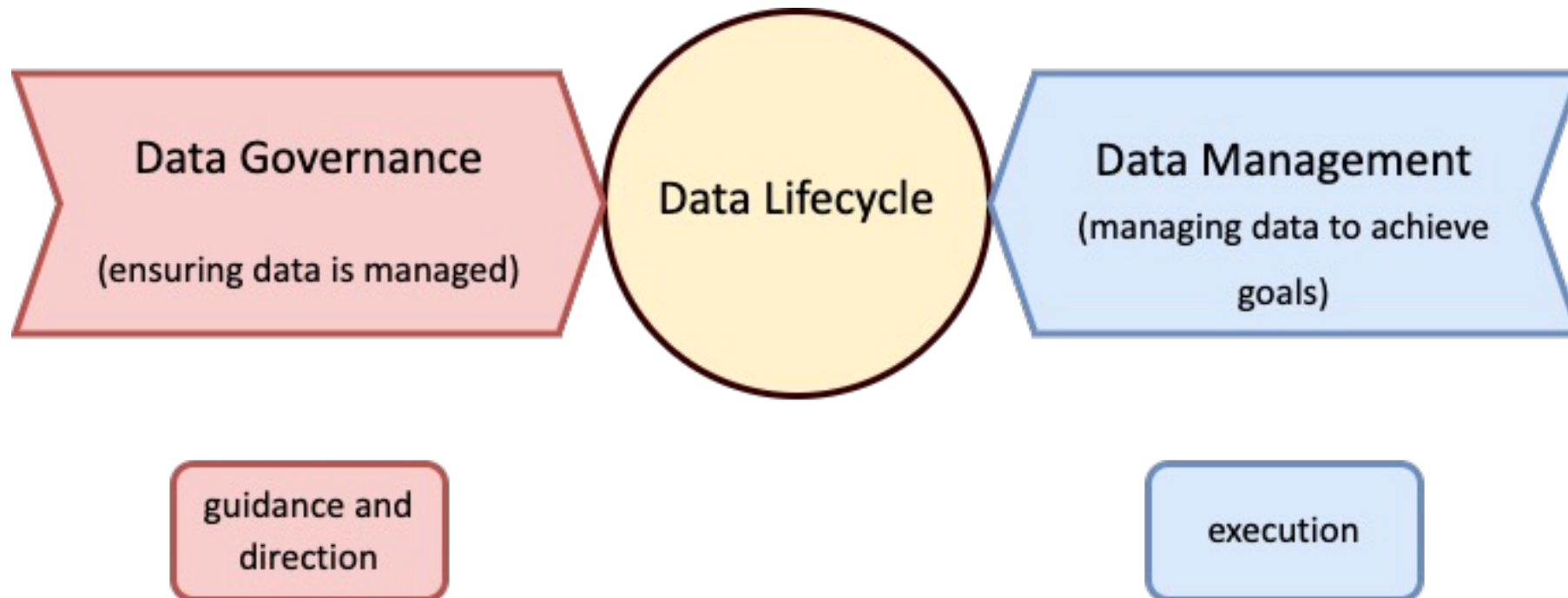
Why do it?

By guiding the other areas of data management, data governance helps organisations increase the value they get from their data.



Data Governance

The role of *data governance* is to **make sure that data is properly managed throughout the lifecycle**, whereas *data management* is about **directly managing the data**.



Next steps

Complete **questions 1 to 4**
in **section 1** of the
'Data Management' workbook.

Data Security



What is it?

Protecting data from unauthorised access or loss, to make sure the right people can access and use data in the right way, and that others can't.

Why do it?

This is needed to make sure an organisation maintains its good reputation and to protect it from the risk of financial losses. It is also needed to satisfy privacy and confidentiality laws e.g., GDPR.



Example



You can find out about some examples of where data hasn't been protected from unauthorised access or loss in the **Data Breaches** section of the **Keeping Organisational Data Secure** lesson.

This includes the top five worst data breaches ever.



Data quality



What is it?

Ensuring that the quality of the data is fit for the purpose it is being used for.

Why do it?

High-quality data is valuable because it allows organisations to **make informed decisions**.

Low-quality (e.g., inaccurate, incomplete or out-of-date) data can **present risks** e.g., bad decisions, poor customer experience or reputational damage.



Example



You can find some examples of where poor quality data resulted in serious consequences in the **Importance of Data Quality** lesson.

This includes a cryptocurrency company who accidentally gave away \$10.7 million.



Document and Content Management



What is it?

