How to Package Data for DREAM (0.2.1)

Labelling of data

Your data consists of polysomnography (PSG) and subjects' dream, mentation, or experience reports following awakening from the same period of sleep. Treat each awakening as one datum, consisting of both the PSG and the associated experience report.

Each datum must be labelled with a unique *Case ID* and a *Subject ID*. The *Subject ID* specifies which subject the datum came from and must be unique to each subject. *Subject ID* can be no longer than 64 characters, and *Case ID* no longer than 54 characters.

If your sleep study makes use of experimental or treatment groups, then each datum should also be labelled with their corresponding *Treatment group*.

These labels will be recorded in an included records table, which you can read about in a later section.

PSG files

When formatting your PSG files to be added to the DREAM, ensure that each awakening case is exported as exactly one EDF or EDF+ file with the *.edf* file extension.

This single EDF(+) file should contain all EEG, EOG, etc., annotations, and hypnograms of the sleep period up to the time of awakening.

The endpoint of all PSGs should be just before the subject is awakened, or otherwise noted. Do **not** include the period when they actually awaken.

Make sure to use EDF+ standard texts where possible for channels and annotations.

Header information

EDF(+) files include information about the recording in its header. You must ensure no personally identifiable information is included in the header. In addition, you should format the header to provide metadata matching the datum's labels. This requirement may become automated in the future.

If you are having trouble dealing with the EDF(+), please let us know.

The following header fields should be formatted as shown:
- *Local patient identification*
  - *Patient code*: Give the *Subject ID* label
  - *Sex*: F, M, or X (if omitted). Other labels may be used if consistent.
  - *Birthdate*: Should always be X
  - *Patient name*: Give the *Case ID* label instead

Note that any spaces used must be replaced with an underscore.

Package directory structure

The overall package of your data set will have the following minimum directory structure and files:
- Data <dir>
  - PSG <dir>
  - Reports <dir>
- Records.csv
- ExperimentalDescription.txt
- README.txt

Those three files in the root directory will be explained later. A template is provided.

/Data/PSG directory

Your PSG files should all be contained in the **/Data/PSG** directory. Within this, they are allowed to have any valid subdirectory organisation and filename, provided that no file path to any PSG is a substring of any other PSG's (excluding file extensions).

For example, **/Data/PSG/1.edf** cannot coexist with **/Data/PSG/10.edf** because the latter begins with "/Data/PSG/1", which matches the former. This can be avoided by padding the first filename with a leading zero ("01.edf").

For the same reason, **/Data/PSG/foo.edf** cannot coexist with **/Data/PSG/foobar/01.edf**, even though the files are in different directories.

This limitation is due to how associated experience reports and other data files will be searched for based on the PSG filename.

/Data/Reports directory

We require you to classify your sleep experience reports according to the DREAM classifications and provide them in the records table (see below), however we also strongly encourage you to provide more detailed or raw reports if possible. Those data should be placed in **/Data/Reports**. If you choose to not give separate data files for those collected reports, **/Data/Reports** may remain empty, but it must not be omitted entirely.

If your collected reports are consolidated into a single data file, then name the file beginning with

"Reports" and place it in the "Data/" directory (e.g. **/Data/Reports.xlsx**); and leave **/Data/Reports** empty.

Otherwise, you must give the experience report data as one or more files for each datum in the **/Data/Reports** directory. Those data files must begin with the same file path as the corresponding PSG files relative to **/Data/PSG** (excluding file extensions).

For example, a datum with a PSG file path of **/Data/PSG/01.edf** can have an experience report file with path **/Data/Reports/01.txt**. It can have multiple files, such as **/Data/Reports/01.xml** and **/Data/Reports/01_translated.txt**. It can even be a subdirectory of files, such as **/Data/Reports/01/***.

Other directories

You may include other data associated with each datum, but that do not count as PSG nor reports (e.g., stimuli, personality tests). To do so, you should create a new subdirectory within **/Data** and organise those data files with the same naming scheme as in **/Data/Reports** above. Also as above, they may alternatively be consolidated into one data file, which should be placed in **/Data**.

An example datum may consist of the following data files: **/Data/PSG/01.edf**, **/Data/Reports/01.txt** and **/Data/Stimuli/01.wav**.

Records table

Your data set package will include a records table that provides accurate metadata about every datum. Please adhere to the given format below, as it will make parsing of your data easier for everyone.

The table will be a comma-separated values (CSV) file – a simple file format compatible with all good spreadsheet editors – with the filename "Records.csv". It has the following 19 columns:

Order	Header name	Value description
1	Filename	Filename of this sample's PSG including the directory path relative to (but not including) the **/Data/PSG** directory; should not include the leading slash
2	Case ID	Unique ID of this sample within the data set
3	Subject ID	Unique ID of this sample's subject; it should match the patient code subfield of the "local patient identification" field of the sample EDF header
4	Experience	The sleep experience reported for this sample (key: 2=experience, 1=without recall, 0=no experience)
5	Treatment group	Unique ID of the condition or treatment group that this sample belongs to; must be documented in the "ExperimentalDescription.txt" file of the data set if used
6	Duration	The duration of the PSG in seconds
7	EEG sample rate	The sampling rate of the EEG in Hertz, not including the unit

8	Number of EEG channels	The number of EEG signals in this sample
9	Last sleep stage	The scored sleep stage of the final epoch in the sample (key: 1=N1, 2=N2, 3=N3/NREM3/NREM4, 5=REM)
10	Has EOG	Whether EOG is included in the PSG (key: 0=no, 1=yes)
11	Has EMG	Whether EMG is included in the PSG (key: 0=no, 1=yes)
12	Has ECG	Whether ECG is included in the PSG (key: 0=no, 1=yes)
13	Proportion artifacts	The proportion of signal data in the EEG that contain obvious artifacts
14	Time of awakening	Time when this sample's PSG ends; leave blank if not known to within 3-hour's precision
15	Subject age	Age of this sample's subject
16	Subject sex	Sex of this sample's subject (key: 0=male, 1=female, 2=other)
17	Subject healthy	Whether the sample's subject is from a relatively healthy population (key: 0=no, 1=yes)
18	Has more data	Whether this sample has more data in the form of files under the **/Data** directory other than the **/Data/PSG** directory (key: 0=no, 1=yes)
19	Remarks	Optional field for remarks

Make use of the records table included in our data set package template.

Experimental description

Information about your data set should be documented in a plain text file named "ExperimentalDescription.txt", located in the root directory. Please refer to the template or use the experimental description tool provided by us.

You should describe your study in as much detail as necessary to allow others to replicate it. If this information cannot be adequately conveyed in plain text, you may substitute or supplement the experimental description with: another file whose filename begins with "ExperimentalDescription" (e.g. "ExperimentalDescription.pdf"), placed in the root directory; or multiple files under a directory named "ExperimentalDescription/", also placed in the root directory. If you include these additional files, please do not omit the "ExperimentalDescription.txt" file, but note within it where these files are located.