# RDM lifecycle

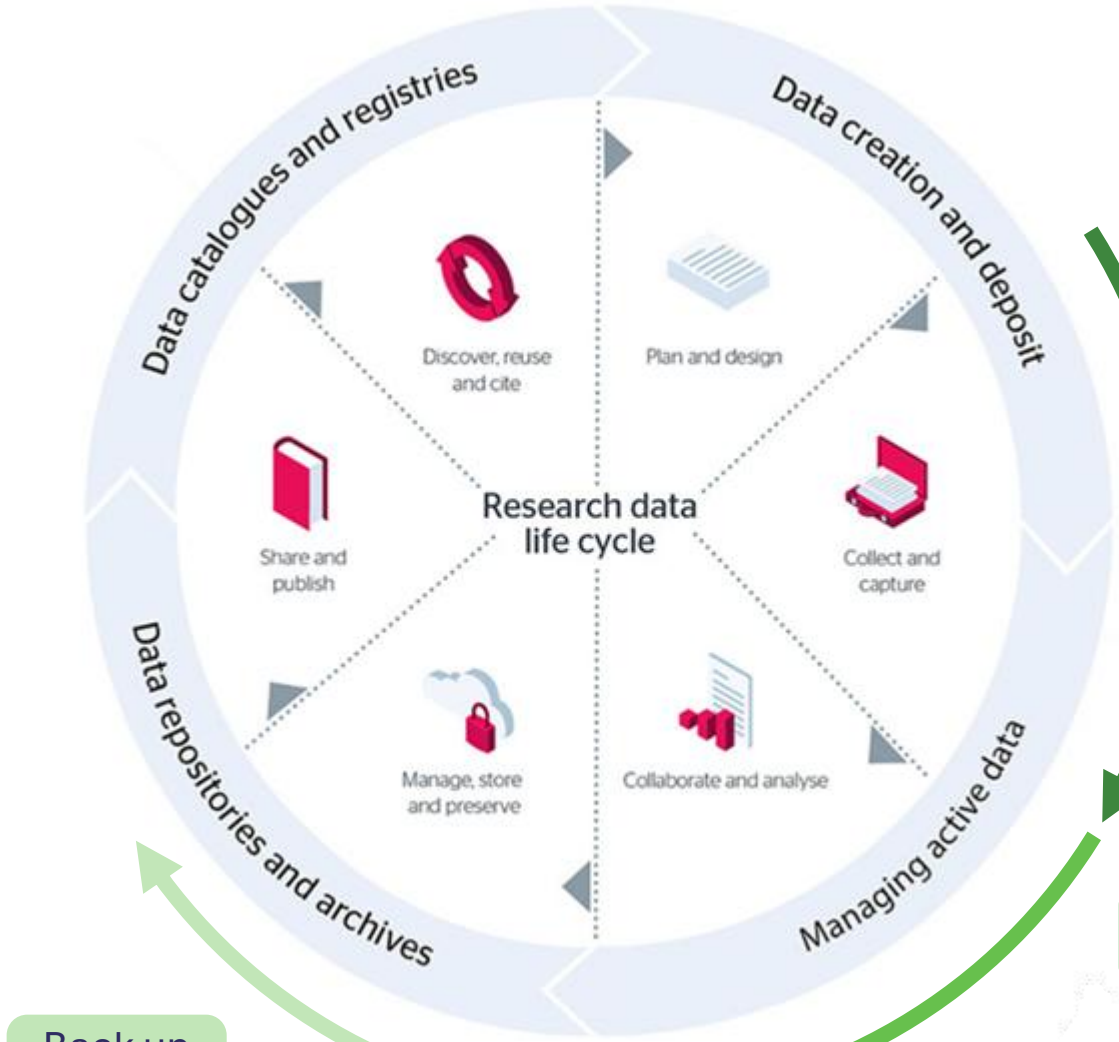
- Data Collection
- Processing and management day-to-day

Describe the data

Process the data

Organise the data

Save the data




Collection

Naming convention

Metadata

Versioning

Data formats

Back up storage

© Jisc

NERC Environmental Data Service

# Collecting data in the field

Understand the deployment environment by asking yourself the following questions:

- Is it a deep field site or a remote field station?
- Data server – may have RAID storage, offline backups
- Time server – run an NTP client to sync accurate time
- Reliable power supply – consider UPS, surge protection, IP PDU
- Remote network comms: scheduled transmission and remote login

**General tips:**

- Storage is cheap, but collecting remote, in situ data, is expensive
- Structure raw data appropriately
- Monitor and alarm propagation
- Have a second storage device
- Ensure available storage can withstand a missed site visit
- Capture metadata along data acquisition

