



## 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 management and archiving 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 We actively manage data so that storage media and formats remain readable and don't go out of date.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. Data can be made accessible for the long-term and to people who may not have been involved in the original collection.

#### 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 straightforward 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. This is because we expect CSV text format to be accessible in the long-term. 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.

File names 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 yyyymmdd.

If there is a filename extension convention for the file format that you are using, 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 [*Note : PDF/A files can be created by doing a “Save as” from Word (or Excel), selecting “Options” and then selecting the “ISO 9005-1 compliant (PDF/A)” option.*]

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. We recommend that 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. Think about what is logical for your data and what will make most sense to someone not involved in the original data collection.

## 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](https://ukerc.rl.ac.uk/DC/cgi-bin/edc_submit.pl)’ at [https://ukerc.rl.ac.uk/DC/cgi-bin/edc\\_submit.pl](https://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 [https://ukerc.rl.ac.uk/DC/cgi-bin/edc\\_search.pl](https://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.

NB The metadata description is dynamic and can be updated at any time – for example as more data is added to the data deposit, or when a DOI is issued.

Examples of published metadata descriptions can be seen at [http://ukerc.rl.ac.uk/DC/cgi-bin/edc\\_search.pl?WantComp=133](http://ukerc.rl.ac.uk/DC/cgi-bin/edc_search.pl?WantComp=133) and [http://ukerc.rl.ac.uk/DC/cgi-bin/edc\\_search.pl?WantComp=137](http://ukerc.rl.ac.uk/DC/cgi-bin/edc_search.pl?WantComp=137)

## HOW TO DEPOSIT YOUR DATA WITH THE EDC

**We strongly encourage you to get in touch with us as early as possible in a project, preferably in the planning / funding request stage so that we can help you plan data management and deposit in to your project.**

- 1) Get in touch with us by emailing [EDCManager@stfc.ac.uk](mailto:EDCManager@stfc.ac.uk) . We will contact you to find out more about your data and ask you to complete a Data Management Plan <link to DMP pro-forma>. **NB We can offer advice and help throughout.**
- 2) We will then review your planned deposit to ensure that we can understand the data set. We may suggest changes to the layout or description.
- 3) 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](#)' 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
Data licence (if not Open Government)	URL or PDF-A

- 4) We will check with you that we have received all the data files
- 5) We will load the data into the archive
- 6) We will create the metadata files for the deposit
- 7) 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 HELD ELSEWHERE

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](https://ukerc.rl.ac.uk/DC/cgi-bin/edc_submit.pl)' form at [https://ukerc.rl.ac.uk/DC/cgi-bin/edc\\_submit.pl](https://ukerc.rl.ac.uk/DC/cgi-bin/edc_submit.pl).

#### IN-PROJECT DATA MANAGEMENT APPROACH

Until the data are supplied to the UKERC Energy Data Centre, you should consider how the data will be stored securely. General advice includes:

- Keep updated anti-virus protection on every computer.
- Use appropriate descriptive file names to avoid over-writing or mixing up results.
- Record all relevant metadata at the same time as the data - it can be hard to remember specifics after the event.
- Record the original creation date and time for files on your systems.
- Version control - keep track of authorship and changes made to data files.
- Physical lab notebooks should be stored in a safe place and digital lab books should be backed up.
- Data files should be backed up regularly and the backup data stored in a secure place physically removed from the original data.
- Samples should be appropriately saved so they will not degrade over time.

The use of robust, managed storage provided by university or business IT teams is preferable. Similarly, it is normally better to use automatic backup services provided by IT Services than rely on manual processes. If you choose to use a third-party service, you should ensure that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.