Moberg Analytics HDF5 Documentation

Author: Zack Goldblum - Moberg Analytics

Contents

1 Overview **1**

1.1 Example HDF5 File Structure 2

1.2 Revision History 2

2 HDF5Content **2**

2.1 Group Functions 3

2.2 Dataset Functions 7

2.3 Misc. Functions 10

3 HDF5Components **15**

3.1 Group Functions 16

3.2 Dataset Functions 20

4 HDF5Helper **23**

4.1 Group Functions 24

4.2 Dataset Functions 25

1 Overview

This documentation details how to use the functions available in the Moberg-Analytics-HDF5 package. The Moberg-Analytics-HDF5 package provides user-friendly functions organized into classes for reading HDF5 file content and components into Python. It is built on top of the h5py package which interfaces directly with the HDF5 file.

The three **HDF5Content**, **HDF5Components**, and **HDF5Helper** sections in the left-hand navigation bar correspond to the three classes within the hdf5_tools module:

- **HDF5Content** contains functions that organize the contents of the HDF5 file into lists and dictionaries.
- **HDF5Components** contains functions that return various components of the HDF5 file to the user including groups, datasets, Pandas/NumPy matrices of dataset values, metadata, and structured dictionaries.
- **HDF5Helper** contains functions for argument, group, dataset, and duplicate checks as well as other methods that add functionality to HDF5Content and HDF5Components.

Each class section is further divided into **Group Functions**, **Dataset Functions**, and **Misc. Functions** sections that contain the relevant functions for working with groups, datasets, and other aspects of the HDF5 file. Every function has a description that details what it does, the parameters it accepts (if any), and what it returns. There are also code examples that demonstrate how the function is called and show what it returns. All of the code examples use the example.h5 HDF5 file, the structure of which is shown below.

1.1 Example HDF5 File Structure

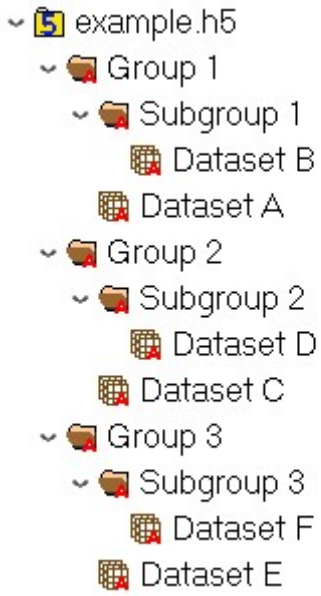

Figure 1: example.h5 in HDFView 3.1.1

1.2 Revision History

Date	Revision Number	Description
01/08/2021	1.0	Documentation created.
03/05/2021	2.0	Updated for package release.

2 HDF5Content

DESCRIPTION:

This class contains functions that organize the contents of the HDF5 file into lists and dictionaries.

PARAMETERS:

- **hdf5_filepath**: HDF5 file path
  path to the user-selected HDF5 file

ATTRIBUTES:

- **hdf5file**: HDF5 file
  user-selected HDF5 file
- **all_group_names_dict**: dict
  dictionary of every group and its associated info. Group names are keys.
- **all_dataset_names_dict**: dict
  dictionary of all datasets and their associated info. Datasetset names are keys.
- **all_dataset_paths_dict**: dict
  dictionary of all datasets and their associated info. Datasetset paths are keys.

EXAMPLE:

```
# Path to the HDF5 file in your system directory
hdf5_filepath = r"resources\example.h5"

# Instantiate the HDF5Content class
hdf5_content = hdf5_tools.HDF5Content(hdf5_filepath=hdf5_filepath)
```

2.1 Group Functions

2.1.1 get_all_group_paths

get_all_group_paths()

DESCRIPTION:

Returns a list of all group paths in the HDF5 file (including the Root group and subgroups).

PARAMETERS:

none

RETURNS:

- **all_group_paths_list**: list
  list of all group paths

EXAMPLE:

```
all_group_paths_list = hdf5_content.get_all_group_paths()
```

```
## ['/',
##  'Group 1',
##  'Group 1/Subgroup 1',
##  'Group 2',
##  'Group 2/Subgroup 2',
##  'Group 3',
##  'Group 3/Subgroup 3']
```

2.1.2 get_all_group_objs

get_all_group_objs()

DESCRIPTION:

Returns a list of all group class objects in the HDF5 file (including the Root group and subgroups).