NERC Environmental Data Service

# Case study: Data collection on a ship

Ship track – position data

'Event' deployments – needs to be linked to other continuous data

Physical sample/lab metadata

Multiple continuous data acquisition systems e.g. Multibeam echosounder, ADCP


I prepare the information and material I need for my analysis, record any information that might impact the  quality of the data, make sure to back up the data

# Data on a day-to-day basis

## DOCUMENTATION AT PROJECT LEVEL

- Data collection methods, including instruments and protocols
- Standards used and replication
- File and data structure
- Procedures for quality control and versioning
- Problems that may affect the use of the data

## DOCUMENTATION AT DATA LEVEL

- Date created and version
- Names and definitions of fields (variables)
- Reasons for missing values
- Problems that may affect specific records
- Code used to generate processed data values

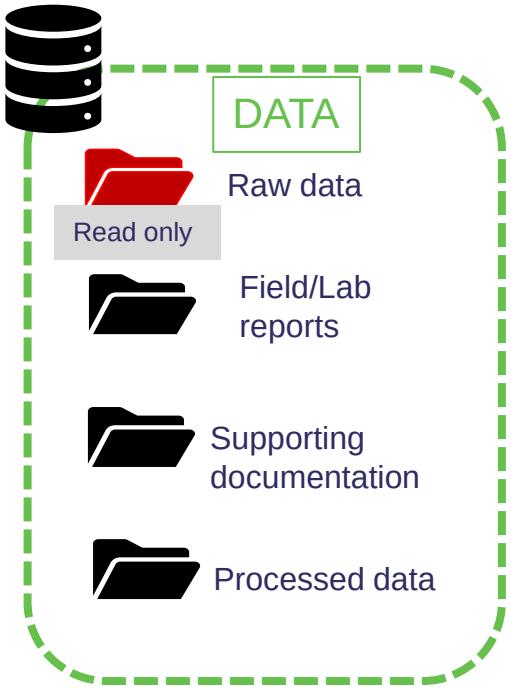
Recording these details will make it easier to create metadata

NERC Environmental Data Service

After Sandy Harrison presentation on RDM

Manage your citations and references using citation software


# Folder structures




## Use folders and structure them hierarchically

- Group files within folders so particular topics are in one place
- Make sure folder structure is understandable by others
- Start with a limited number of generic folders, then create more specific sub-folders

## Adhere to existing procedures

- Check for established team/department approaches which you can adopt
- Agree on a common folder structure that makes sense to everyone
- Document strategy in DMP

## Name folders and file appropriately

- Be consistent when you name your data and the folders they're stored in
- Avoid spaces and special characters
- Use a file naming convention

## Separate raw and working data

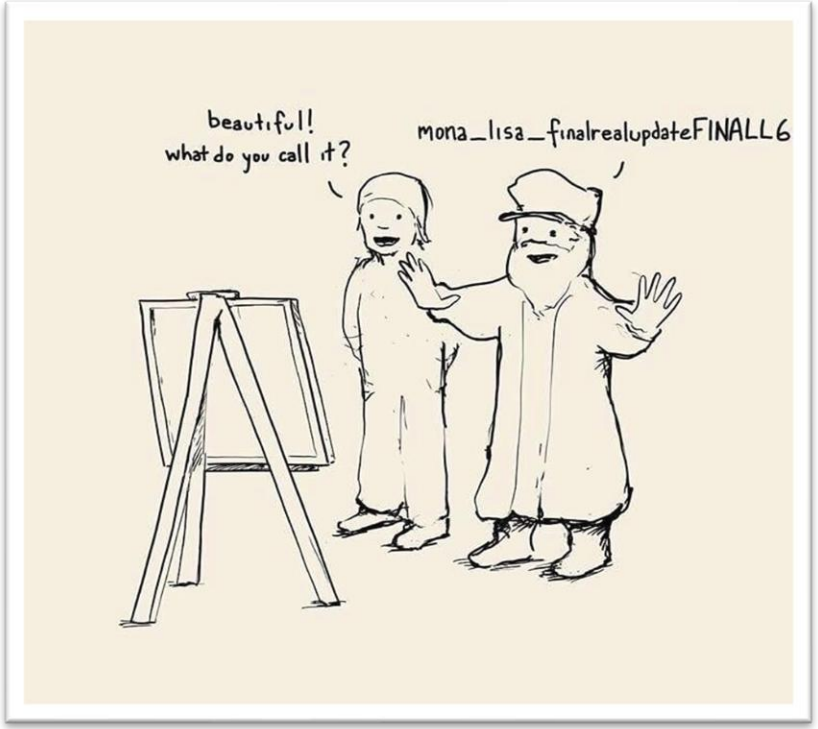
- Do not make any changes / corrections to the original raw data files to preserve original files
- Consider having your raw data read only to avoid deletion or changes.

