Exchange-format for underwater sound monitoring 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_beginofmeasurement_endofmeasurement.h5'***
Example: ***'06-DE-FN1_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'}

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

measuring_institution # institution, which acquired the data
MANDATORY
<HDF dataset, type string>
{'BSH; jon.snow@bsh.de'}

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

/dataset_ambient_noise # group for better overview
MANDATORY

averaging_time # avg. time in seconds
MANDATORY
<HDF dataset, type float>
{'1.0'} or {'20.0'} etc.

/calibration # for details on this refer to Jomopans_MEASUREMENTS_CALIBRATION_v2.docx

calibration_file # assignment of calibration procedure to recorded calibration signals (.wav files)
OPTIONAL
<HDF dataset, type string>
{example: filename, calibration procedure
AA000.wav, pistonphon …}

calibration_procedure # method used to check the measuring chain. e.g. point calibration with pistonphone, functionality test with microphone and loudspeaker (frequency dependent), or other.
MANDATORY
<HDF dataset, type string>
either {'pistophone'}, {'frequency dependent'} or {'other'}

other_calibration_method # please describe the used method of your calibration
OPTIONAL
<HDF dataset, type string>

e.g. {‘sent our instruments to company xy to get the calibration certificate…’} or {‘went to a quiet lake to…’} or {‘laboratory ring test to…’}

reference_frequencies_levels # frequencies in Hz and reference sound pressure levels in dB re 1 µPa, point calibration: data from used pistonphone: 250.0 156.2 (1row array with 2 columns), frequency dependent: e.g. array with third band frequencies and corresponding levels (frequency count rows array with 2 columns).
OPTIONAL
<HDF dataset, type float>
{frequency(ies) | level(s)}

comments
OPTIONAL
<HDF dataset, type string>

construction_design # description of deployment construction
MANDATORY
<HDF dataset, type string>
either {‘bottom frame’}, {‘mooring with floating buoy’} or {‘other’}

coordinates_measurement_position # station coordinates in decimal degrees WGS84, 6 decimals
MANDATORY
<HDF dataset, type float(latitude, longitude)>
{‘56.926667, 11.202333’}

count # number of measurement values
MANDATORY
<HDF dataset, type int>
{‘3295800’}

dataset_type
MANDATORY
<HDF dataset, type string>
{‘ambient_noise’}

dataset_version
MANDATORY
<HDF dataset, type float>
e.g. {‘2’}, where ‘2’ indicates the version of the submitted dataset (how often you needed to send it to us)

datetime_index # format ‘yyyymmddHHMMSS’ in UTC+0
MANDATORY
<HDF dataset, shape(count,1), type int>
e.g. {20180731121535} for the 31st of July 2018 12:15:35

device_manufacturer # recording and battery unit
MANDATORY
<HDF dataset, type string>
{‘WILDLIFE’}

device_serial_number
MANDATORY
<HDF dataset, type string>
{‘SN758’}

device_type
MANDATORY
<HDF dataset, type string>
{‘SM2M’}

duty_cycle # description of one hour, one entry per minute, which can be either on (1) or off (0)
MANDATORY
<HDF dataset, shape(60,1), type int>
e.g. {‘1,1,1,1,1,0,0,0,0,0,1,1,1,1,1,…’} for alternating 5 minutes on and off

frequency_count # number of frequency bands, please use ‘34’ for every JOMOPANS delivery
MANDATORY
<HDF dataset, type int>
fixed {‘34’}

frequency_index # center frequency bands, please use all 34 predefined freq. bands for every file, fill unused with NANs in 'spectral_temporal_stats' and 'spectral_temporal_values'

MANDATORY

<HDF dataset, shape (frequency_count,1), type float>

fixed, see table under 'Centre frequency'