Documents contain **unstructured or semi-structured** data (i.e., content). Like structured data, this data needs to be managed. This area covers the protection, storage, retrieval and disposal of documents.

Why do it?

Typically, organisations have large volumes of data stored within documents. Managing these documents enables people to locate, access and share them more efficiently and reduces the risk they will become lost.

emails

word-processing documents

printed documents

chats

websites

Structured data Ordered in rows and columns (tabular) with clearly defined data types

Semi-structured data Uses markers to organise and describe the structure giving it flexibility. Data can vary in order and length.

Unstructured data No pre-defined structure, typically contains lots of text

Example



Here's a fictitious example to highlight the consequences of document and content management being done badly.

You join In2Space, a new start-up spacecraft manufacturing company, as a software engineering manager. Launch date is 12 months from now.

Your job is to lead the team of 30 software engineers who are writing the code for the mission- (and life!) critical lunar-landing module.



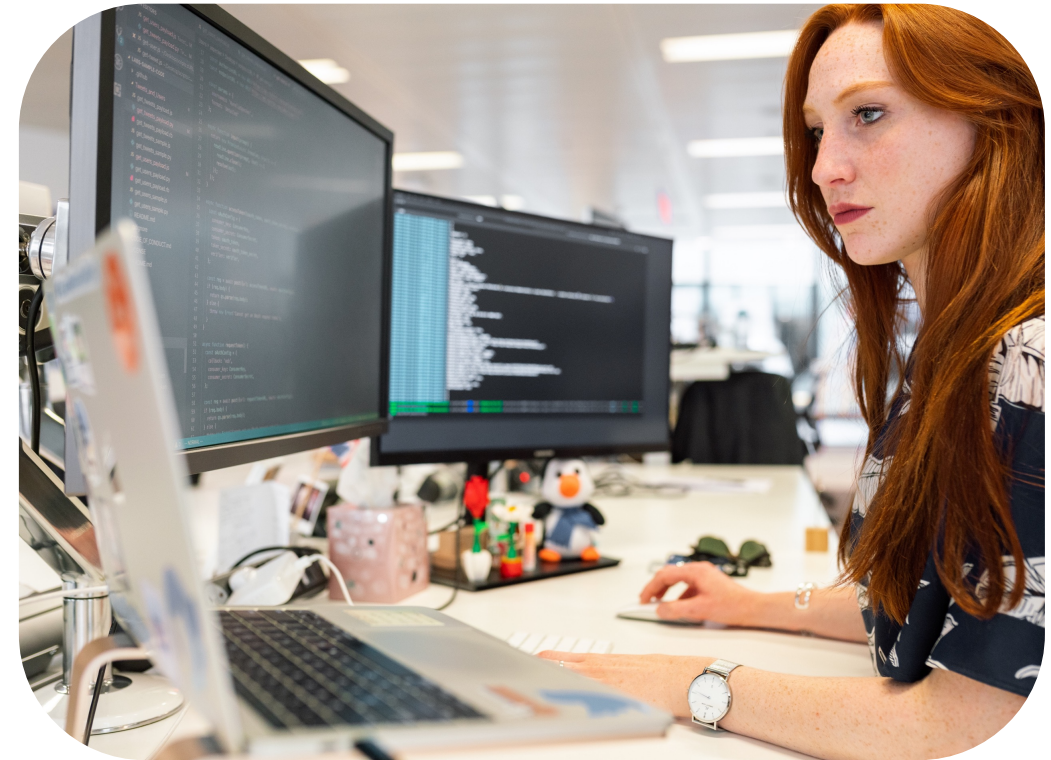
Example



The team who were previously developing this module have used over **200 specification documents** which provide the requirements for this module. Unfortunately, this team have all now left the company.

Your new team of gets to work on development. After 2 months, you realise that **something is wrong**.

Due to poor document management, your team have been using old versions of the documentation.



Your turn....



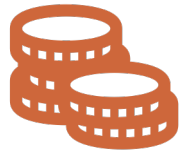
What might the consequences be for In2Space?

Some aspects to consider are:

- costs
- productivity and efficiency
- safety
- staff morale
- reputation



Your turn....



financial costs – a large team have been paid to develop the wrong thing



productivity and efficiency – time is limited and hasn't been used well. Will the launch be need to be delayed?



safety – the safety of the users of the spacecraft could be at risk



staff morale – you need your engineers to be positive and focused



reputation – what will others in the business think of your team?

Next steps

Complete **questions 1 to 3**
in **section 2** of the
'Data Management' workbook.