PARAMETERS:

none

RETURNS:

- **all_group_objs_list**: list
  list of all HDF5 group class objects

EXAMPLE:

```
all_group_objs_list = hdf5_content.get_all_group_objs()
```

```
## [<HDF5 group "/" (3 members)>,
##  <HDF5 group "/Group 1" (2 members)>,
##  <HDF5 group "/Group 1/Subgroup 1" (1 members)>,
##  <HDF5 group "/Group 2" (2 members)>,
##  <HDF5 group "/Group 2/Subgroup 2" (1 members)>,
##  <HDF5 group "/Group 3" (2 members)>,
##  <HDF5 group "/Group 3/Subgroup 3" (1 members)>]
```

2.1.3 get_all_group_names

get_all_group_names()

DESCRIPTION:

Returns a list of all group names in the HDF5 file (including the Root group and subgroups).

PARAMETERS:

none

RETURNS:

- **all_group_names_list**: list
  list of all group names

EXAMPLE:

```
all_group_names_list = hdf5_content.get_all_group_names()
```

```
## ['/', 'Group 1', 'Subgroup 1', 'Group 2', 'Subgroup 2', 'Group 3', 'Subgroup 3']
```

2.1.4 get_all_group_names_dict

get_all_group_names_dict()

DESCRIPTION:

Returns a dictionary of all group names (including the Root group and subgroups) and their associated info from the HDF5 file.

PARAMETERS:

none

RETURNS:

- **all_group_names_dict**: dict
  dictionary of all group names and their associated info

EXAMPLE:

```
all_groups_dict = hdf5_content.get_all_group_names_dict()
```

```
## {'/': {'group_obj': <HDF5 group "/" (3 members)>,
##        'group_path': '/',
##        'subgroups': {'Group 1': <HDF5 group "/Group 1" (2 members)>,
##                      'Group 2': <HDF5 group "/Group 2" (2 members)>,
##                      'Group 3': <HDF5 group "/Group 3" (2 members)>},
##        'datasets': {}},
##  'Group 1': {'group_obj': <HDF5 group "/Group 1" (2 members)>,
##              'group_path': 'Group 1',
##              'subgroups': {'Subgroup 1': <HDF5 group "/Group 1/Subgroup 1" (1 members)>},
##              'datasets': {'Dataset A': <HDF5 dataset "Dataset A": shape (1, 5), type "<i4">}},
##  'Subgroup 1': {'group_obj': <HDF5 group "/Group 1/Subgroup 1" (1 members)>,
##                 'group_path': 'Group 1/Subgroup 1',
##                 'subgroups': {},
##                 'datasets': {'Dataset B': <HDF5 dataset "Dataset B": shape (1, 5), type "<i4">}},
##  'Group 2': {'group_obj': <HDF5 group "/Group 2" (2 members)>,
##              'group_path': 'Group 2',
##              'subgroups': {'Subgroup 2': <HDF5 group "/Group 2/Subgroup 2" (1 members)>},
##              'datasets': {'Dataset C': <HDF5 dataset "Dataset C": shape (1, 5), type "<i4">}},
##  'Subgroup 2': {'group_obj': <HDF5 group "/Group 2/Subgroup 2" (1 members)>,
##                 'group_path': 'Group 2/Subgroup 2',
##                 'subgroups': {},
##                 'datasets': {'Dataset D': <HDF5 dataset "Dataset D": shape (1, 5), type "<i4">}},
##  'Group 3': {'group_obj': <HDF5 group "/Group 3" (2 members)>,
##              'group_path': 'Group 3',
##              'subgroups': {'Subgroup 3': <HDF5 group "/Group 3/Subgroup 3" (1 members)>},
##              'datasets': {'Dataset E': <HDF5 dataset "Dataset E": shape (1, 5), type "<i4">}},
##  'Subgroup 3': {'group_obj': <HDF5 group "/Group 3/Subgroup 3" (1 members)>,
##                 'group_path': 'Group 3/Subgroup 3',
##                 'subgroups': {},
##                 'datasets': {'Dataset F': <HDF5 dataset "Dataset F": shape (1, 5), type "<i4">}}}
```

2.1.5 get_all_group_objs_dict

get_all_group_objs_dict()

DESCRIPTION:

Returns a dictionary of all group objects (including the Root group and subgroups) and their associated info from the HDF5 file.

