
Pydap Documentation

Release 3.2

Roberto De Almeida

Mar 30, 2017

Contents

| | | |
|----------|-----------------------------------|-----------|
| 1 | Quickstart | 3 |
| 2 | Help | 5 |
| 3 | Documentation | 7 |
| 3.1 | Using the client | 7 |
| 3.2 | Running a server | 16 |
| 3.3 | Handlers | 18 |
| 3.4 | Responses | 22 |
| 3.5 | Developer documentation | 24 |
| 3.6 | License | 34 |
| 4 | Indices and tables | 35 |

Pydap is a pure [Python](#) library implementing the [Data Access Protocol](#), also known as **DODS** or **OPeNDAP**. You can use Pydap as a client to access hundreds of scientific datasets in a transparent and efficient way through the internet; or as a server to easily distribute your data from a variety of formats.

CHAPTER 1

Quickstart

You can install the latest version (3.2) using `pip`. After installing `pip` you can install Pydap with this command:

```
$ pip install Pydap
```

This will install Pydap together with all the required dependencies. You can now open any remotely served dataset, and Pydap will download the accessed data on-the-fly as needed:

```
>>> from pydap.client import open_url
>>> dataset = open_url('http://test.opendap.org/dap/data/nc/coads_climatology.nc')
>>> var = dataset['SST']
>>> var.shape
(12, 90, 180)
>>> var.type
<class 'pydap.model.Float32'>
>>> print var[0,10:14,10:14] # this will download data from the server
<class 'pydap.model.GridType'>
  with data
[[ -1.26285708e+00  -9.99999979e+33  -9.99999979e+33  -9.99999979e+33]
 [ -7.69166648e-01  -7.79999971e-01  -6.75454497e-01  -5.95714271e-01]
 [  1.28333330e-01  -5.00000156e-02  -6.36363626e-02  -1.41666666e-01]
 [  6.38000011e-01   8.95384610e-01   7.21666634e-01   8.10000002e-01]]
  and axes
366.0
[-69. -67. -65. -63.]
[ 41.  43.  45.  47.]
```

For more information, please check the documentation on using Pydap as a client. Pydap also comes with a simple server, implemented as a [WSGI](#) application. To use it, you first need to install a data handler:

```
$ pip install pydap.handlers.netcdf
```

This will install support for [netCDF](#) files; more handlers for different formats are available, if necessary. Now create a directory for your server data:

```
$ paster create -t pydap myserver
```

To run the server just issue the command:

```
$ paster serve ./myserver/server.ini
```

This will start a standalone server running on <http://localhost:8001/>, serving netCDF files from `./myserver/data/`, similar to the test server at <http://test.pydap.org/>. Since the server uses the WSGI standard, it can easily be run behind Apache. The server documentation has more information on how to better deploy Pydap.

CHAPTER 2

Help

If you need any help with Pydap, please feel free to send an email to the [mailing list](#).

Using the client

Pydap can be used as a client to inspect and retrieve data from any of the [hundreds of scientific datasets](#) available on the internet on [OPeNDAP](#) servers. This way, it's possible to introspect and manipulate a dataset as if it were stored locally, with data being downloaded on-the-fly as necessary.

Accessing gridded data

Let's start accessing gridded data, i.e., data that is stored as a regular multidimensional array. Here's a simple example where we access the [COADS](#) climatology from the official OPeNDAP server:

```
>>> from pydap.client import open_url
>>> dataset = open_url('http://test.opendap.org/dap/data/nc/coads_climatology.nc')
>>> print type(dataset)
<class 'pydap.model.DatasetType'>
```

Here we use the `pydap.client.open_url` function to open an URL specifying the location of the dataset; this URL should be stripped of the extensions commonly used for OPeNDAP datasets, like `.dds` or `.das`. When we access the remote dataset the function returns a `DatasetType` object, which is a *Structure* – a fancy dictionary that stores other variables. We can check the names of the store variables like we would do with a Python dictionary:

```
>>> print dataset.keys()
['SST', 'AIRT', 'UWND', 'VWND']
```

Let's work with the SST variable; we can reference it using the usual dictionary syntax of `dataset['SST']`, or using the “lazy” syntax `dataset.SST`:

```
>>> sst = dataset['SST'] # or dataset.SST
>>> print type(sst)
<class 'pydap.model.GridType'>
```

Note that the variable is of type `GridType`, a multidimensional array with specific axes defining each of its dimensions:

```
>>> print sst.dimensions
('TIME', 'COADSY', 'COADSX')
>>> print sst.maps
{'TIME': <pydap.model.BaseType object at ...>, 'COADSY': <pydap.model.BaseType object
↪at ...>, 'COADSX': <pydap.model.BaseType object at ...>}
```