NERC Environmental Data Service

# File-naming conventions

Decide on a file naming convention at the start of your project. Useful file names are:

- ✓ Consistent
- ✓ Meaningful to you and your colleagues
- ✓ Allow you to find the file easily




**Some tips:**

- ✓Start your file name with the date, formatted as YYYYMMDD, to display your files in chronological order
- ✓Don't make file names too long
- ✓Special characters should be avoided
- ✓ For sequential numbering, use leading zeros to ensure files sort properly."01, 02…99, etc" instead of "1, 2…99, etc."
- ✓Do not use spaces: Instead use underscore (file_name), dashes (file--name), no separation (filename or FileName)
- ✓Include versioning when appropriate

| File name | Changes to file |
|---|---|
| ResearchReport_1.0 | Original document |
| ResearchReport_1.1 | Minor revision made |
| ResearchReport_1.2 | Further minor revisions |
| ResearchReport_2.0 | Second approved version |

https://help.eds.ukri.org/article/5182-file-naming

# Units

- Describing parameters in a reusable way
- Making data understandable for others
- Placed in the header, not in the data
- Unified Code for Units of Measurement (UCUM)
  - All units of measure
  - International science, engineering
  - 7-bit US-ASCII character set

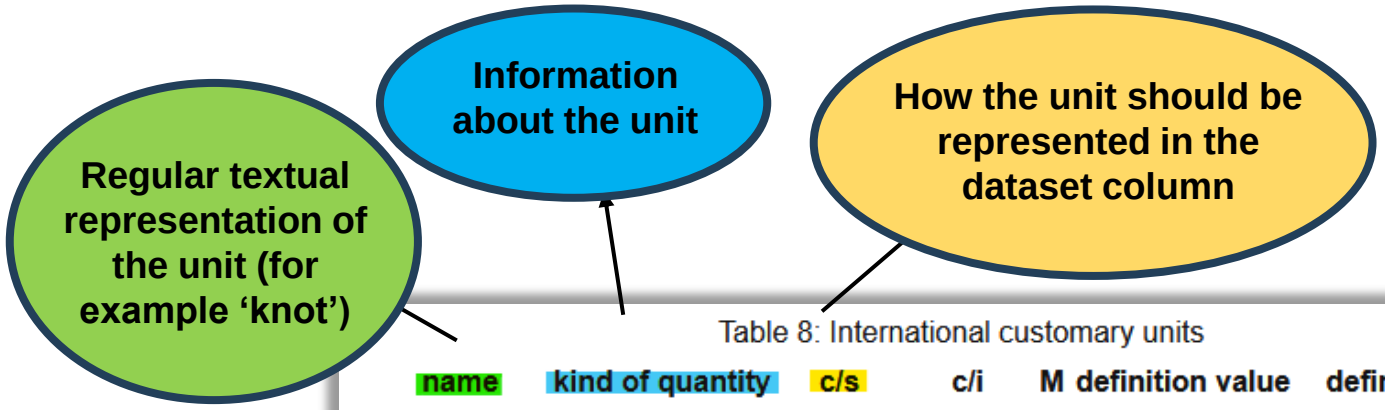



Table 8: International customary units

| name | kind of quantity | c/s | c/i | M | definition value | definition unit |
|---|---|---|---|---|---|---|
| inch | length | [in_i] | [IN_I] | no | 2.54 | cm |
| foot | length | [ft_i] | [FT_I] | no | 12 | [in_i] |
| yard | length | [yd_i] | [YD_I] | no | 3 | [ft_i] |
| mile | length | [mi_i] | [MI_I] | no | 5280 | [ft_i] |
| fathom | depth of water | [fth_i] | [FTH_I] | no | 6 | [ft_i] |
| nautical mile | length | [nmi_i] | [NMI_I] | no | 1852 | m |
| knot | velocity | [kn_i] | [KN_I] | no | 1 | [nmi_i]/h |
| square inch | area | [sin_i] | [SIN_I] | no | 1 | [in_i]2 |
| square foot | area | [sft_i] | [SFT_I] | no | 1 | [ft_i]2 |
| square yard | area | [syd_i] | [SYD_I] | no | 1 | [yd_i]2 |
| cubic inch | volume | [cin_i] | [CIN_I] | no | 1 | [in_i]3 |
| cubic foot | volume | [cft_i] | [CFT_I] | no | 1 | [ft_i]3 |
| cubic yard | volume | [cyd_i] | [CYD_I] | no | 1 | [yd_i]3 |
| board foot | volume | [bf_i] | [BF_I] | no | 144 | [in_i]3 |
| cord | volume | [cr_i] | [CR_I] | no | 128 | [ft_i]3 |
| mil | length | [mil_i] | [MIL_I] | no | $1 \times 10^{-3}$ | [in_i] |
| circular mil | area | [cml_i] | [CML_I] | no | 1 | [pi]/4.[mil_i]2 |
| hand | height of horses | [hd_i] | [HD_I] | no | 4 | [in_i] |

NERC Environmental Data Service