Data Architecture



What is it?

An architect creates a plan for a building, such as a house. No-one would build a house *without* a plan. The plan helps make sure the building meets the owner's needs (e.g., doesn't fall down and keeps them dry). It's similar to data architecture. Data architecture is **the plan for the data management resources that an organisation should build** to meet its needs.

Why do it?

Data architecture is needed to translate business needs into data and system requirements to fulfil these needs.

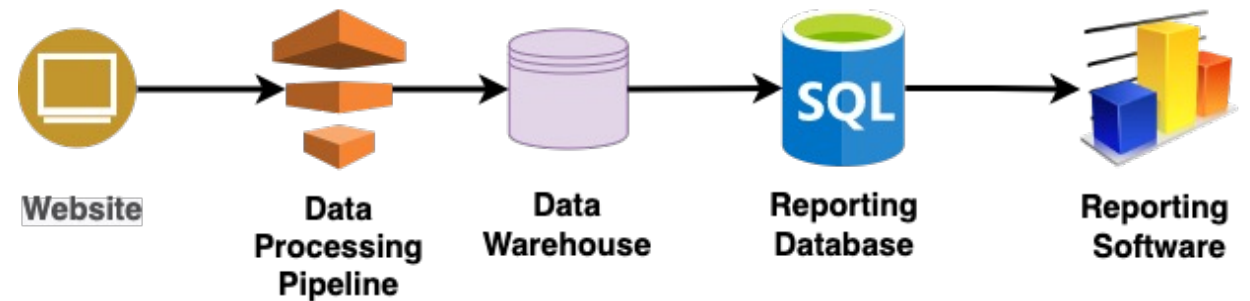


Example

Here is an example of a simple architecture used by a small e-commerce company that sells products on its website.

The company's data needs are to store orders and sales data in its data warehouse and be able to report on them.

The architecture shows how data flows through the system and how resources (e.g., databases) will be used to meet the data needs of the organisation.



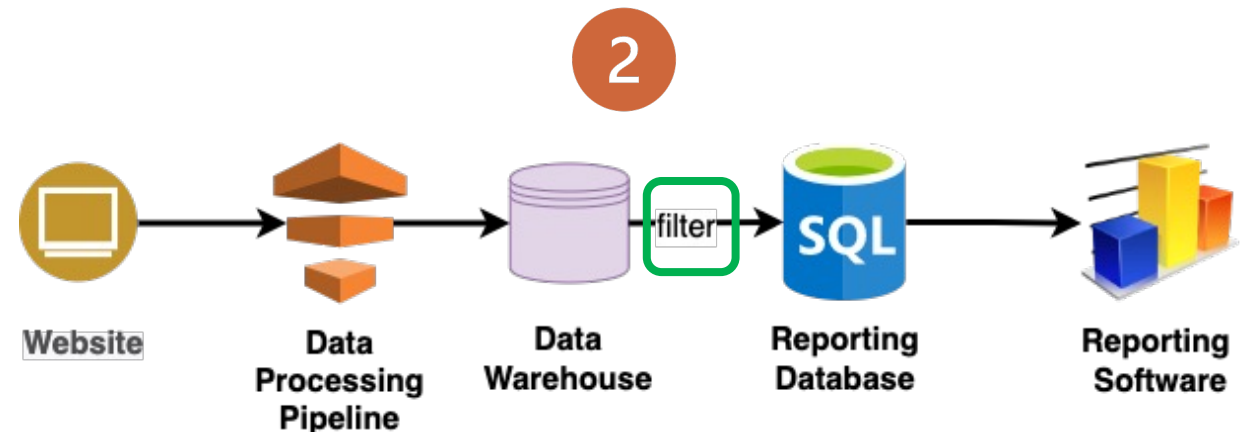
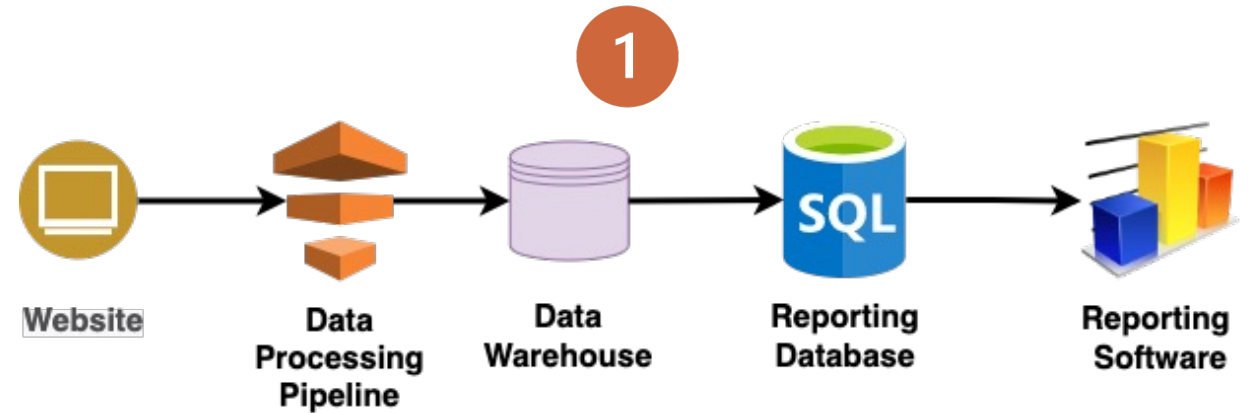
Example