Each map is also, in turn, a variable that can be accessed using the same syntax used for Structures:

```
>>> print sst.TIME
<class 'pydap.model.BaseType'>
  with data
<ArrayProxy pointing to variable "SST.TIME" at "http://test.opendap.org/dap/data/nc/
↪coads_climatology.nc">
```

The axes are all of type `BaseType`. This is the OPeNDAP equivalent of a multidimensional array, with a specific shape and type. Even though no data have been downloaded up to this point, we can introspect these attributes from the axes or from the Grid itself:

```
>>> print sst.shape
(12, 90, 180)
>>> print sst.dtype
float32
>>> print sst.TIME.shape
(12,)
>>> print sst.TIME.dtype
float64
```

We can also introspect the variable attributes; they are stored in an attribute appropriately called `attributes`, and they can also be accessed with a “lazy” syntax:

```
>>> import pprint
>>> pprint.pprint(sst.attributes)
{'_FillValue': -9.999999790214768e+33,
 'history': 'From coads_climatology',
 'long_name': 'SEA SURFACE TEMPERATURE',
 'missing_value': -9.999999790214768e+33,
 'units': 'Deg C'}
>>> print sst.units
Deg C
```

Finally, we can also download some data. To download data we simply access it like we would access a `Numpy` array, and the data for the corresponding subset will be downloaded on the fly from the server:

```
>>> print sst.shape
(12, 90, 180)
>>> grid = sst[0,10:14,10:14] # this will download data from the server
>>> print grid
<class 'pydap.model.GridType'>
  with data
[[ -1.26285708e+00  -9.99999979e+33  -9.99999979e+33  -9.99999979e+33]
 [ -7.69166648e-01  -7.79999971e-01  -6.75454497e-01  -5.95714271e-01]
 [  1.28333330e-01  -5.00000156e-02  -6.36363626e-02  -1.41666666e-01]
 [  6.38000011e-01   8.95384610e-01   7.21666634e-01   8.10000002e-01]]
  and axes
366.0
```

```
[-69. -67. -65. -63.]
[ 41.  43.  45.  47.]
```

The data itself can be accessed in the `array` attribute of the `Grid`, and also on the individual axes:

```
>>> print grid.array[:]
[[ -1.26285708e+00  -9.99999979e+33  -9.99999979e+33  -9.99999979e+33]
 [ -7.69166648e-01  -7.79999971e-01  -6.75454497e-01  -5.95714271e-01]
 [  1.28333330e-01  -5.00000156e-02  -6.36363626e-02  -1.41666666e-01]
 [  6.38000011e-01   8.95384610e-01   7.21666634e-01   8.10000002e-01]]
>>> print grid.COADSX[:]
[ 41.  43.  45.  47.]
```

Alternatively, we could have downloaded the data directly, skipping the axes:

```
>>> print sst.array[0,10:14,10:14]
[[ -1.26285708e+00  -9.99999979e+33  -9.99999979e+33  -9.99999979e+33]
 [ -7.69166648e-01  -7.79999971e-01  -6.75454497e-01  -5.95714271e-01]
 [  1.28333330e-01  -5.00000156e-02  -6.36363626e-02  -1.41666666e-01]
 [  6.38000011e-01   8.95384610e-01   7.21666634e-01   8.10000002e-01]]
```

Instead of indexes we can also subset the data using its maps, in a more natural way. Just keep in mind that sometimes axes can be cyclic, like longitude, and you may have to download two separate parts and concatenate them together. This is not the case here:

```
>>> print sst[ 0 , (-10 < sst.COADSY) & (sst.COADSY < 10) , (sst.COADSX > 320) & (sst.
↳COADSX < 328) ]
<class 'pydap.model.GridType'>
  with data
[[ -9.99999979e+33  -9.99999979e+33   2.75645447e+01   2.74858131e+01]
 [ -9.99999979e+33  -9.99999979e+33   2.75924988e+01   2.74538631e+01]
 [  2.74333324e+01   2.75676193e+01   2.75849991e+01   2.72220459e+01]
 [  2.74995346e+01   2.75236359e+01   2.75734081e+01   2.71845455e+01]
 [  2.75163631e+01   2.74263630e+01   2.73368282e+01   2.72538090e+01]
 [  2.74848824e+01   2.74654541e+01   2.72157135e+01   2.71914806e+01]
 [  2.75176182e+01   2.74858055e+01   2.71117859e+01   2.71154156e+01]
 [  2.74184361e+01   2.71918182e+01   2.70971432e+01   2.68821430e+01]
 [  2.66373062e+01   2.65258331e+01   2.66468735e+01   2.65185719e+01]
 [  2.56100006e+01   2.62577419e+01   2.62805882e+01   2.62171783e+01]]
  and axes
366.0
[-9. -7. -5. -3. -1.  1.  3.  5.  7.  9.]
[ 321.  323.  325.  327.]
```