Table A1 One third octave frequency bands

Band index	Lower bound	Centre frequency	Upper bound	Nominal centre frequency
n	f_{min}/Hz	f_c/Hz	f_{max}/Hz	$f_{c,norm}$/Hz
-20	8.9125	10	11.22	10
-19	11.22	12.589	14.125	12.5
-18	14.125	15.849	17.783	16
-17	17.783	19.953	22.387	20
-16	22.387	25.119	28.184	25
-15	28.184	31.623	35.481	31.5
-14	35.481	39.811	44.668	40
-13	44.668	50.119	56.234	50
-12	56.234	63.096	70.795	63
-11	70.795	79.433	89.125	80
-10	89.125	100	112.2	100
-9	112.2	125.89	141.25	125
-8	141.25	158.49	177.83	160
-7	177.83	199.53	223.87	200
-6	223.87	251.19	281.84	250
-5	281.84	316.23	354.81	315
-4	354.81	398.11	446.68	400
-3	446.68	501.19	562.34	500
-2	562.34	630.96	707.95	630
-1	707.95	794.33	891.25	800
0	891.25	1000	1122.0	1000
1	1122.0	1258.9	1412.5	1250
2	1412.5	1584.9	1778.3	1600
3	1778.3	1995.3	2238.7	2000
4	2238.7	2511.9	2818.4	2500
5	2818.4	3162.3	3548.1	3150
6	3548.1	3981.1	4466.8	4000
7	4466.8	5011.9	5623.4	5000
8	5623.4	6309.6	7079.5	6300
9	7079.5	7943.3	8912.5	8000
10	8912.5	10000	11220	10000
11	11220	12589	14125	12500
12	14125	15849	17783	15000
13	17783	19953	22387	20000

hydrophone_decoupling # description whether hydrophone is decoupled from recording/battery unit or not

MANDATORY

<HDF dataset, type string>

either {'yes'} or {'no'}

hydrophone_manufacturer

MANDATORY

<HDF dataset, type string>

e.g. {'B&K'}

hydrophone_sensitivity # in dB re 1 µPa

MANDATORY

<HDF dataset, type float>

e.g. {'-160.1'}

hydrophone_serial_number

MANDATORY

<HDF dataset, type string>

e.g. {'SN45736'}

hydrophone_type
MANDATORY
<HDF dataset, type string>
e.g. {‘8106’}

measurement_height # in meters, height above ground
MANDATORY
<HDF dataset, type float>
e.g. {‘3.0’}

measurement_purpose
MANDATORY
<HDF dataset, type string>
{‘Research and Development’}

measurement_setup # description of deployment
MANDATORY
<HDF dataset, type string>
e.g. {‘autonomous’} or {‘cable mounted’}

name_measurement_position # see table on page 1
MANDATORY
<HDF dataset, type string>
{‘01-SE-VIN’}

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

rawdata_timestamp # assignment of spectral_temporal_stats to recorded signals (WAV files), timestamp format ‘yyyymmddHHMMSS’ in UTC+0
OPTIONAL, NOT NECESSARY FOR JOMOPANS
<HDF dataset, type string>
e.g. filename | start time | end time
{FINO1_20190615030000.wav | 20190615030000 | 20190615031500…}

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-affd24f86f85’}

spectral_analysis_tool # tool used for conversion from temporal to spectral domain – to create the spectral temporal values
MANDATORY
<HDF dataset, type string>
{‘JOMOPANS_processing_function.m, version 2.3’}

\spectral_temporal_stats # percentiles with 34 values per entry, in dB re 1 µPa; fill unused freq. bands with NANs

MANDATORY

L01

<HDF dataset, shape (34,1), type float>

{'98.5 97.9 95.1 92.7…'}

L05

L10

L25

L50

L75

L90

L95

L99

LMin

LMax

spectral_temporal_values # matrix with 34 frequency rows and 'count' number of columns; fill unused freq. bands with NANs

MANDATORY

<HDF dataset, shape (34,count), type float>

water_depth # in meters, depth must be related to LAT

MANDATORY

<HDF dataset, type float>

e.g. {'55.0'}

water_depth_method # how/when was water depth derived?

MANDATORY

<HDF dataset, type string>

e.g. {'Measured by echosounder and related to LAT on 12th of September 2018'} or {'derived from position and nautical chart on 12th of September 2018'} or {'by pressure sensor and related to LAT on 12th of September 2018'}…