

Mapping HDF5 object/regional reference to OPeNDAP

MuQun Yang, Hyo-kyung Lee

The HDF Group

James Gallagher

OPeNDAP

I. Object reference and regional reference

Object reference and regional reference are HDF5 datatypes that can be used to access HDF5 objects. Primarily object reference can be used to access HDF5 group and dataset. Regional reference will be used to access regions of HDF5 dataset.

II. DAP URL

URL is a DAP datatype. It is a string containing an OPeNDAP Uniform Resource Locator (URL). A user can specify an OPeNDAP URL to indicate some data file on a remote host machine. The data can be accessed through the network via the OPeNDAP handler. A typical OPeNDAP URL[1] is listed as:

```
>dncview http://dods.gso.uri.edu/cgi-bin/nph-
nc/data/fnoc1.nc.das
      ^      ^      ^      ^      ^      ^      ^
      |      |      |      |      |      |      |
Program |      |      |      |      |      |      |
Protocol--      |      |      |      |      |      |
Machine Name-----      |      |      |      |      |
Server-----      |      |      |      |      |
Directory-----      |      |      |      |      |
Filename-----      |      |      |      |      |
URL Suffix-----      |      |      |      |      |
```

Parts of an OPeNDAP URL (without a constraint expression)

However, URL type is not restricted to only OPeNDAP URL. It can essentially serve as a reference for any DAP objects. This is the feature we use to map HDF5 object reference/regional reference to OPeNDAP.

III. Mapping HDF5 object reference and regional reference to DAP URL

1. Using URL to encode HDF5 objects

The DAP URL type can reference a DAP object. In the context of the HDF5 handler, HDF5 handler will map object reference and regional reference to DAP URL. That is, HDF5 object reference and regional reference will be represented in the DDS/DDX as a URL. When a DAP client sends out a request to decode URL type via HDF5 handler, the handler will treat the URL as either an object reference or regional reference. One should be aware that one should not use the dereferencing operator of OPeNDAP to dereference the HDF5 object reference and regional reference URL points to. We will illustrate the process of using URL to encode HDF5 object reference and regional reference in the next section.

2. Examples to illustrate the usage of URL to encode HDF5 objects

2.1 An example to encode object reference

Supposed that there is an HDF5 object reference array `/foo/od1` in an HDF5 file called `foo.h5`. This object reference array includes three elements and it refers to three HDF5 datasets: `foo1[100]`, `foo2[10][10]` and `foo3[1]`.

When accessing DAP DDS,

`http://hdfdap.hdfgroup.uiuc.edu:8080/cgi-bin/nph-dods/foo.h5.dds`

The object reference will be displayed as:

URL `/foo/od1[3]`

To obtain the data the object reference refers to will consist of two steps:

a) Obtain the absolute HDF5 path(s) the object references refer to

When an OPeNDAP client sends a request for an object reference type:

`http://hdfdap.hdfgroup.uiuc.edu:8080/cgi-bin/nph-dods/foo.h5.dods?/foo/od1`

The handler will return the absolute path of these three HDF5 datasets:

`{/foo/foo1,/foo/foo2,/foo/foo3}`

The client can also request one element of the object reference type:

`http://hdfdap.hdfgroup.uiuc.edu:8080/cgi-bin/nph-dods/foo.h5.dods?/foo/od1[1]`

The handler will then just return the absolute path of the dataset of that object reference refers to. In this case `{/foo/foo2}`.

b) Access the HDF5 datasets with the regular syntax

`http://hdfdap.hdfgroup.uiuc.edu:8080/cgi-bin/nph-dods/foo.h5.dods?/foo/foo1`

The real data will then be fetched.

2.2 An example to encode regional reference

Using the same example file as section 2.1; inside `foo.h5` we have a regional reference array `/foo/rd1`. This array refers to subsets of two HDF5 dataset array. In this case, `/foo/foo1[0:20]` and `/foo/foo2/[0:2][0:5]`.

When accessing DAP DDS,

`http://hdfdap.hdfgroup.uiuc.edu:8080/cgi-bin/nph-dods/foo.h5.dds`

The regional reference will be displayed as:

URL `/foo/rd1[2]`

When an OPeNDAP client send a request for an object reference type:

`http://hdfdap.hdfgroup.uiuc.edu:8080/cgi-bin/nph-dods/foo.h5.dods?/foo/rd1`

The handler will return the absolute path of these two HDF5 datasets:

`{/foo/foo1[0:20],/foo/foo2[0:2][0:5]}`

Then the client can have access to the “region” of the dataset with the regular expression constraint:

`http://hdfdap.hdfgroup.uiuc.edu:8080/cgi-bin/nph-dods/foo.h5.dods?/foo/foo1[0:20]`