Older Servers

Some servers using a very old OPeNDAP application might run out of memory when attempting to retrieve both the data and the coordinate axes of a variable. The work around is to simply disable the retrieval of coordinate axes by using the `output_grid` option to open url:

```
>>> from pydap.client import open_url
>>> dataset = open_url('http://test.opendap.org/dap/data/nc/coads_climatology.nc',
↳output_grid=False)
```

Accessing sequential data

Now let's see an example of accessing sequential data. Sequential data consists of one or more records of related variables, such as a simultaneous measurements of temperature and wind velocity, for example. In this example we're going to access data from the [Argo project](#), consisting of profiles made by autonomous buoys drifting on the ocean:

```
>>> from pydap.client import open_url
>>> dataset = open_url('http://dapper.pmel.noaa.gov/dapper/argo/argo_all.cdp')
```

This dataset is fairly complex, with several variables representing heterogeneous 4D data. The layout of the dataset follows the [Dapper in-situ conventions](#), consisting of two nested sequences: the outer sequence contains, in this case, a latitude, longitude and time variable, while the inner sequence contains measurements along a z axis.

The first thing we'd like to do is limit our region; let's work with a small region in the Tropical Atlantic:

```
>>> print type(dataset.location)
<class 'pydap.model.SequenceType'>
>>> print dataset.location.keys()
['LATITUDE', 'JULD', 'LONGITUDE', '_id', 'profile', 'attributes', 'variable_attributes
↵']
>>> my_location = dataset.location[
...     (dataset.location.LATITUDE > -2) &
...     (dataset.location.LATITUDE < 2) &
...     (dataset.location.LONGITUDE > 320) &
...     (dataset.location.LONGITUDE < 330)]
```

Note that the variable `dataset.location` is of type `SequenceType` – also a `Structure` that holds other variables. Here we're limiting the sequence `dataset.location` to measurements between given latitude and longitude boundaries. Let's access the identification number of the first 10-or-so profiles:

```
>>> for i, id_ in enumerate(my_location['_id']):
...     print id_
...     if i == 10:
...         print '...'
...         break
1125393
835304
839894
875344
110975
864748
832685
887712
962673
881368
1127922
...
>>> print len(my_location['_id'])
623
```

Note that calculating the length of a sequence takes some time, since the client has to download all the data and do the calculation locally. This is why you should use `len(my_location['_id'])` instead of `len(my_location)`. Both should give the same result (unless the dataset changes between requests), but the former retrieves only data for the `_id` variable, while the later retrieves data for all variables.

We can explicitly select just the first 5 profiles from our sequence:

```
>>> my_location = my_location[:5]
>>> print len(my_location['_id'])
5
```

And we can print the temperature profiles at each location. We're going to use the `coards` module to convert the time to a Python `datetime` object:

```
>>> from coards import from_udunits
>>> for position in my_location:
...     date = from_udunits(position.JULD.data, position.JULD.units.replace('GMT',
↪ '+0:00'))
...     print
...     print position.LATITUDE.data, position.LONGITUDE.data, date
...     print '=' * 40
...     i = 0
...     for pressure, temperature in zip(position.profile.PRES, position.profile.
↪ TEMP):
...         print pressure, temperature
...         if i == 10:
...             print '...'
...             break
...         i += 1
```

```
-1.01 320.019 2009-05-03 11:42:34+00:00
```

```
=====
```

```
5.0 28.59
10.0 28.788
15.0 28.867
20.0 28.916
25.0 28.94
30.0 28.846
35.0 28.566
40.0 28.345
45.0 28.05
50.0 27.595
55.0 27.061
...
```

```
-0.675 320.027 2006-12-25 13:24:11+00:00
```

```
=====
```

```
5.0 27.675
10.0 27.638
15.0 27.63
20.0 27.616
25.0 27.617
30.0 27.615
35.0 27.612
40.0 27.612
45.0 27.605
50.0 27.577
55.0 27.536
...
```

```
-0.303 320.078 2007-01-12 11:30:31.001000+00:00
```

```
=====
```

```
5.0 27.727
10.0 27.722
15.0 27.734
```

```

20.0 27.739
25.0 27.736
30.0 27.718
35.0 27.694
40.0 27.697
45.0 27.698
50.0 27.699
55.0 27.703
...

-1.229 320.095 2007-04-22 13:03:35.002000+00:00
=====
5.0 28.634
10.0 28.71
15.0 28.746
20.0 28.758
25.0 28.755
30.0 28.747
35.0 28.741
40.0 28.737
45.0 28.739
50.0 28.748
55.0 28.806
...

-1.82 320.131 2003-04-09 13:20:03+00:00
=====
5.1 28.618
9.1 28.621
19.4 28.637
29.7 28.662
39.6 28.641
49.6 28.615
59.7 27.6
69.5 26.956
79.5 26.133
89.7 23.937
99.2 22.029
...

