Exchange-format for oceanographic auxiliary data

Hosting continuous sound data requires storage and organization of large amounts of data. The data formats HDF5 and NetCDF are both well suited for this purpose and technically widely supported. There is a high level of compatibility between both formats and an abundance of technical support for format conversion and write/read support.

The HDF5 format provides objects called *groups*, *datasets* and *attributes*. A *group* is comparable to a folder in a file system. *Datasets* can be e.g. matrices or single values/strings. *Attributes* can be used to store metadata of datasets. Resources in HDF5 files can be accessed using a POSIX-like syntax e.g. */filename/group/specific_resource*.

For a lean exchange format definition, the hdf5-exchange-format can be composed by *groups* and *datasets* only:

- Group: a container structures composed of datasets and other groups
- Dataset: single value/multidimensional arrays of a homogeneous type

For more information on the hdf5 format and supporting software follow the link www.hdfgroup.org.

The following format definition describes the suggested hierarchy in the HDF-file and specifies the suggested datatype (e.g. int, float, string and bool).
All **bold** names followed by a <HDF dataset …> are *datasets*. All names followed by further **names** are *groups*. The highest parent node is the hdf-file in the file system "result_file.h5":

{Examples are in orange!}

File naming convention:
The exchange file name has to be:
'***stationshortcut_aux_beginofmeasurement_endofmeasurement.h5'***
Example: ***'06-DE-FN1_aux_20190101_20190131.h5'***

Dataset length per file:
File size should be one month only.
Also use a single file for every different month and measurement station.

Country	Name	Position
Sweden	Vinga	**01-SE-VIN**
Denmark	Anholt	**02-DK-ANH**
Denmark	Horns Reef	**03-DK-HRF**
Germany	FINO3	**04-DE-FN3**
Germany	ES01	**05-DE-ES1**
Germany	FINO1	**06-DE-FN1**
Netherlands	West of Texel	**07-NL-TEX**
Belgium	Westhinder	**08-BE-WST**
England	Dowsing	**09-UK-DOW**
Scotland	Arbroath 10	**10-SC-ARB**
Scotland	Helmsdale 5	**11-SC-HEL**
Scotland	Moray Firt	**12-SC-MOR**
Norway	LoVe	**13-NO-LOV**
Norway	Norwegian Trench	**14-NO-NTR**
Scotland	Central North Sea	**15-SC-CNS**

result_file.h5/

author # creator of the HDF5 file, responsible for evaluations
MANDATORY
<HDF dataset, type string>
{'Jens-Georg Fischer; jens.fischer@bsh.de'}

comments
OPTIONAL
<HDF dataset, type string>

date_of_creation # of this file
MANDATORY
<HDF dataset, type int>
{'20190131'} for the 31st of January 2019

name_measurement_position # see Table on page 1, name of JOMOPANS measurement station near which aux data was collected
OPTIONAL
<HDF dataset, type string>
{'01-SE-VIN'}

name_measurement_project
OPTIONAL
<HDF dataset, type string>
{'JOMOPANS'}

point_of_contact # contact for all external queries in the future
MANDATORY
<HDF dataset, type string>
{'BSH; department M23…'}

rawdata_uuid # generate a unique version 4 uuid (random) for each dataset version – matlab function available:
uuid = char(java.util.UUID.randomUUID);
MANDATORY
<HDF dataset, type string>
e.g. {' 0bc179e4-e533-4fd3-ae6c-affd24f86f86'}

/impulsive_noise # description of permanent and temporal impulsive noise activities near (<50km) measurement area (piling, seismic, hydrographic surveying, …) + Indication of days when these activities occurred.

impnoise_event # description of event: when, where, what
<HDF dataset, type string>
OPTIONAL
e.g. {'pile driving at windpark xyz from 2019-03-25 13:30:00 till 2019-03-28 16:45:00'}

/SSP_1 # sound speed profile during deployment and recovery of station or during inspection

ctd_profile # in case the sound speed profile was derived from ctd measurement please state the raw ctd profile here
OPTIONAL
<HDF dataset, shape (3,), type float>
e.g. for 'pressure | conductivity | temperature'
{5.8, 42.0, 4.3;
6.8, 47.2, 4.2;…}

ctd_units # in case the sound speed profile was derived from ctd measurement (see above) please state the used units for ctd_profile
OPTIONAL (MANDATORY if ctd_profile was used)
<HDF dataset, type string>
e.g. {dBar, S/m, °C}

ctd_datetime_index # starttime CTD profile was taken, format 'yyyymmddHHMMSS' in UTC+0
OPTIONAL (MANDATORY if ctd_profile was used)
<HDF dataset, type int>
e.g. {20180731121535} for the 31st of July 2018 12:15:35

ssp # sound speed in m/s over depth in m
MANDATORY
<HDF dataset, shape (2,), type float>
e.g. {1500.1, 2.1; 1500.2, 2.2;…}

ssp_method # either derived or measured directly with a SVP (sound velocity probe)
MANDATORY
<HDF dataset, type string>
e.g. {'derived from CTD'} or {'measured by SVP'}

/SSP_2 # if more than 1 ssp was taken per month, please use another subfolder

/tidal # tidal height as a function of time near JOMOPANS measurement station

coordinates_tidal # coordinates of tidal measurement
OPTIONAL (MANDATORY if tidal_height was used)
<HDF dataset, shape (2,), type float
(latitude, longitude) in Decimal degrees WGS84, 6 decimals>
{'56.926667 11.202333'}

tidal_height # in metres above (LAT) (hourly mean) over time (format 'yyyymmddHHMMSS' in UTC+0)
OPTIONAL
<HDF dataset, shape (,2), type float>
e.g. {20180731121500, 2.0; 20180731131500, 3.1;…}

/waves # wave parameters, hourly mean

waves_coordinates_measurement_position # wave measurement station coordinates
OPTIONAL
<HDF dataset, shape (2,), type float
(latitude, longitude) in Decimal degrees WGS84, 6 decimals>
{‘56.926667 11.202333’}

wave_datetime_index # format ‘yyyymmddHHMMSS’ in UTC+0
OPTIONAL
<HDF dataset, shape(), type int>
e.g. {20180731121535} for the 31st of July 2018 12:15:35

wave_dir # mean wave direction per hour in degree, 0°/360° (vector needs to have same length as “wave_datetime_index”)
OPTIONAL
<HDF dataset, shape(), type int>

wave_hs # sign. wave height in metres per hour (vector needs to have same length as “wave_datetime_index”)
OPTIONAL
<HDF dataset, shape(), type float>

wave_tm # wave period in seconds per hour (vector needs to have same length as “wave_datetime_index”)
OPTIONAL
<HDF dataset, shape(), type float>

/wind # wind parameters, averages per hour, 10m above sea level

wind_coordinates_measurement_position # wave measurement station coordinates
OPTIONAL
<HDF dataset, shape (2,), type float
(latitude, longitude) in Decimal degrees WGS84, 6 decimals>
{‘56.926667 11.202333’}

wind_datetime_index # format ‘yyyymmddHHMMSS’ in UTC+0
OPTIONAL
<HDF dataset, shape(), type int>
e.g. {20180731121535} for the 31st of July 2018 12:15:35

wind_dir # mean wind direction per hour in degree, 0°/360° (vector needs to have same length as “wind_datetime_index”)
OPTIONAL
<HDF dataset, shape(), type int>

wind_speed # wind speed in metres per second - averaged per hour (vector needs to have same length as “wind_datetime_index”)
OPTIONAL

```
<HDF dataset, shape(), type float>
```