PARAMETERS:

none

RETURNS:

- **all_group_objs_dict**: dict
  dictionary of all group objects and their associated info

EXAMPLE:

```
all_group_objs_dict = hdf5_content.get_all_group_objs_dict()
```

```
## {<HDF5 group "/" (3 members)>: {'group_name': '/', 'group_path': '/'},
##  <HDF5 group "/Group 1" (2 members)>: {'group_name': 'Group 1',
##                                        'group_path': 'Group 1'},
##  <HDF5 group "/Group 1/Subgroup 1" (1 members)>: {'group_name': 'Subgroup 1',
##                                                   'group_path': 'Group '
##                                                                 '1/Subgroup 1'},
##  <HDF5 group "/Group 2" (2 members)>: {'group_name': 'Group 2',
##                                        'group_path': 'Group 2'},
##  <HDF5 group "/Group 2/Subgroup 2" (1 members)>: {'group_name': 'Subgroup 2',
##                                                   'group_path': 'Group '
##                                                                 '2/Subgroup 2'},
##  <HDF5 group "/Group 3" (2 members)>: {'group_name': 'Group 3',
##                                        'group_path': 'Group 3'},
##  <HDF5 group "/Group 3/Subgroup 3" (1 members)>: {'group_name': 'Subgroup 3',
##                                                   'group_path': 'Group '
##                                                                 '3/Subgroup 3'}}
```

2.1.6 get_subgroups_list

get_subgroups_list(group_path)

DESCRIPTION:

Returns a list of subgroup names within the group at the "group_path" location in the HDF5 file.

PARAMETERS:

- **group_path**: str
  path of the group in the HDF5 file

RETURNS:

- **subgroup_list**: list
  list of subgroups within a group

EXAMPLE:

```
subgroup_list = hdf5_content.get_subgroups_list(group_path="Group 1")
```

```
## ['Subgroup 1']
```

2.2 Dataset Functions

2.2.1 get_all_dataset_paths

get_all_dataset_paths()

DESCRIPTION:

Returns a list of all dataset paths in the HDF5 file.

PARAMETERS:

none

RETURNS:

- **all_dataset_paths_list**: list
  list of all dataset paths

EXAMPLE:

```
all_dataset_paths_list = hdf5_content.get_all_dataset_paths()
```

```
## ['Group 1/Dataset A',
##  'Group 1/Subgroup 1/Dataset B',
##  'Group 2/Dataset C',
##  'Group 2/Subgroup 2/Dataset D',
##  'Group 3/Dataset E',
##  'Group 3/Subgroup 3/Dataset F']
```

2.2.2 get_all_dataset_objs

get_all_dataset_objs()

DESCRIPTION:

Returns a list of all dataset class objects in the HDF5 file.

PARAMETERS:

none

RETURNS:

- **all_dataset_objs_list**: list
  list of all dataset class objects

EXAMPLE:

```
all_dataset_objs_list = hdf5_content.get_all_dataset_objs()
```

```
## [<HDF5 dataset "Dataset A": shape (1, 5), type "<i4">, <HDF5 dataset "Dataset B": shape (1, 5), type
```

2.2.3 get_all_dataset_names

get_all_dataset_names()

DESCRIPTION:

Returns a list of all dataset names in the HDF5 file.

PARAMETERS:

none

RETURNS:

- **all_dataset_names_list**: list
  list of all dataset names

EXAMPLE:

```
all_dataset_names_list = hdf5_content.get_all_dataset_names()
```

```
## ['Dataset A', 'Dataset B', 'Dataset C', 'Dataset D', 'Dataset E', 'Dataset F']
```

2.2.4 get_all_dataset_paths_dict

get_all_dataset_paths_dict()

DESCRIPTION:

Returns a dictionary of every dataset path and its associated info from the HDF5 file.

PARAMETERS:

none

RETURNS:

- **all_dataset_paths_dict**: dict
  dictionary of every dataset path and its associated info

EXAMPLE:

```
all_dataset_paths_dict = hdf5_content.get_all_dataset_paths_dict()
```

```
## {'Group 1/Dataset A': {'dataset_name': 'Dataset A',
##                        'dataset_obj': <HDF5 dataset "Dataset A": shape (1, 5), type "<i4">},
##  'Group 1/Subgroup 1/Dataset B': {'dataset_name': 'Dataset B',
##                                   'dataset_obj': <HDF5 dataset "Dataset B": shape (1, 5), type "<i4">
##  'Group 2/Dataset C': {'dataset_name': 'Dataset C',
##                        'dataset_obj': <HDF5 dataset "Dataset C": shape (1, 5), type "<i4">},
##  'Group 2/Subgroup 2/Dataset D': {'dataset_name': 'Dataset D',
##                                   'dataset_obj': <HDF5 dataset "Dataset D": shape (1, 5), type "<i4">
##  'Group 3/Dataset E': {'dataset_name': 'Dataset E',
##                        'dataset_obj': <HDF5 dataset "Dataset E": shape (1, 5), type "<i4">},
##  'Group 3/Subgroup 3/Dataset F': {'dataset_name': 'Dataset F',
##                                   'dataset_obj': <HDF5 dataset "Dataset F": shape (1, 5), type "<i4">
```

2.2.5 get_all_dataset_names_dict

get_all_dataset_names_dict()

DESCRIPTION:

Returns a dictionary of every dataset name and its associated info from the HDF5 file.