```

These profiles could be easily plotted using `matplotlib`:

```

>>> for position in my_location:
...     plot(position.profile.TEMP, position.profile.PRES)
>>> show()

```

You can also access the deep variables directly. When you iterate over these variables the client will download the data as nested lists:

```

>>> for value in my_location.profile.PRES:
...     print value[:10]
[5.0, 10.0, 15.0, 20.0, 25.0, 30.0, 35.0, 40.0, 45.0, 50.0]
[5.0, 10.0, 15.0, 20.0, 25.0, 30.0, 35.0, 40.0, 45.0, 50.0]
[5.0, 10.0, 15.0, 20.0, 25.0, 30.0, 35.0, 40.0, 45.0, 50.0]
[5.0, 10.0, 15.0, 20.0, 25.0, 30.0, 35.0, 40.0, 45.0, 50.0]
[5.0999999, 9.1000004, 19.4, 29.700001, 39.599998, 49.599998, 59.700001, 69.5, 79.5,
↵89.699997]

```


Pydap 3.0 has been rewritten to make it easier to work with Dapper datasets like this one, and it should be intuitive¹ to work with these variables.

Authentication

To use Basic and Digest authentication, simply add your username and password to the dataset URL. Keep in mind that if the server only supports Basic authentication your credentials will be sent as plaintext, and could be sniffed on the network.

```
>>> from pydap.client import open_url
>>> dataset = open_url('http://username:password@server.example.com/path/to/dataset')
```

The [Central Authentication Service \(CAS\)](#) is a single sign-on protocol for the web, usually involving a web browser and cookies. Nevertheless it's possible to use Pydap with an OPeNDAP server behind a CAS. The function `install_cas_client` below replaces Pydap's default HTTP function with a new version able to submit authentication data to an HTML form and store credentials in cookies. (In this particular case, the server uses Javascript to redirect the browser to a new location, so the client has to parse the location from the Javascript code; other CAS would require a tweaked function.)

To use it, just attach a web browsing `session` with authentication cookies:

```
>>> from pydap.client import open_url
>>> from pydap.cas.get_cookies import setup_session
>>> session = setup_session(authentication_url, username, password)
>>> dataset = open_url('http://server.example.com/path/to/dataset', session=session)
```

This method could work but each CAS is slightly different and might require a specifically designed `setup_session` instance. Two CAS are however explicitly supported by pydap:

URS NASA EARTHDATA

Authentication is done through a username and a password:

```
>>> from pydap.client import open_url
>>> from pydap.cas.urs import setup_session
>>> dataset_url = 'http://server.example.com/path/to/dataset'
>>> session = setup_session(username, password, check_url=dataset_url)
>>> dataset = open_url(dataset_url, session=session)
```

Earth System Grid Federation (ESGF)

Authentication is done through an openid and a password:

```
>>> from pydap.client import open_url
>>> from pydap.cas.esgf import setup_session
>>> session = setup_session(openid, password)
>>> dataset = open_url('http://server.example.com/path/to/dataset', session=session)
```

If your `openid` contains the string `ceda.ac.uk` authentication requires an additional `username` argument:

¹ But please check this quote.

```
>>> from pydap.client import open_url
>>> from pydap.cas.esgf import setup_session
>>> session = setup_session(openid, password, username=username)
>>> dataset = open_url('http://server.example.com/path/to/dataset', session=session)
```

Advanced features

When you open a remote dataset, the `DatasetType` object has a special attribute named `functions` that can be used to invoke any server-side functions. Here's an example of using the `geogrid` function from Hyrax:

```
>>> dataset = open_url('http://test.opendap.org/dap/data/nc/coads_climatology.nc')
>>> new_dataset = dataset.functions.geogrid(dataset.SST, 10, 20, -10, 60)
>>> print new_dataset.SST.shape
(12, 12, 21)
>>> print new_dataset.SST.COADSY[:]
[-11.  -9.  -7.  -5.  -3.  -1.   1.   3.   5.   7.   9.  11.]
>>> print new_dataset.SST.COADSX[:]
[ 21.  23.  25.  27.  29.  31.  33.  35.  37.  39.  41.  43.  45.  47.  49.
  51.  53.  55.  57.  59.  61.]
```

Unfortunately, there's currently no standard mechanism to discover which functions the server support. The `function` attribute will accept any function name the user specifies, and will try to pass the call to the remote server.

You can pass any URL to the `open_url` function, together with any valid constraint expression. Here's an example of restricting values for the months of January, April, July and October:

```
>>> dataset = open_url('http://test.opendap.org/dap/data/nc/coads_climatology.nc?
↳SST[0:3:11][0:1:89][0:1:179]')
>>> print dataset.SST.shape
(4, 90, 180)
```

This can be extremely useful for server side-processing; for example, we can create and access a new variable `A` in this dataset, equal to twice `SSH`:

```
>>> dataset = open_url('http://hycom.coaps.fsu.edu:8080/thredds/dodsC/las/dynamic/
↳data_A5CDC5CAF9D810618C39646350F727FF.jnl_expr_%7B%7D%7Blet%20A=SSH*2%7D?A')
>>> print dataset.keys()
['A']
```

In this case, we're using the Ferret syntax `let A=SSH*2` to define the new variable, since the data is stored in an [F-TDS server](#). Server-side processing is useful when you want to reduce the data before downloading it, to calculate a global average, for example.