The architecture shown in diagram 1 is fine when there is a small amount of data.

However, as the amount of data grows this will become more **expensive**, because *all* of the data is transferred from the data warehouse to the reporting database.

The architecture could be changed to 2 so that the data is *filtered* before being stored in the reporting database. This will reduce the amount of data stored, which will lower the cost.



Data Modelling and Design

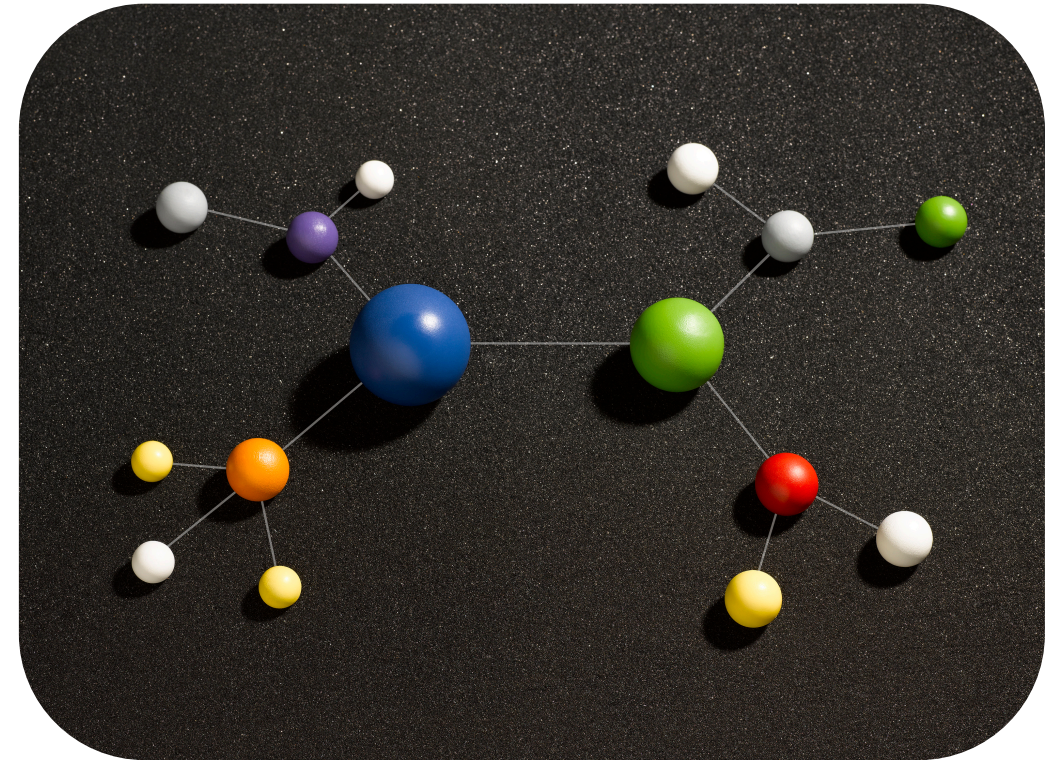


What is it?

Modelling and design involves the discovery of data requirements and the design and planning of a model that specifies these requirements. A model is a **diagram** that shows **what data there is, how it fits together** and **how it relates to things in the real world**.

Why do it?

Modelling is needed to help an organisation understand what data it has and how it fits together. A data model is a useful tool for communicating this, and helps with the design of the systems that will use this data.