PARAMETERS:

none

RETURNS:

- **all_dataset_names_dict**: dict
  dictionary of every dataset name and its associated info

EXAMPLE:

```
all_dataset_names_dict = hdf5_content.get_all_dataset_names_dict()
```

```
## {'Dataset A': {'dataset_path': 'Group 1/Dataset A',
##                'dataset_obj': <HDF5 dataset "Dataset A": shape (1, 5), type "<i4">},
##  'Dataset B': {'dataset_path': 'Group 1/Subgroup 1/Dataset B',
##                'dataset_obj': <HDF5 dataset "Dataset B": shape (1, 5), type "<i4">},
##  'Dataset C': {'dataset_path': 'Group 2/Dataset C',
##                'dataset_obj': <HDF5 dataset "Dataset C": shape (1, 5), type "<i4">},
##  'Dataset D': {'dataset_path': 'Group 2/Subgroup 2/Dataset D',
##                'dataset_obj': <HDF5 dataset "Dataset D": shape (1, 5), type "<i4">},
##  'Dataset E': {'dataset_path': 'Group 3/Dataset E',
##                'dataset_obj': <HDF5 dataset "Dataset E": shape (1, 5), type "<i4">},
##  'Dataset F': {'dataset_path': 'Group 3/Subgroup 3/Dataset F',
##                'dataset_obj': <HDF5 dataset "Dataset F": shape (1, 5), type "<i4">}}
```

2.2.6 get_group_dataset_names

get_group_dataset_names(group_name, group_path)

DESCRIPTION:

Returns a list of dataset names within the group at the "group_path" location in the HDF5 file. Does not include datasets within subgroups.

PARAMETERS:

- **group_name**: str
  name of the group in the HDF5 file
- **group_path**: str
  path of the group in the HDF5 file

RETURNS:

- **dataset_name_list**: list
  list of all datasets within the group

EXAMPLE:

```
dataset_name_list = hdf5_content.get_group_dataset_names(group_path="Group 1")
```

```
## ['Dataset A']
```

2.3 Misc. Functions

2.3.1 get_path

get_path(group_name, dataset_name)

DESCRIPTION:

Returns the HDF5 file path to a group or dataset.

PARAMETERS:

- **group_name**: str
  group name to get the path to

-or-

- **dataset_name**: str
  dataset name to get the path to

RETURNS:

- **hdf5_path**: str
  path to the group or dataset location in the HDF5 file

EXAMPLES:

```
hdf5_path = hdf5_content.get_path(group_name="Group 1")
```

```
## Group 1
```

2.3.2 get_metadata

get_metadata(group_name=None, dataset_name=None, dataset_path=None)

DESCRIPTION:

Returns a dictionary of metadata attributes.

PARAMETERS:

- **group_name**: str
  group name to get the metadata of

-or-

- **dataset_name**: str
  dataset name to get the metadata of

-or-

- **dataset_path**: str
  dataset path to get the metadata of

RETURNS:

- **metadata_dict**: dict
  dictionary of metadata attributes

EXAMPLES:

```
group_metadata = hdf5_content.get_metadata(group_name="Group 1")
```

```
## {'group1_attr1': array([123]),
##  'group1_attr2': array([b'test string 1'], dtype='|S14')}
```

```
dataset_metadata = hdf5_content.get_metadata(dataset_name="Dataset A")
```

```
## {'datasetA_attr1': array([345]),
##  'datasetA_attr2': array([b'test string 3'], dtype='|S14')}
```

2.3.3 get_hdf5_filename

get_hdf5_filename(hdf5_filepath)

DESCRIPTION:

Parses the HDF5 file path (on the system) and returns the name of the HDF5 file.

