Knowing Your Data

What we need to know

- Purpose

- Content

- Method of Creation

- Attributes

- Relationships

- Constraints and Dependencies

Purpose

- Why was the data created?

- Who needs it?

- How long is it useful for?

- Why should it be retained?

- Is it for personal, business, or public use?
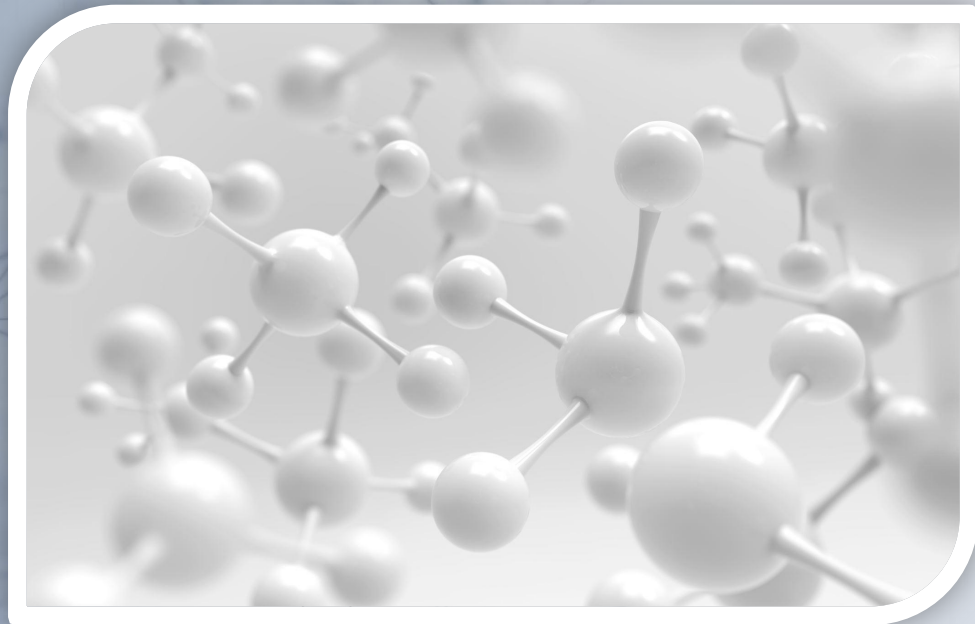
Content

- What is it about?

- What type of content is it?

- Does it contain sensitive data?

- Does it contain personal data?

Method of Creation

- Is it raw data or has it gone through levels of processing or interpretation?

- Was it created using proprietary software?

- Has it been exported in a reliable method?

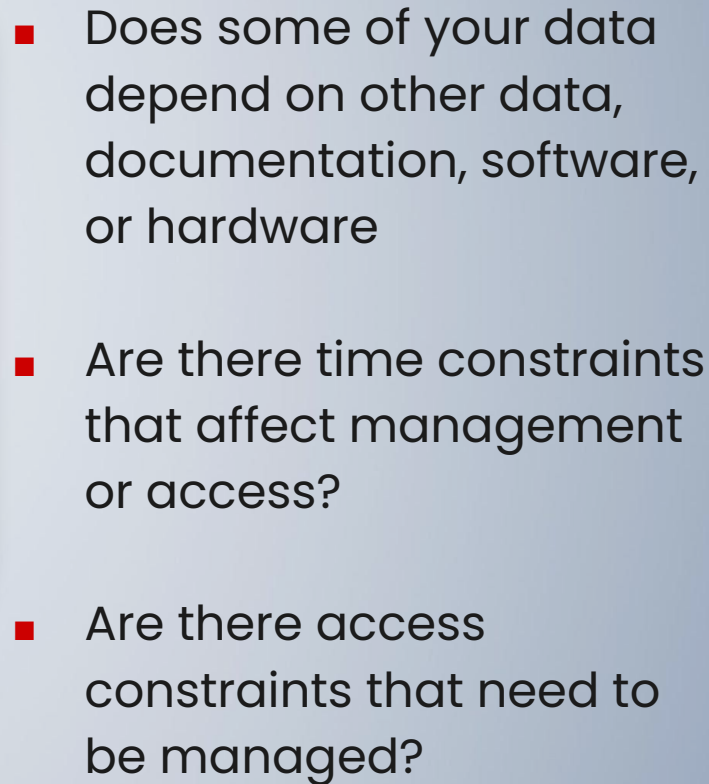- What are the chances of data loss or corruption?

Properties

- What significant properties need to be preserved?

- What is important for the end user?