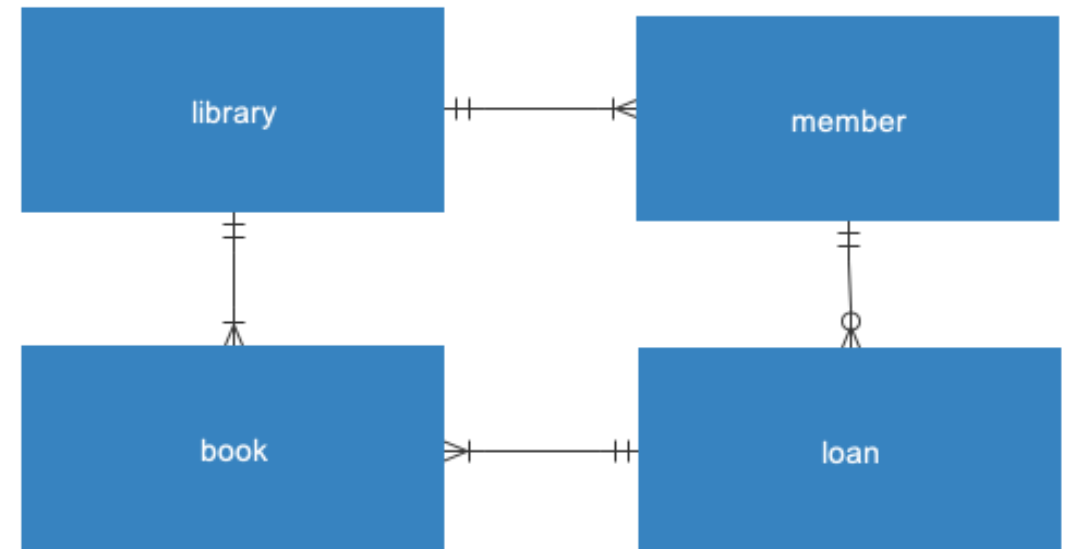


Example

Here is an example of a simple data model for a library.

It shows **the data that is needed**, based on **things in the real world** (i.e., the library, its books, its members and the book loans they have).

It also shows **how the data fits together** e.g., a library has 1 or more members and 1 or more books, and a member of the library has 0 or more loans.



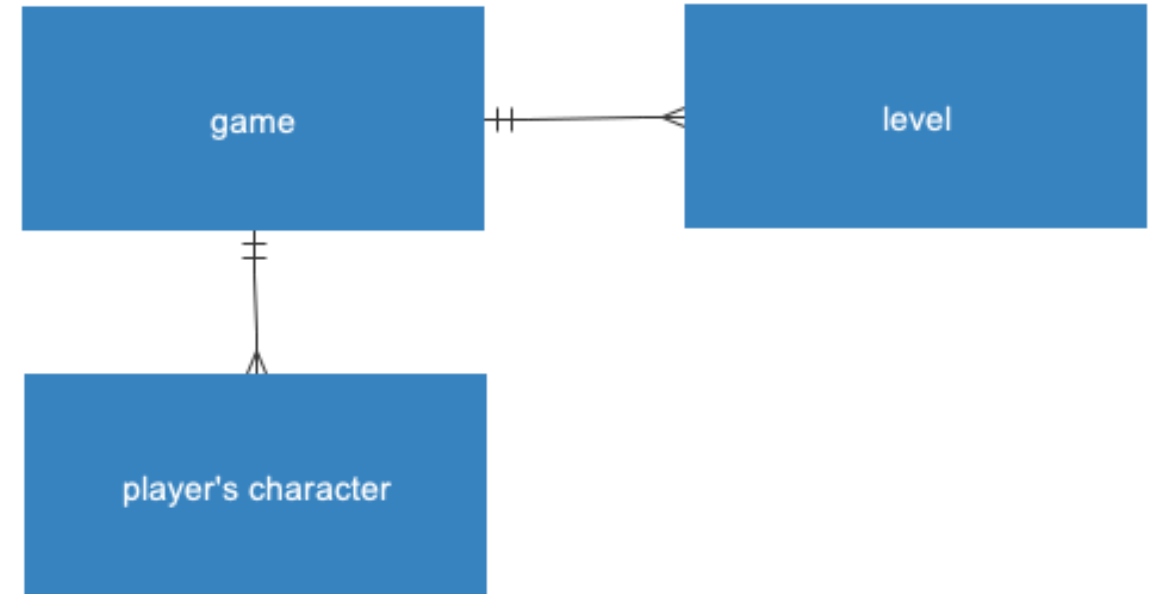
Your turn....



A video games development company are designing a data model for a new 2-player, multi-level puzzle game, where there's a different kind of puzzle to be solved on each level.