PARAMETERS:

- **hdf5_filepath**: HDF5 file path
  path to the user-selected HDF5 file

RETURNS:

- **hdf5_filename**: str
  name of the HDF5 file

EXAMPLE:

```
hdf5_filename = hdf5_content.get_hdf5_filename(hdf5_filepath=hdf5_filepath)

## example.h5
```

2.3.4 get_name

get_name(hdf5_path)

DESCRIPTION:

Returns the group or dataset name from a HDF5 file path.

PARAMETERS:

- **hdf5_path**: str
  HDF5 path to the group or dataset

RETURNS:

- **name**: str
  name of the group or dataset

EXAMPLE:

```
dataset_name = hdf5_content.get_name(hdf5_path="/Group 1/Dataset A")

## Dataset A
```

2.3.5 split_hdf5_path

split_hdf5_path(hdf5_path)

DESCRIPTION:

Parses an HDF5 group or dataset path and creates a list of the path components.

PARAMETERS:

- **hdf5_path**: str
  HDF5 path to the group or dataset

RETURNS:

- **split_path**: list
  list of path components

EXAMPLE:

```
split_path = hdf5_content.split_hdf5_path(hdf5_path="Group 1/Dataset A")
```

```
## ['Group 1', 'Dataset A']
```

2.3.6 get_dup_dataset_names

get_dup_dataset_names()

DESCRIPTION:

Returns a list of all duplicate dataset names in the HDF5 file.

PARAMETERS:

none

RETURNS:

- **dup_dataset_names_list**: list
  list of all duplicate dataset names

EXAMPLE:

```
dup_dataset_names_list = hdf5_content.get_dup_dataset_names()
```

```
## []
```

2.3.7 get_dup_dataset_dict

get_dup_dataset_dict()

DESCRIPTION:

Returns a dictionary of every duplicate dataset name and its associated info from the HDF5 file.

PARAMETERS:

none

RETURNS:

- **dup_dataset_dict**: dict
  dictionary of every duplicate dataset name and its associated info

EXAMPLE:

```
dup_dataset_dict = hdf5_content.get_dup_dataset_dict()
```

```
## {}
```

2.3.8 dup_dataset_check

dup_dataset_check(dataset_name)

DESCRIPTION:

Returns True if "dataset_name" is a duplicate dataset in the HDF5 file. Otherwise returns False.

PARAMETERS:

- **dataset_name**: str
  name of the dataset to check

RETURNS:

- **True**
  -or-
- **False**

EXAMPLE:

```
dataset_check = hdf5_content.dup_dataset_check(dataset_name="Dataset A")
```

```
## False
```

2.3.9 zip_datasets

zip_datasets()

DESCRIPTION:

Returns a zip object of the dataset_path_list, dataset_name_list, and dataset_list.

PARAMETERS:

none

RETURNS:

- **zipped_datasets**: zip object
  iterator of a tuple of dataset_path_list, dataset_name_list, and dataset_list

EXAMPLE:

```
zipped_datasets = hdf5_content.zip_datasets()
```

```
## <zip object at 0x0000000061FC94C0>
```

```
## [('Group 1/Dataset A',
##   'Dataset A',
##   <HDF5 dataset "Dataset A": shape (1, 5), type "<i4">),
##  ('Group 1/Subgroup 1/Dataset B',
##   'Dataset B',
##   <HDF5 dataset "Dataset B": shape (1, 5), type "<i4">),
##  ('Group 2/Dataset C',
##   'Dataset C',
##   <HDF5 dataset "Dataset C": shape (1, 5), type "<i4">),
##  ('Group 2/Subgroup 2/Dataset D',
##   'Dataset D',
##   <HDF5 dataset "Dataset D": shape (1, 5), type "<i4">),
##  ('Group 3/Dataset E',
##   'Dataset E',
##   <HDF5 dataset "Dataset E": shape (1, 5), type "<i4">),
##  ('Group 3/Subgroup 3/Dataset F',
##   'Dataset F',
##   <HDF5 dataset "Dataset F": shape (1, 5), type "<i4">)]
```

3 HDF5Components

DESCRIPTION:

This class contains functions that return various components of the HDF5 file to the user including groups, datasets, Pandas/NumPy matrices of dataset values, metadata, and structured dictionaries.

PARAMETERS:

- **hdf5_filepath**: HDF5 file path
  path to the user-selected HDF5 file

ATTRIBUTES:

- **hdf5file**: HDF5 file
  user-selected HDF5 file
- **all_group_names_dict**: dict
  dictionary of every group and its associated info. Group names are keys.
- **all_dataset_names_dict**: dict
  dictionary of all datasets and their associated info. Datasetset names are keys.
- **all_dataset_paths_dict**: dict
  dictionary of all datasets and their associated info. Datasetset paths are keys.