The client module has a special function called `open_dods`, used to access raw data from a DODS response:

```
>>> from pydap.client import open_dods
>>> dataset = open_dods(
...     'http://test.opendap.org/dap/data/nc/coads_climatology.nc.dods?
↳SST[0:3:11][0:1:89][0:1:179]')
```

This function allows you to access raw data from any URL, including appending expressions to [F-TDS](#) and [GDS](#) servers or calling server-side functions directly. By default this method downloads the data directly, and skips meta-data from the DAS response; if you want to investigate and introspect datasets you should set the `get_metadata` parameter to `true`:

```
>>> dataset = open_dods(
...     'http://test.opendap.org/dap/data/nc/coads_climatology.nc.dods?
↳SST[0:3:11][0:1:89][0:1:179]',
...     get_metadata=True)
>>> print dataset.attributes['NC_GLOBAL']['history']
FERRET V4.30 (debug/no GUI) 15-Aug-96
```

You can specify a cache directory in the `pydap.lib.CACHE` global variable. If this value is different than `None`, the client will try (if the server headers don't prohibit) to cache the result, so repeated requests will be read from disk instead of the network:

```
>>> import pydap.lib
>>> pydap.lib.CACHE = "/tmp/pydap-cache/"
```

To specify a timeout for the client, just set the global variable `pydap.lib.TIMEOUT` to the desired number of seconds; after this time trying to connect the client will give up. The default is `None` (never timeout).

```
>>> import pydap.lib
>>> pydap.lib.TIMEOUT = 60
```

It's possible to configure Pydap to access the network through a proxy server. Here's an example for an HTTP proxy running on localhost listening on port 8000:

```
>>> import httplib2
>>> from pydap.util import socks
>>> import pydap.lib
>>> pydap.lib.PROXY = httplib2.ProxyInfo(
...     socks.PROXY_TYPE_HTTP, 'localhost', 8000)
```

This way, all further calls to `pydap.client.open_url` will be routed through the proxy server. You can also authenticate to the proxy:

```
>>> pydap.lib.PROXY = httplib2.ProxyInfo(
...     socks.PROXY_TYPE_HTTP, 'localhost', 8000,
...     proxy_user=USERNAME, proxy_pass=PASSWORD)
```

A user has reported that `httplib2` has problems authenticating against a NTLM proxy server. In this case, a simple solution is to change the `pydap.http.request` function to use `urllib2` instead of `httplib2`, monkeypatching the code like in the *CAS authentication example above*:

```
import urllib2
import logging

def install_urllib2_client():
    def new_request(url):
        log = logging.getLogger('pydap')
        log.INFO('Opening %s' % url)

        f = urllib2.urlopen(url.rstrip('?&'))
        headers = dict(f.info().items())
        body = f.read()
        return headers, body

    from pydap.util import http
    http.request = new_request
```

The function `install_urllib2_client` should then be called before doing any requests.

Running a server

Pydap comes with a lightweight and scalable OPeNDAP server, implemented as a [WSGI application](#). Being a WSGI application, Pydap can run on a variety of [servers](#), including Apache, IIS or even as a standalone Python process. It can also be seamlessly combined with different [middleware](#) for authentication/authorization, GZip compression, and much more.

In order to distribute your data first you need to install a proper handler, that will convert the data format to the Pydap data model.

Running standalone

If you just want to quickly test the Pydap server, you can run it as a standalone Python application using the server that comes with [Python Paste](#):

```
$ paster create -t pydap myserver
```

To run the server just issue the command:

```
$ paster serve ./myserver/server.ini
```

This will run the server on <http://localhost:8001/>, serving files from `./myserver/data/`. By default the server will listen only to local requests, ie, from the same machine. You can change this by editing the `server.ini` file; and you can also change the port number, though for ports lower than 1024 you will probably need to run the script as root.

To change the default directory listing, the help page and the HTML form, simply edit the corresponding templates in `./myserver/templates/`. The HTML form template is fairly complex, since it contains some application logic and some Javascript code, so be careful to not break anything.