Here's where they've got to so far.

What other data do you think needs to be included in the data model? What have they missed?



Your turn....



Some additional data that could be included in the data model are:

- Players
- Scores
- Puzzles
- Activities
- Skills

You can probably think of many more!



Example



Imagine the games company create a data model that *doesn't* capture **the data, the relationships between them** and **how they relate to real-world items** well.

Because the data model will be used to design how the data will be stored and processed, the company will design and build systems **which are not suitable for their needs**.

This could be a very expensive mistake!



Next steps

Complete **questions 1 to 5**
in **section 3** of the
'Data Management' workbook.

Data Storage and Operations



What is it?

The technical design, implementation and support for the storage of data. This covers the whole lifecycle including backup, archiving and deletion of data.

Why do it?

Organisations rely on the availability of data. The systems in which data is stored need to have high availability and perform well. Systems which do not are likely to severely impair the operations of the organisation.



Example



Think of your favourite social media platform. A massive volume of user data is stored by them (posts/photos/videos etc).

Imagine they had not accurately predicted the growth of the platform's user-base and had not planned to sufficiently increase the data storage needed.

This is likely to impact on the speed at which data can be accessed, including how responsive the platform feels to a user.



Your turn....



What might the consequences be for the business if large numbers of users have a poor experience of the platform due to slow response times?



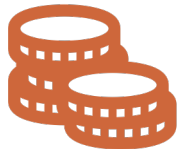
Your turn....



Users of a social media platform expects it to be responsive (i.e., fast). If it is not, the business may experience:



customer dissatisfaction – users won't put up with a poor user experience for long. They may leave the platform if the issue isn't addressed.



financial costs – if users do leave, this will have a financial impact.

A solution would be to use data storage technologies which automatically **scale** (i.e. increase or decrease resources) as demand changes.

Data Integration and interoperability

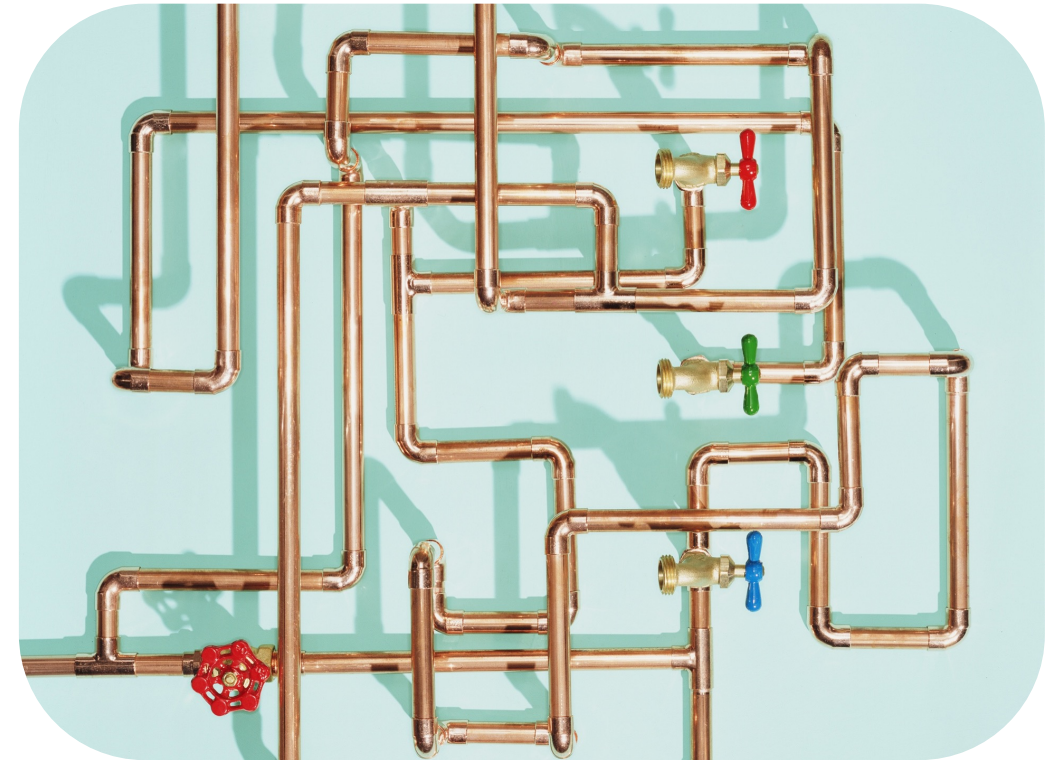


What is it?