EXAMPLE:

```
# Path to the HDF5 file in your system directory
hdf5_filepath = r"resources\example.h5"

# Instantiate the HDF5Components class
hdf5_comps = hdf5_tools.HDF5Components(hdf5_filepath=hdf5_filepath)
```

3.1 Group Functions

3.1.1 get_group_info

get_group_info(group_name)

DESCRIPTION:

Returns a dictionary of the group info.

PARAMETERS:

- **group_name**: str
  name of the group to get the info of

RETURNS:

- **group_info_dict**: dict
  dictionary of group info

EXAMPLE:

```
group_info = hdf5_comps.get_group_info(group_name="Subgroup 1")
```

```
## {'group_name': 'Subgroup 1',
##  'group_path': 'Group 1/Subgroup 1',
##  'group_metadata': {'subgroup1_attr1': array([234]),
##                     'subgroup1_attr2': array([b'test string 2'], dtype='|S14')},
##  'subgroups': {},
##  'datasets': {'Dataset B': <HDF5 dataset "Dataset B": shape (1, 5), type "<i4">}}
```

3.1.2 get_group_obj

get_group_obj(group_name)

DESCRIPTION:

Returns the HDF5 group class object for the "group_name" group.

PARAMETERS:

- **group_name**: str
  name of the group to get the HDF5 class of

RETURNS:

- **group_obj**: HDF5 group class object
  instance of the HDF5 group class

EXAMPLE:

```
group_obj = hdf5_comps.get_group_obj(group_name="Subgroup 1")
```

```
## <HDF5 group "/Group 1/Subgroup 1" (1 members)>
```

3.1.3 get_subgroup_dict

get_subgroup_dict(group_path)

DESCRIPTION:

Returns a dictionary of subgroup info for each subgroup in the group at the group_path location in the HDF5 file.

PARAMETERS:

- **group_path**: str
  path of the subgroup in the HDF5 file

RETURNS:

- **subgroup_dict**: dict
  dictionary of subgroup info for each subgroup in the group

EXAMPLE:

```
subgroup_dict = hdf5_comps.get_subgroup_dict(group_path="Group 1")
```

```
## {'Subgroup 1': {'group_name': 'Subgroup 1',
##                 'group_path': 'Group 1/Subgroup 1',
##                 'group_metadata': {'subgroup1_attr1': array([234]),
##                                    'subgroup1_attr2': array([b'test string 2'], dtype='|S14')},
##                 'subgroups': {},
##                 'datasets': {'Dataset B': <HDF5 dataset "Dataset B": shape (1, 5), type "<i4">}}}
```

3.1.4 get_group_dict

get_group_dict(group_name)

DESCRIPTION:

Returns a dictionary of the group info and its subgroups, and datasets.

PARAMETERS:

- **group_name**: str
  name of the group to get

RETURNS:

- **group_dict**: dict
  dictionary of group info, subgroups, and datasets

EXAMPLE:

```
group_dict = hdf5_comps.get_group_dict(group_name="Subgroup 1")
```

```
## {'group_name': 'Subgroup 1',
##  'group_path': 'Group 1/Subgroup 1',
##  'group_metadata': {'subgroup1_attr1': array([234]),
##                     'subgroup1_attr2': array([b'test string 2'], dtype='|S14')},
##  'subgroups': {},
##  'datasets': {'Dataset B': <HDF5 dataset "Dataset B": shape (1, 5), type "<i4">},
##  'subgroup_dict': {},
##  'dataset_dict': {'Dataset B': {'dataset_name': None,
##                                 'dataset_path': 'Group 1/Subgroup 1/Dataset B',
##                                 'dataset_metadata': {'datasetB_attr1': array([456]),
##                                                      'datasetB_attr2': array([b'test string 4'], dty
##                                 'column_names': [],
##                                 'dataset': <HDF5 dataset "Dataset B": shape (1, 5), type "<i4">}}}
```

3.1.5 get_parent_group_obj

get_parent_group_obj(group_name=None, dataset_name=None, dataset_path=None)

DESCRIPTION:

Returns the parent group class object of the dataset or group.

