



## UKERC ENERGY DATA CENTRE

# GUIDANCE FOR DATA PROVIDERS

### SCOPE

[STFC's Energy Research Unit](#) provides the UKERC Energy Data Centre (EDC) facility. The EDC functions include long-term archive of energy data sets to secure them in the long-term and to make them as accessible as possible to anyone with permission to make use of them.

The purpose of these guidance notes is to encourage and enable researchers to deposit data in the EDC.

## WHY DEPOSIT YOUR DATA WITH THE ENERGY DATA CENTRE?

### FOR SECURITY

Data deposited with the Energy Data Centre will be backed up and managed for the long-term. We actively manage data so that storage media and formats remain readable and don't go out of date.

### FOR ACCESSIBILITY AND ACCESS CONTROL

Data in the Energy Data Centre can be accessed online from anywhere by permitted users. Data can be made publically available, or constrained to specific users where required.

### PUBLICATION REQUIREMENTS

It is increasingly common for data publication to be required, either by funding bodies or to support a journal publication. Depositing data in the Energy Data Centre can be a straight-forward way to meet such requirements.

In addition, we are able to create a Digital Object Identifier (DOI) for data lodged with us, providing a persistent identifier for linking to, and citing, your data set.

## DATA REQUIREMENTS

### DATA FORMATS

Date and time data formats must be very clearly defined and information about the time zone used must be given. They must be consistent throughout the data. Ideally the timestamp should follow ISO 8601 ASCII format (and not be a date number).

ISO 8601 format principles:

- a) date and time values are ordered from the largest to smallest unit of time
- b) the separator used between date values (year, month, week, and day) is the hyphen, while the colon is used as the separator between time values (hours, minutes, and seconds)
- c) seconds can have a decimal i.e. ss.sss

For example : yyyy-mm-dd hh:mm:ss

The ISO 8601 principles should be used even if the full date is not included in the data e.g. ss.sss only or hh:ss.sss only

### FILE FORMATS

We strongly encourage depositing data in CSV (comma-separated variable) format where possible, using ASCII (American Standard for Information Interchange) character set. We can also accept data in standard data dissemination formats: that is in formats that are clearly defined and widely used by a relevant community (eg Matlab, NetCDF). Please contact us about other formats.

However, Excel is not recommended as it does not behave consistently across software or operating system versions and has a tendency to conceal values through formatting in some circumstances.

## CSV EXAMPLE

### TABLE OF DATA

Column 1 (units)	Column 2 (units)	Column 3 (units)	Column 4 (units)
2018-12-31 12:34:59.3	1000.4	8888.9	5.7
2018-12-31 12:35:59.3	12.7	145.8	12345.8

### THE SAME DATA IN CSV FORMAT

```
`Column1 (units)', 'Column 2 (units)', 'Column 3 (units)', 'Column 4 (units)'  
'2018-12-31 12:34:59.3', 1000.4, 8888.9, 5.7  
'2018-12-31 12:35:59.3', 12.7, 145.8, 12345.8
```

Note that text values, including the timestamp, are quoted. This example includes a header row as the first row and this is good practice.

## FILENAMES

File and directory names in the Energy Data Centre use a restricted character set in order to avoid problems that can arise from characters that have a special meaning in URLs or different computer systems.

Filenames can only contain the following ASCII characters, with only one full-stop that comes before the file extension:

a-z A-Z 0-9 - _ .
-------------------

We encourage the use of meaningful filenames which will assist an end-user in making sense of the data. For example including within the filename the date information in the ISO date order format yyyy-mm-dd or yyyyymmdd.

If there is a filename extension convention for this type of data, please use it.

## DOCUMENTATION

If you have documentation or reports about the data or the project that generated it and that may be useful to an end user, we encourage you to deposit these in the Energy Data Centre alongside the data. Our preferred format for these is PDF-A which is an archive suitable PDF format.

We will also work with you to provide plain ASCII text files (known as 'ReadMe' files) in data directories that describe to a data user what the contents of the directory are. Ideally there should be a **Folder\_Readme.txt** describing the folder structure, and a **Format\_Readme.txt** describing the format layout – columns, units, how missing data is indicated, timing conventions etc - basically everything that an informed but un-involved person will need to be able to use the data. If a common format is being used throughout the experiment this can be combined with the folder structure **Folder\_Readme.txt**, otherwise each folder/sub-folder will need its own **Format\_Readme.txt**.

## FILE STRUCTURE PLAN

We are flexible about how the data is stored – sometimes people decide to make a folder for each experiment, others have a folder for each year and then sub-folders for each month/day/experiment.

## METADATA REQUIREMENTS

Metadata is information about the data. It enables data discovery and gives an end-user the information they need to read and make sense of the data.

Data depositors must also complete our web form to ‘[Submit a Link or Dataset](http://ukerc.rl.ac.uk/DC/cgi-bin/edc_submit.pl)’ at [http://ukerc.rl.ac.uk/DC/cgi-bin/edc\\_submit.pl](http://ukerc.rl.ac.uk/DC/cgi-bin/edc_submit.pl). This will provide the metadata for the Energy Data Centre’s data catalogue and will make data discovery possible.

Examples of published Metadata descriptions can be found via the Energy Data Centre’s Data Catalogue at [http://ukerc.rl.ac.uk/DC/cgi-bin/edc\\_search.pl](http://ukerc.rl.ac.uk/DC/cgi-bin/edc_search.pl)

It is important that you include appropriate keywords in the data description as this will enable users to find relevant datasets.

## HOW TO DEPOSIT YOUR DATA WITH THE EDC

- 1) Get in touch with us by emailing [EDCManager@stfc.ac.uk](mailto:EDCManager@stfc.ac.uk). It would be useful to advise us if you wish to restrict access to the data in any way, give us an approximate size of the deposit, and indicate if the data will be deposited all at once, or will be added to during the project. **NB We can offer advice and help throughout.**
- 2) We will then make contact with you to discuss your proposed folder structure, data formats, proposed documentation and to request a sample data file (with the word DRAFT included in the filename to ensure it is not subsequently archived)
- 3) We will then review your planned deposit to ensure that we can understand the data layout. We may suggest changes to the layout or description.
- 4) We will then ask you to :
  - a) Put the data you wish to deposit in the suitable format,
  - b) Give it suitable and meaningful filenames,
  - c) Create the necessary readme files, and other descriptive documentation and reports.
  - d) Complete our online form ‘[Submit a Link or Dataset](http://ukerc.rl.ac.uk/DC/cgi-bin/edc_submit.pl)’ to provide the metadata for the EDC catalogue.
  - e) Send us a copy of the data, and all documentation files. This can be by email, FTP, or on a physical storage device, depending on the size.

Summary of files and information to be provided

File	Preferred Format
Datafile(s).csv	csv
Reports_Or_Documentation.pdf	PDF-A
Folder_Readme.txt	ASCII text
Format_Readme.txt	ASCII text
Metadata information	Submit via on-line form

- 5) We will check with you that we have received all the data files
- 6) We will load the data into the archive
- 7) We will create the Metadata files for the deposit
- 8) We will publish the Metadata files onto the Energy Data Centre website (at which time the data will be available to data users)

## LINKING TO DATA

If you have a data set already published online elsewhere that you would like us to link to, please tell us about it by completing our online '[Submit a Link or Dataset](http://ukerc.rl.ac.uk/DC/cgi-bin/edc_submit.pl)' form at [http://ukerc.rl.ac.uk/DC/cgi-bin/edc\\_submit.pl](http://ukerc.rl.ac.uk/DC/cgi-bin/edc_submit.pl).