Most organisations have multiple systems in which data is stored. Data moves between these systems. This area involves ensuring these systems can communicate with one another.

Why do it?

Data that is well-managed as it moves between systems is more likely to be of a higher quality and more be suitable for analysis.



Example (1)



A theatre stores data on ticket sales on their ticketing system. It is operated by box-office team.

When the sales team wish to analyse the ticketing sales data they email the box-office team and request a spreadsheet. The box-office team export the spreadsheet from the ticketing system and email it back.

The process is **manual**. The systems (ticketing system and spreadsheet software) do not communicate with each other. There is **no data integration**.



Your turn....



1. What are the issues with the theatre's process for accessing the ticketing sales data?
2. What could be done to resolve these issues?



Your turn....



The main issues with the theatre's process for accessing data are:

1. it's not easy for the people who need the data (the sales team) to **access** it.
2. it's not **efficient**, because it's a manual process
3. it's prone to **errors**, because it's a manual process

A solution would be to **automatically export the data** to a system the sales team can access.



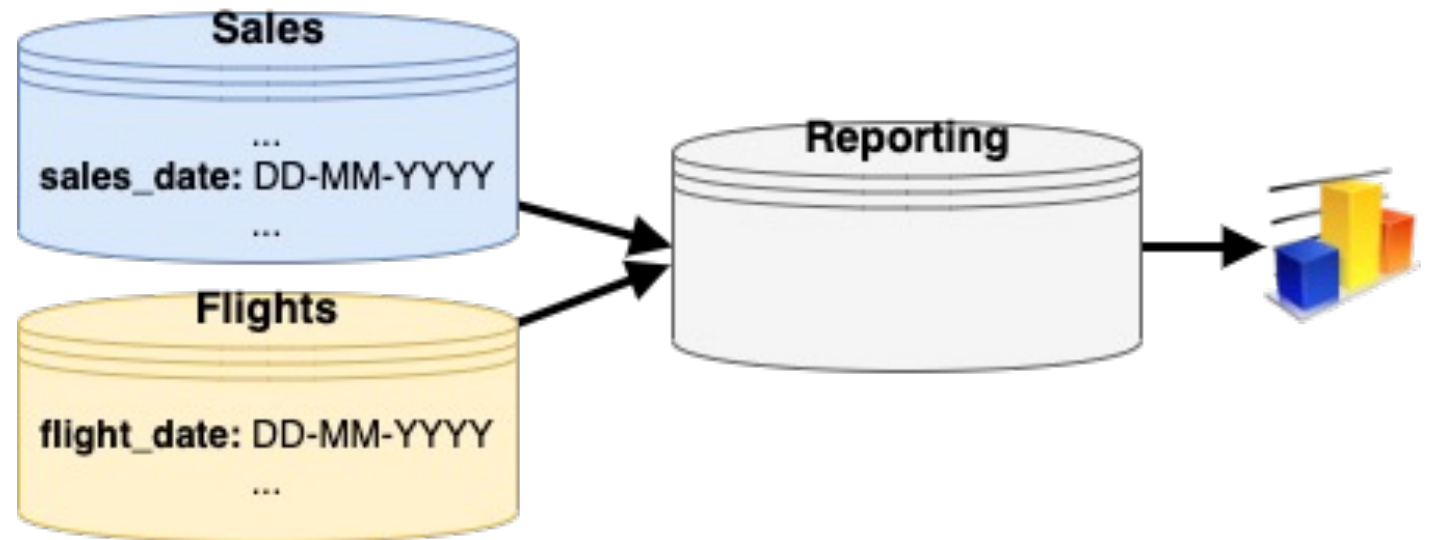
Example (2)



Data analysts working for an airline use a reporting database to undertake their analyses. Data in the reporting database is imported from the Sales and Flights databases.

Some reports compare flight dates to the dates on which flight tickets were purchased.

Both these dates are in the same format, DD-MM-YYYY e.g. 23-12-2022.