PARAMETERS:

- **group_name**: str
  name of the group to get the parent group of
  -or-
- **dataset_name**: str
  name of the dataset to get the parent group of
  -or-
- **dataset_path**: str
  path of the dataset to get the parent group of

RETURNS:

- **parent_group_obj**: HDF5 group class object
  instance of the class object of the parent group

EXAMPLE:

```
parent_group_obj = hdf5_comps.get_parent_group_obj(group_name="Subgroup 1")
```

```
## <HDF5 group "/Group 1" (2 members)>
```

3.1.6 get_parent_group_path

get_parent_group_path(group_name=None, dataset_name=None, dataset_path=None)

DESCRIPTION:

Returns the path of the parent group of the dataset or group.

PARAMETERS:

- **group_name**: str
  name of the group to get the parent group of
  -or-
- **dataset_name**: str
  name of the dataset to get the parent group of
  -or-
- **dataset_path**: str
  path of the dataset to get the parent group of

RETURNS:

- **parent_group_path**: str
  path in HDF5 file to the parent group

EXAMPLE:

```
parent_group_path = hdf5_comps.get_parent_group_path(group_name="Subgroup 1")
```

```
## /Group 1
```

3.1.7 get_parent_group_name

get_parent_group_name(group_name=None, dataset_name=None, dataset_path=None)

DESCRIPTION:

Returns the path of the parent group of the dataset or group.

PARAMETERS:

- **group_name**: str
  name of the group to get the parent group of
  -or-
- **dataset_name**: str
  name of the dataset to get the parent group of
  -or-
- **dataset_path**: str
  path of the dataset to get the parent group of

RETURNS:

- **parent_group_name**: str
  name of the parent group

EXAMPLE:

```
parent_group_name = hdf5_comps.get_parent_group_name(group_name="Subgroup 1")
```

```
## Group 1
```

3.1.8 get_eeg_matrix

get_eeg_matrix(group_name, matrix_type="pandas")

DESCRIPTION:

Combines all CNS EEG datasets (EEG channels) within a group into one 2D matrix. Returns a matrix of all EEG channel values for the following CNS groups: Impedance, NeonatalParamas, SampleSeries.

PARAMETERS:

- **group_name**: str
  name of the CNS group containing the datasets to be converted
- **matrix_type**: str
  matrix type to convert the dataset values into - "pandas" or "numpy" default value: "pandas"

RETURNS:

- **eeg_matrix**: Pandas DataFrame or NumPy Array
  matrix of all EEG channel values of a group

EXAMPLE:

```
# NOTE: this is currently an internal Moberg Analytics function
#eeg_matrix = hdf5_comps.get_eeg_matrix(group_name="SampleSeries", matrix_type="pandas")
```

3.2 Dataset Functions

3.2.1 get_dataset_info

get_dataset_info(dataset_name=None, dataset_path=None)

DESCRIPTION:

Returns a dictionary of the dataset info.

PARAMETERS:

- **dataset_name**: str
  name of the dataset to get the info of
  -or-

- **dataset_path**: str
  path of the dataset to get the parent group of

RETURNS:

- **dataset_info_dict**: dict
  dictionary of dataset info

EXAMPLE:

```
dataset_info_dict = hdf5_comps.get_dataset_info(dataset_name="Dataset A")
```

```
## {'dataset_name': 'Dataset A',
##  'dataset_path': 'Group 1/Dataset A',
##  'dataset_metadata': {'datasetA_attr1': array([345]),
##                       'datasetA_attr2': array([b'test string 3'], dtype='|S14')},
##  'column_names': []}
```

3.2.2 get_dataset_obj

get_dataset_obj(dataset_name=None, dataset_path=None)

DESCRIPTION:

Returns the HDF5 dataset class object for the "dataset_name" dataset.

PARAMETERS:

- **dataset_name**: str
  name of the dataset to get HDF5 class object of
  -or-
- **dataset_path**: str
  path of the dataset to get HDF5 class object of

RETURNS:

- **dataset_obj**: HDF5 dataset class object
  instance of the HDF5 dataset class

EXAMPLE:

```
dataset_obj = hdf5_comps.get_dataset_obj(dataset_name="Dataset A")
```

```
## <HDF5 dataset "Dataset A": shape (1, 5), type "<i4">
```

3.2.3 get_dataset_dict

get_dataset_dict(dataset_name=None, dataset_path=None)

DESCRIPTION:

Returns a dictionary of the dataset info and values.

