# Guidance

This template may help you structure the supporting information you should deposit with a repository, alongside the data you have generated, if you haven’t already produced a suitable document. Please overwrite the grey guidance text with your own text and delete any sections that are not applicable to the resource you are depositing. You can also add any additional sections or information that would aid the reuse of the data you are depositing**. Any sections where no conditions are specified should be included**. Always refer to the individual repository’s guidance on supporting information before proceeding.

# Working Title

Describe the data resource, not the project that created it or the findings from an experiment. Location and temporal extent should be included if relevant. You may be required to modify the title of the data resource depending on the requirements of the data repository when depositing.

# Summary

A short summary about what data are included and why they were collected/generated.

# File format

List the format(s) the data file(s) are in e.g., CSV, NetCDF, GeoTIFF. Repositories may only accept non-proprietary formats. Check guidance on suitable formats from the repository you are using.

# File names and/or naming convention

List the file names of the data files (including the file extension).

If the files are arranged in a folder (and subfolders), that hierarchy needs to be described here.

If you have a lot of files (e.g., more than 20), you may prefer to specify any file naming conventions used instead.

Example: There are 20 files with the following name convention: YYYYMMDD\_rainfall\_StationName.csv

Where YYYYMMDD refers to the date the measurements were recorded and StationName is the name of the station at which the measurements were recorded.

# Nature and units of recorded values

Include information about the structure of the data (variable name(s), description, units).

* For tabular data (e.g., spreadsheets or databases) we recommend a structure such as the table below. If you have multiple files with different variables in each, use one table per file.
* For image data, you should include information such as resolution, focal length, etc.
* If applicable, include details of valid range of values for the variables, lowest level of detection, determinands etc.
* If there are missing data, state the value used to represent them (e.g., "*Missing values are indicated by 'N/A'*")
* If available, include URLs for vocabulary terms

Table 1. Description of variables in [insert filename].

| Variable/Column header | Description | Units | Vocabulary URL (if applicable) |
| --- | --- | --- | --- |
| *Dissolved nitrate* | *Nitrate measured in soil* | *mg/l* | *http://onto.nerc.ac.uk/CAST/16* |
|  |  |  |  |
|  |  |  |  |

# Spatial coverage **[CONDITIONAL – Required for spatial data resources]**

Include the following:

* name and bounding coordinates of the region OR for sampling locations provide the coordinates for each location OR if those aren’t possible provide the name of the general region/locations.
* Coordinate reference system and/or grid used e.g., WGS84, OSGB 1936 / British National Grid
* If applicable, relevant geographical information e.g., aspect, elevation, surface area, volume etc.

# Temporal coverage and resolution **[CONDITIONAL – Required if applicable]**

Include the time period during which the data collection took place e.g., 01 January 2023 to 31 December 2023.

Include how often the measurements/sampling/experiment/monitoring took place e.g., annually, daily, hourly, every 5 minutes.

# Models **[CONDITIONAL – required where models are being deposited]**

Include information on the model:

* Brief description of the methods behind the model, including references/links to the authors/source of the model
* Any training or calibration steps that were taken prior to running the model
* Where and how the model was run e.g., platform used and how many times it was run
* Details of model configurations used during model runs
* Model version

## Input data **[CONDITIONAL- required when depositing input data with a model]**

Describe the input data used in the model, this can include the following details:

* Source(s) of the input data, include links/references to the source(s) of the data
* Any pre-processing steps applied to the data before they were used in the model
* Parameters used from the input data

## Output data **[CONDITIONAL – required if depositing model output data]**

Describe any post-processing steps applied to the data once it was output from the model, this can include the following details:

* Any processing/analysis of the output data that changed the data values e.g., gap filling, converting to different units, calculating uncertainty.
* Transformations to the output file formats to make them suitable for deposit with a repository.

# Collection/Generation/Transformation methods

Describe how the data were collected/generated, and if applicable how they were transformed, this can include the following details:

* Description of the methods/techniques used
* Standard Operating Procedures (SOPs) and references of methods used.
* Sample storage details
* Date of analysis of samples if different from date of sampling.
* How/what format the data were collected in e.g., in a database, on paper then converted to digital, etc.
* Any processing/analysis of the raw data that changed the raw measurements/data values e.g., gap filling, converting to different units, calculating uncertainty.
* Transformations of the file formats for deposit at a repository e.g., converting from Excel spreadsheet to CSV format.

# Experimental design/sampling regime **[CONDITIONAL - required for monitoring/experimental data]**

Describe the experimental design and/or sampling regime, this can include the following details:

* Any treatments applied (how they were applied/created/managed/verified)
* Any sample/observation replication, including explanations for any missing samples/observations
* Any control methods employed
* Total number of samples/observations collected
* Frequency of sampling

# Fieldwork and/or laboratory instrumentation **[CONDITIONAL - required for experiments where instruments used]**

Include information on instruments/machines used for collection/analysis of samples/observations. This should include the type/make, model and serial number of each particular instrument/machine and any persistent identifiers (PID) of instruments, where known.

# Calibration steps and values **[CONDITIONAL – required where instruments have been used]**

Include details of steps taken to calibrate any instruments/machine used, including use of any blanks, and the values used for calibration.

# Quality control

Describe any quality control measures undertaken to ensure the quality of data values. This can include the following details:

* Methods of quality control e.g., identifying missing/incorrect values
* Explanation of quality codes
* Factors affecting the data e.g., reasons for missing data, or incorrect values, adverse fieldwork conditions

# Miscellaneous

Any additional information necessary for someone else to understand and reuse the data. E.g., any known limitations on reuse of the data, references to any papers which use the data etc.

# References

If you have included any references, please list them here.