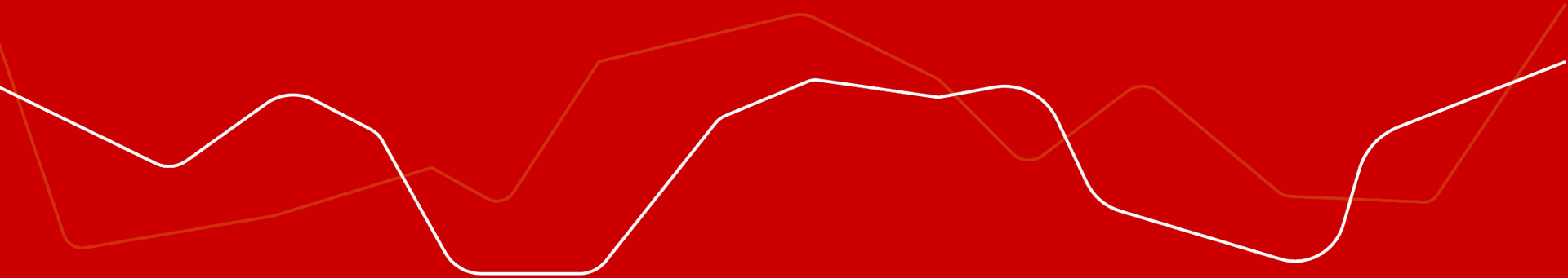- What do the significant properties require?

Relationships

- Which relationships between files or groups of files are important?

- Which files belong together?

- Which files or collections of files are associated with others?

- Do you need to relate your data to external collections or resources?

Constraints & Dependencies

- Does some of your data depend on other data, documentation, software, or hardware

- Are there time constraints that affect management or access?

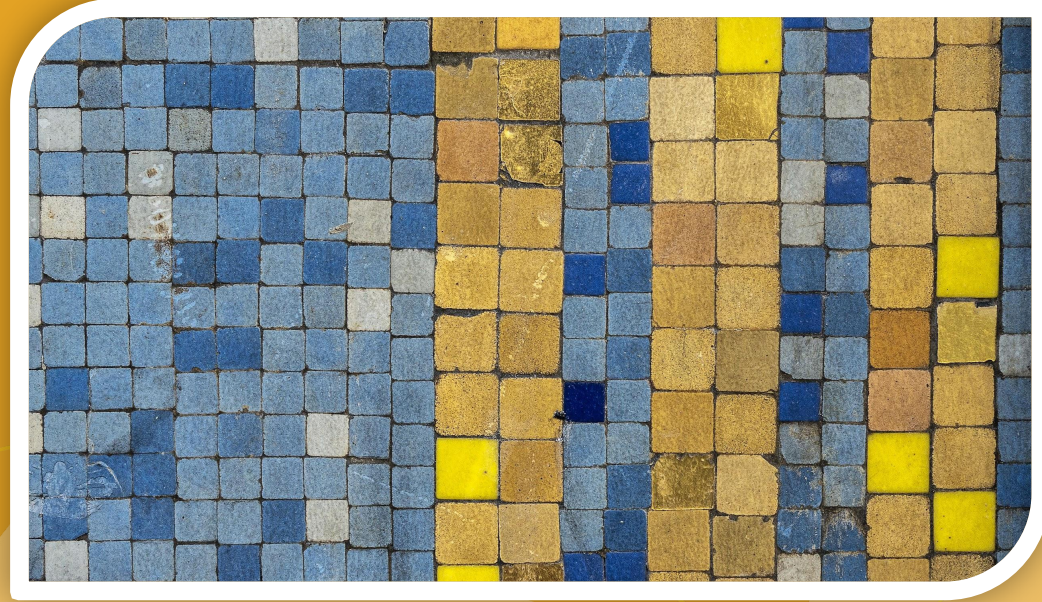- Are there access constraints that need to be managed?

Organising Your Data

Establishing a system

- File Structure
- Filenames
- Version Control
- Discard policy
- Backup
- Security
- Documentation

File Structure

- Group files with common traits:

- Purpose

- Content

- Source or method of creation

- Attributes

- Relationships

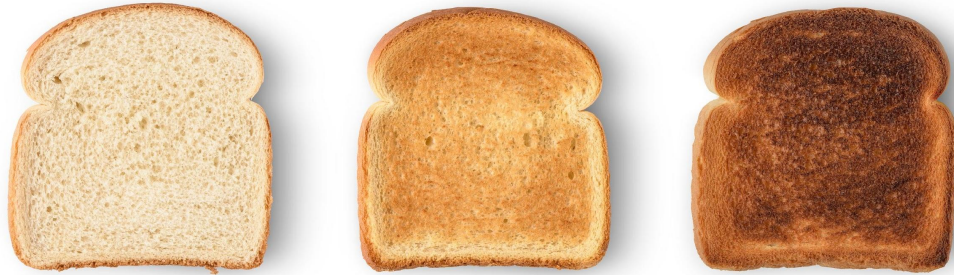- Access constraints or dependencies

Filenames

Version control

Group or name files so that the version is clear, is the file or data in it's raw state? Is it being actively worked on? Is it a representation of a completed draft or process? Or is it a closed final version, ready for public use?

Discard policy

The way you organise and structure your data can inform and maintain your discard policy. Clear filenames and version control can indicate when a file or group of files have exceeded their purpose and no longer have any reuse value.

Backup

- Consider organising your data to facilitate data backup – thinking about backup frequency, cost, access downtime, and retrieval methods

Security

Using data structure and organisation to manage access, restricting access to sensitive data

Documentation

All data and policies will need some form of documentation to transfer knowledge about everything that I've mentioned so far. So consider documentation storage and how relevant documentation is related to the data itself.