PARAMETERS:

- **dataset_name**: str
  name of the dataset to get
  -or-
- **dataset_path**: str
  path of the dataset to get

RETURNS:

- **dataset_dict**: HDF5 dataset class object
  dictionary of dataset info and values

EXAMPLE:

```
dataset_dict = hdf5_comps.get_dataset_dict(dataset_name="Dataset A")
```

```
## {'dataset_name': None,
##  'dataset_path': 'Group 1/Dataset A',
##  'dataset_metadata': {'datasetA_attr1': array([345]),
##                       'datasetA_attr2': array([b'test string 3'], dtype='|S14')},
##  'column_names': [],
##  'dataset': <HDF5 dataset "Dataset A": shape (1, 5), type "<i4">}
```

3.2.4 get_column_names

get_column_names(dataset_name=None, dataset_path=None)

DESCRIPTION:

Returns a list of dataset column names.

PARAMETERS:

- **dataset_name**: str
  name of the dataset to get the column names from
  -or-
- **dataset_path**: str
  path of the dataset to get the column names from

RETURNS:

- **column_names_list**: list
  list of column names

EXAMPLE:

```
column_names_list = hdf5_comps.get_column_names(dataset_name="Dataset A")
```

```
## []
```

3.2.5 get_values

get_values(dataset_name=None, dataset_path=None, matrix_type="pandas")

DESCRIPTION:

Returns a matrix of the values in the dataset.

PARAMETERS:

- **dataset_name**: str
  name of the dataset to get the values of
  -or-
- **dataset_path**: str
  path of the dataset to get the values of
- **matrix_type**: str
  matrix type to get the dataset values in - "pandas" or "numpy"

RETURNS:

- **dataset_values**: Pandas DataFrame or NumPy Array
  matrix of dataset values

EXAMPLE:

```
dataset_values = hdf5_comps.get_values(dataset_name="Dataset A", matrix_type="numpy")
```

```
## [[0 1 2 3 4]]
```

4 HDF5Helper

DESCRIPTION:

This class contains functions for argument, group, dataset, and duplicate checks as well as other methods that add functionality to HDF5Content and HDF5Components.

PARAMETERS:

none

ATTRIBUTES:

none

EXAMPLE:

```
# Path to the HDF5 file in your system directory
hdf5_filepath = r"resources\example.h5"

# Instantiate the HDF5Helper class
hdf5_helper = hdf5_tools.HDF5Helper()
```

4.1 Group Functions

4.1.1 check_group_name

check_group_name(group_names_list, group_name)

DESCRIPTION:

Returns True if "group_name" is a group name in the HDF5 file. Otherwise returns False and raises a GroupNameError.

PARAMETERS:

- **group_names_list**: list
  list of group names in the HDF5 file
- **group_name**: str
  name of the group to check

RETURNS:

- **True**
  -or-
- **False**

EXAMPLE:

```
group_names_list = hdf5_content.get_all_group_names()
valid_group_name = hdf5_helper.check_group_name(group_names_list, group_name="Group 1")
```

```
## True
```

4.1.2 is_group

is_group(hdf5_obj)

DESCRIPTION:

Returns True if the "hdf5_obj" object a HDF5 group object. Otherwise returns False.

PARAMETERS:

- **test_obj**: HDF5 class object
  HDF5 class object to check

RETURNS:

- **True**

  -or-

- **False**

EXAMPLE:

```
group_obj = hdf5_content.get_all_group_objs()[0]
group_obj_bool = hdf5_helper.is_group(hdf5_obj=group_obj)
```

```
## True
```

4.2 Dataset Functions

4.2.1 check_dataset_name

check_dataset_name(dataset_names_list, dataset_name)

DESCRIPTION:

Returns True if "dataset_name" is a dataset name in the HDF5 file. Otherwise returns False and raises a DatasetNameError.

PARAMETERS:

- **dataset_names_list**: list

  list of dataset names in the HDF5 file

- **dataset_name**: str

  name of the dataset to check

RETURNS:

- **True**

  -or-

- **False**

EXAMPLE:

```
dataset_names_list = hdf5_content.get_all_dataset_names()
valid_dataset_name = hdf5_helper.check_dataset_name(dataset_names_list, dataset_name="Dataset A")
```

```
## True
```

4.2.2 is_dataset

is_dataset(hdf5_obj)

DESCRIPTION:

Returns True if the "hdf5_obj" object a HDF5 group object. Otherwise returns False.

PARAMETERS:

- **test_obj**: HDF5 class object
  HDF5 class object to check

RETURNS:

- **True**
  -or-
- **False**

EXAMPLE:

```
dataset_obj = hdf5_content.get_all_dataset_objs()[0]
dataset_obj_bool = hdf5_helper.is_dataset(hdf5_obj=dataset_obj)
```

```
## True
```