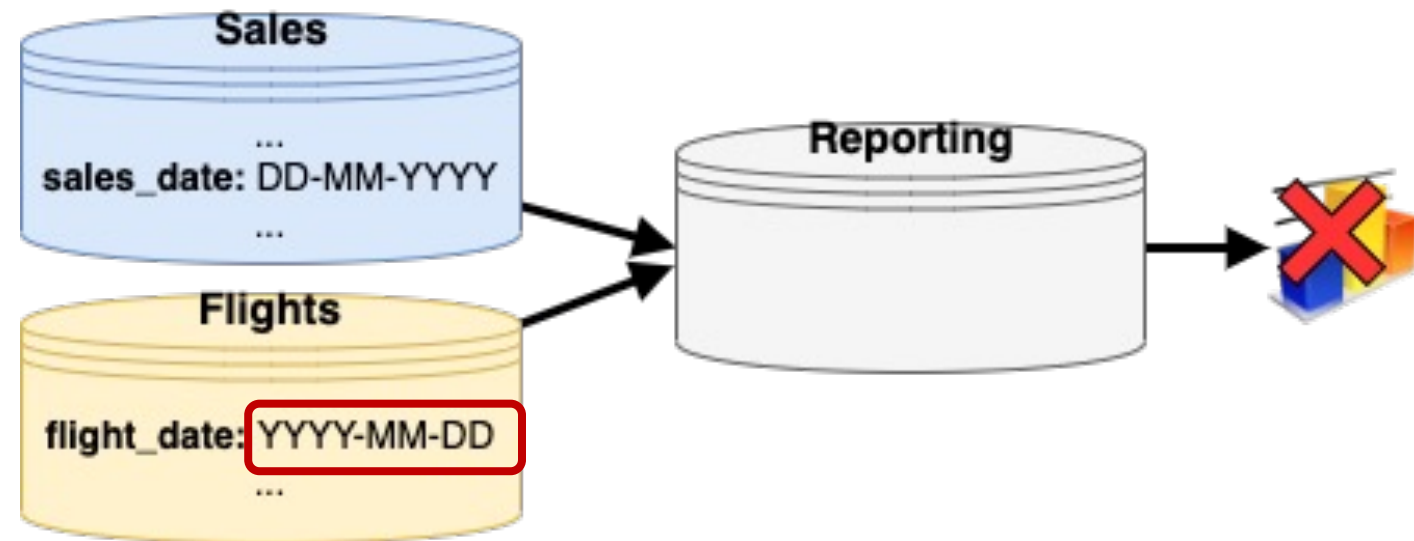
Example (2) - contd



One day, **the reporting application suddenly stopped working and failed to generate any reports.**

Engineers trace the issue back to **the format of dates** in the Flights database. **They had changed.**

The reporting database expected the data in one format but received it in another format.



The mismatch caused the reporting application to stop working.

Your turn....



1. What is the issue with the integration between the Flight/Sales databases and the Reporting database?
2. What could be done to resolve this issue?



Your turn....



The main issue with the airline's data integration is that the structure of the data sent from one system to another changed without warning.

The **interface** between the system *sending* the data and the system *consuming* the data changed.

A solution would be to **plan for changes** between the systems and **communicate planned changes** to those who need to know.



Other Data Management Areas



- **Metadata** - data about data. Metadata is vital for an organisation to know what data it has, what it represents and how it is accessed and used.
- **Reference and Main Data** - data used by other data sources (*reference data*) or data that contains the core information that is important to an organisation (*main data*).
- **Data Warehousing & Business Intelligence** – a data warehouse is a large data store used for business intelligence activities, where the data is aggregated from multiple data sources. Business Intelligence involves the use of analytics to make informed business decisions.

More information about metadata, and reference and main data can be found in the **Caring for Data** lesson.

Summary

It's easy think of using data as just collecting it, preparing it and analysing it.

However, **using data well involves much more than this**, as you've learned in this lesson. It **needs to be cared for**, in lots of different ways.

When we take care of it, and manage it well, data can be incredibly valuable and bring lots of benefits to the people and organisations who use it.

However, it's easy *not* to manage it well, and when this happens the consequences can be disastrous.



Next steps

Complete **questions 1 to 3**
in **section 4** of the
'Data Management' workbook.

Learning checklist

I can *describe* the areas of data management based on the DAMA Wheel

I can *explain* why it is important for organisations to manage data

I can *describe* some data management activities organisations undertake

I can *describe* some consequences of not managing data

How you can use this lesson



You are free to:

- **Share** – copy and redistribute the material in any medium or format
- **Adapt** – remix, transform and build upon the material

Under the following terms:

- **Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **NonCommercial** — You may not use the material for [commercial purposes](#).
- **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

© 2022. This work is licensed under a [CC BY-NC-SA 4.0 license](#).

Created by effini in partnership with The Data Lab.



Alternative format

If you require this document in an alternative format, such as large print or a coloured background, please contact

hello@effini.com

or

4th Floor, The Bayes Centre
47 Potterrow
Edinburgh
EH8 9BT

