

Data Description in SENSORWeb

We require a data description document. This description shall describe the data which is stored on the MCS. It is required for each sensor URN which has data stored. The complexity of such a data description depends on the complexity of your data. The data description document shall be uploaded to your sensor in sensor.awi.de, e.g. as a pdf file. Please use the resources tab and use "Data Description" as document type. We require this to prepare the raw data publication in PANGAEA for your proofread, to ensure all the data and metadata, including the authorship and contact information is correct. We will not publish your data open access without prior contacting you. As a default, the open access publication date is 2023-01-01. You can instruct us to release the data earlier.

- Why is a Data Description required?
 - To prepare the PANGAEA publication. Examples: <https://doi.pangaea.de/10.1594/PANGAEA.919999?format=html#download>
 - To identify and find the relevant data for the publication besides all other data of your sensor on the mass storage.
- What shall it contain? (This depends highly on your sensors data.)
 - If your sensor produces only one file type and files are all stored in your sensors root directory under **platforms/urn/exdata/**, there are no duplicates and all existing files shall be published, then then we require the following:
 - A description how we can extract date and time information from your filenames
 - A data format description, if it is a non-standard data format. This can be a description or a persistent link to a format description or a software which can be used to open and process this data.
 - If your sensor produces more than one file type we additionally need:
 - A file name description for how to distinguish between the different files. (E.g. file endings, prefix in filename, ...)
 - If there are directories below the sensors root directory
 - A description if we shall archive files from all directories or if to discard certain folders
 - A description if we can extract information such as Expedition, Event, Date and Time from the directory name
- What is it used for? We use this data description to create an Ingest template. Your descriptions are used to create Regular Expressions to match your files in your directories. We appreciate you to provide us with such regex repressions directly in your data description if you are familiar with them. Your data description is translated into such a template:

```

    "_ingest": {
      "_sourceRegex": ".*(?P<campaign>PS[0-9]{2,3})",
      "_columns": [
        {
          "regex": "^((Screenshots|screenshots))\\\\.\\.\\.*\\.*(JPG|jpg|PNG|png))",
          "column": "Binary Object []",
          "comment": "Screenshots",
          "description": ""
        },
        {
          "regex": "^PHF_ASD_[1-2][0-9][0-9][0-9][0-1][0-9][0-3][0-9].*\\./HS3PHF_(?P<year>[1-2][0-9][0-9][0-9])-(?P<month>[0-1][0-9])-(?P<day>[0-3][0-9])T(?P<hour>[0-2][0-9])(?P<minute>[0-5][0-9])(?P<second>[0-5][0-9]).*Z_[0-9]*\\.\\.asd",
          "column": "Binary Object []",
          "comment": "PHF_ASD files",
          "description": ""
        },
        {
          "regex": "^PHS_ASD_[1-2][0-9][0-9][0-9][0-1][0-9][0-3][0-9].*\\./HS3PHS_(?P<year>[1-2][0-9][0-9][0-9])-(?P<month>[0-1][0-9])-(?P<day>[0-3][0-9])T(?P<hour>[0-2][0-9])(?P<minute>[0-5][0-9])(?P<second>[0-5][0-9]).*Z_[0-9]*\\.\\.asd",
          "column": "Binary Object []",
          "comment": "PHS_ASD files",
          "description": ""
        },
        {
          "regex": "^S7K_[1-2][0-9][0-9][0-9][0-1][0-9][0-3][0-9].*\\./(?P<year>[1-2][0-9][0-9][0-9])(?P<month>[0-1][0-9])(?P<day>[0-3][0-9])_(?P<hour>[0-2][0-9])(?P<minute>[0-5][0-9])(?P<second>[0-5][0-9])_.*\\.*(s7k|S7K))$",
          "column": "Binary Object []",
          "comment": "RESON-S7K files",
          "description": ""
        }
      ]
    }
  }
}

```

- the key "_sourceRegex" can be used to extract campaign or event information from the directory. This is important if you wish to publish data from several campaigns (expedition legs) or events together. The following regex groups are supported by our framework: "campaign", "leg", "science_operation", "device_operation"
 - each element in the list below "columns" : [describes one file type. "regex": provides the regular expression to match this certain file. The following regex groups are supported to extract information here: "year", "month", "hour", "minute", "second"
- Finally this template is used to create the data table for the PANGAEA publication page and to archive the listed files accordingly. The date time information from the file can be used to provide a georeference based on the campaign's mastertrack if desired.

1 ⓘ Date/Time	2 ⓘ Latitude	3 ⓘ Longitude	4 ⓘ URL raw (raw format, see hdl:10013/epi...)	5 ⓘ URL raw (idx format)
2019-12-13T08:21:52			Link	Link
2019-12-13T08:31:38			Link	Link
2019-12-13T08:41:44			Link	Link
2019-12-13T08:51:30			Link	Link
2019-12-13T09:01:15			Link	Link
2019-12-13T09:11:01	86.5932320	119.2369840	Link	Link
2019-12-13T09:20:47	86.5933160	119.2296640	Link	Link
2019-12-13T09:30:25	86.5933950	119.2227150	Link	Link
2019-12-13T09:40:11	86.5934860	119.2159130	Link	Link
2019-12-13T09:49:57	86.5935890	119.2090590	Link	Link