Running Pydap with Apache

For a robust deployment you should run Pydap with Apache, using `mod_wsgi`. After installing `mod_wsgi`, create a sandbox in a directory *outside* your DocumentRoot, say `/var/www/pydap/`, using `virtualenv`:

```
$ mkdir /var/www/pydap
$ python virtualenv.py /var/www/pydap/env
```

If you want the sandbox to use your system installed packages (like Numpy, e.g.) you can use the `--system-site-packages` flag:

```
$ python virtualenv.py --system-site-packages /var/www/pydap/env
```

Now let's activate the sandbox and install Pydap – this way the module and its dependencies can be isolated from the system libraries:

```
$ source /var/www/pydap/env/bin/activate.sh
(env)$ pip install Pydap
```

And now we can install the basic server files:

```
(env)$ cd /var/www/pydap
(env)$ paster create -t pydap server
```

Now edit the file `/var/www/pydap/server/apache/pydap.wsgi` and insert the two following lines in the beginning of the file, forcing `mod_wsgi` to use the Python modules from the sandbox:

```
import site
site.addsitedir('/var/www/pydap/env/lib/python2.X/site-packages')
```

You'll need to insert the correct path (including Python version) to your sandbox `site-packages` directory, of course. After this, your file should look like this:

```
import site
site.addsitedir('/var/www/pydap/env/lib/python2.X/site-packages')

import os
from paste.deploy import loadapp

config = os.path.join(os.path.dirname(__file__), '../server.ini')
application = loadapp('config:%s' % config)
```

Now create an entry in your Apache configuration pointing to the `pydap.wsgi` file you just edited. To mount the server on the URL `/pydap`, for example, you should configure it like this:

```
WSGIScriptAlias /pydap /var/www/pydap/server/apache/pydap.wsgi

<Directory /var/www/pydap/server/apache>
    Order allow,deny
    Allow from all
</Directory>
```

This is the file I use for the `test.pydap.org` virtualhost:

```
<VirtualHost *:80>
    ServerAdmin rob@pydap.org
    ServerName test.pydap.org

    DocumentRoot /var/www/sites/test.pydap.org/server/data

    <Directory /var/www/sites/test.pydap.org/server/data>
        Order allow,deny
        Allow from all
    </Directory>

    WSGIScriptAlias / /var/www/sites/test.pydap.org/server/apache/pydap.wsgi

    <Directory /var/www/sites/test.pydap.org/server/apache>
        Order allow,deny
        Allow from all
    </Directory>

    ErrorLog /var/log/apache2/test.pydap.org.error.log

    # Possible values include: debug, info, notice, warn, error, crit,
    # alert, emerg.
    LogLevel warn

    CustomLog /var/log/apache2/test.pydap.org.access.log combined
    ServerSignature On
</VirtualHost>
```

You can find more information on the [mod_wsgi configuration guide](#). Just remember that Pydap is a WSGI application like any other else, so any information on WSGI applications applies to it as well.

Running Pydap with uWSGI

uWSGI is a “fast, self-healing and developer/sysadmin-friendly application container server coded in pure C” that can run Pydap. This is the recommended way to run Pydap if you don’t have to integrate it with other web applications. Simply install uWSGI, follow the instructions in the last section in order to create a virtualenv and Pydap installation:

```
$ mkdir /var/www/pydap
$ python virtualenv.py /var/www/pydap/env
$ source /var/www/pydap/env/bin/activate.sh
(env)$ pip install Pydap uWSGI
(env)$ cd /var/www/pydap
(env)$ paster create -t pydap server
```

Now create a file in `/etc/init/pydap.conf` with the content:

```
description "uWSGI server for Pydap"

start on runlevel [2345]
stop on runlevel [!2345]

respawn

exec /var/www/pydap/env/bin/uwsgi \
    --http-socket 0.0.0.0:80 \
    -H /var/www/pydap/env \
    --master --processes 4 \
    --paste config:/var/www/pydap/server/server.ini
```

In order to make it run automatically during boot on Linux you can type:

```
$ sudo initctl reload-configuration
```

Handlers

Handlers are special Python modules that convert between a given data format and the data model used by Pydap (defined in the `pydap.model` module). They are necessary in order to Pydap be able to actually serve a dataset. There are handlers for NetCDF, HDF 4 & 5, Matlab, relational databases, Grib 1 & 2, CSV, Seabird CTD files, and a few more.

Installing data handlers

NetCDF

NetCDF is a format commonly used in oceanography, meteorology and climate science to store data in a machine-independent format. You can install the NetCDF handler using `pip`:

```
$ pip install pydap.handlers.netcdf
```

This will take care of the necessary dependencies. You don't even need to have NetCDF libraries installed, since the handler will automatically install a pure Python NetCDF library called `pupynere`.

The NetCDF handler uses a buffered reader that access the data in contiguous blocks from disk, avoiding reading everything into memory at once. You can configure the size of the buffer by specifying a key in the `server.ini` file:

```
[app:main]
use = egg:pydap#server
root = %(here)s/data
templates = %(here)s/templates
x-wsgiorg.throw_errors = 0
pydap.handlers.netcdf.buf_size = 10000
```

In this example, the handler will read 10 thousand values at a time, converting the data and sending to the client before reading more blocks.

NCA

The `pydap.handlers.nca` is a simple handler for NetCDF aggregation (hence the name). The configuration is extremely simple. As an example, to aggregate model output in different files (say, `output1.nc`, `output2.nc`, etc.) along a new axis "ensemble" just create an INI file with the extension `.nca`:

```
; output.nca
[dataset]
match = /path/to/output*.nc
axis = ensemble
; below optional metadata:
history = Test for NetCDF aggregator

[ensemble]
values = 1, 2, ...
long_name = Ensemble members
```

This will assign the values 1, 2, and so on to each ensemble member. The new, aggregated dataset, will be accessed at the location of the INI file:

```
http://server.example.com/output.nca
```

Another example: suppose we have monthly data in files `data01.nc`, `data02.nc`, ..., `data12.nc`, and we want to aggregate along the time axis:

```
[dataset]
match = /path/to/data*.nc
axis = TIME # existing axis
```

The handler only works with NetCDF files for now, but in the future it should be changed to work with any other Pydap-supported data format. As all handlers, it can be installed using `pip`:

```
$ pip install pydap.handlers.nca
```

CDMS

This is a handler that uses the `cdms2.open` function from `CDAT/CdatLite` to read files in any of the self-describing formats netCDF, HDF, GrADS/GRIB (GRIB with a GrADS control file), or PCMDI DRS. It can be installed using `pip`:

```
$ pip install pydap.handlers.cdms
```

The handler will automatically install `CdatLite`, which requires the NetCDF libraries to be installed on the system.

SQL

The SQL handler reads data from a relation database, as the name suggests. It works by reading a file with the extension `.sql`, defining the connection to the database and other metadata using either YAML or INI syntax. Below is an example that reads data from a SQLite database:

```
# please read http://groups.google.com/group/pydap/browse_thread/thread/
↪c7f5c569d661f7f9 before
# setting your password on the DSN
database:
  dsn: 'sqlite://simple.db'
  table: test

dataset:
  NC_GLOBAL:
    history: Created by the Pydap SQL handler
    dataType: Station
    Conventions: GrADS

    contact: roberto@dealmeida.net
    name: test_dataset
    owner: Roberto De Almeida
    version: 1.0
    last_modified: !Query 'SELECT time FROM test ORDER BY time DESC LIMIT 1;'

sequence:
  name: simple
  items: !Query 'SELECT COUNT(id) FROM test'

_id:
  col: id
  long_name: sequence id
  missing_value: -9999

lon:
  col: lon
  axis: X
  grads_dim: x
  long_name: longitude
  units: degrees_east
  missing_value: -9999
  type: Float32
  global_range: [-180, 180]
  valid_range: !Query 'SELECT min(lon), max(lon) FROM test'

lat:
  col: lat
  axis: Y
  grads_dim: y
  long_name: latitude
  units: degrees_north
  missing_value: -9999
  type: Float32
```



```

global_range: [-90, 90]
valid_range: !Query 'SELECT min(lat), max(lat) FROM test'

time:
  col: time
  axis: T
  grads_dim: t
  long_name: time
  missing_value: -9999
  type: String

depth:
  axis: Z
  col: depth
  long_name: depth
  missing_value: -9999
  type: Float32
  units: m

temp:
  col: temp
  long_name: temperature
  missing_value: -9999
  type: Float32
  units: degc

```

The handler works with SQLite, MySQL, PostgreSQL, Oracle, MSSQL and ODBC databases. To install the handler use pip; you should also install the dependencies according to the database used:

```

$ pip install pydap.handlers.sql
$ pip install "pydap.handlers.sql[oracle]"
$ pip install "pydap.handlers.sql[postgresql]"
$ pip install "pydap.handlers.sql[mysql]"
$ pip install "pydap.handlers.sql[mssql]"

```

Proxy

This is a simple handler intended to serve remote datasets locally. For example, suppose you want to serve [this dataset](#) on your Pydap server. The URL of the dataset is:

```
http://test.opendap.org:8080/dods/dts/D1
```

So we create an INI file called, say, D1.url:

```

[dataset]
url = http://test.opendap.org:8080/dods/dts/D1
pass = dds, das, dods

```

The file specifies that requests for the DDS, DAS and DODS responses should be passed directly to the server (so that the data is downloaded directly from the remote server). Other requests, like for the HTML form or a WMS image are built by Pydap; in this case Pydap acts as an Opendap client, connecting to the remote server and downloading data to fulfill the request.

CSV

This is a handler for files with comma separated values. The first column should contain the variable names, and subsequent lines the data. Metadata is not supported. The handler is used mostly as a reference for building handlers for sequential data. You can install it with:

```
$ pip install pydap.handlers.csv
```

HDF5

A handler for HDF5 files, based on `h5py`. In order to install it:

```
$ pip install pydap.handlers.hdf5
```

SQLite

This is a handler very similar to the SQL handler. The major difference is that data and metadata are all contained in a single `.db` SQLite file. Metadata is stored as JSON in a table called `attributes`, while data goes into a table `data`.

The handler was created as a way to move sequential data from one server to another. It comes with a script called `freeze` which will take an Opendap dataset with sequential data and create a `.db` file that can be served using this handler. For example:

```
$ freeze http://opendap.ccst.inpe.br/Observations/PIRATA/pirata_stations.sql
```

This will create a file called `pirata_stations.db` that can be served using the SQLite handler.

Responses

Like handlers, responses are special Python modules that convert between the Pydap data model and an external representation. For instance, to access a given dataset an Opendap client request two different representations of the dataset: a *Dataset Attribute Structure* (DAS) response, describing the attributes of the dataset, and a *Dataset Descriptor Structure* (DDS), describing its structure (shape, type, hierarchy). These responses are returned by appending the extension `.das` and `.dds` to the dataset URL, respectively.

Other common responses include the ASCII (`.asc` or `.ascii`) response, which returns an ASCII representation of the data; and an HTML form for data request using the browser, at the `.html` extension. And perhaps the most important response is the `.dods` response, which actually carries the data in binary format, and is used when clients request data from the server. All these responses are standard and come with Pydap.

There are other extension responses available for Pydap; these are not defined in the DAP specification, but improve the user experience by allowing data to be accessed in different formats.

Installing additional responses

Web Map Service

This response enables Pydap to act like a [Web Map Service 1.1.1](#) server, returning images (maps) of the available data. These maps can be visualized in any WMS client like Openlayers or Google Earth.

You can install the WMS response using `pip`:

```
$ pip install pydap.responses.wms
```

This will take care of the necessary dependencies, which include [Matplotlib](#) and Pydap itself. Once the response is installed you can introspect the available layers at the URL:

```
http://server.example.com/dataset.wms?REQUEST=GetCapabilities
```

The response will create valid layers from any [COARDS](#) compliant datasets, ie, grids with properly defined latitude and longitude axes. If the data is not two-dimensional it will be averaged along each axis except for the last two, so the map represents a time and/or level average of the data. Keep in mind that Opendap constraint expressions apply before the map is generated, so it's possible to create a map of a specific level (or time) by constraining the variable on the URL:

```
http://server.example.com/dataset.wms?var3d[0]&REQUEST=GetCapabilities
```

You can specify the default colormap and the DPI resolution in the `server.ini` file:

```
[app:main]
use = egg:pydap#server
root = %(here)s/data
templates = %(here)s/templates
x-wsgiorg.throw_errors = 0
pydap.responses.wms.dpi = 80
pydap.responses.wms.cmap = jet
```

Google Earth

This response converts a Pydap dataset to a [KML](#) representation, allowing the data to be visualized using Google Earth as a client. Simply install it with `pip`:

```
$ pip install pydap.responses.kml
```

And open a URL by appending the `.kml` extension to the dataset, say:

```
http://server.example.com/dataset.kml
```

For now, the KML response will only return datasets that have a valid WMS representation. These datasets can be overlaid as layers on top of Google Earth, and are presented with a nice colorbar. In the future, [Dapper](#)-compliant datasets should be supported too.

NetCDF

This response allows data to be downloaded as a NetCDF file; it works better with gridded data, but sequential data will be converted into 1D variables. To install it, just type:

```
$ pip install pydap.responses.netcdf
```

And try to append the extension `.nc` to a request. The data will be converted on-the-fly to a NetCDF file.

Matlab

The Matlab response returns data in a Matlab v5 file. It is returned when the file is requested with the extension `.mat`, and can be installed by with:

```
$ pip install pydap.responses.matlab
```

Excel spreadsheet

This response returns sequential data as an Excel spreadsheet when `.xls` is appended to the URL. Install with:

```
$ pip install pydap.responses.xls
```

Developer documentation

This documentation is intended for other developers that would like to contribute with the development of Pydap, or extend it in new ways. It assumes that you have a basic knowledge of Python and HTTP, and understands how data is stored in different formats. It also assumes some familiarity with the [Data Access Protocol](#), though a lot of its inner workings will be explained in detail here.

The DAP data model

The DAP is a protocol designed for the efficient transmission of scientific data over the internet. In order to transmit data from the server to a client, both must agree on a way to represent data: *is it an array of integers?, a multi-dimensional grid?* In order to do this, the specification defines a *data model* that, in theory, should be able to represent any existing dataset.

Metadata

Pydap has a series of classes in the `pydap.model` module, representing the DAP data model. The most fundamental data type is called `BaseType`, and it represents a value or an array of values. Here an example of creating one of these objects:

```
>>> from pydap.model import *
>>> a = BaseType(
...     name='a',
...     data=1,
...     shape=(),
...     dimensions=(),
...     type=Int32,
...     attributes={'long_name': 'variable a'})
```

All Pydap types have five attributes in common. The first one is the name of the variable; in this case, our variable is called “a”:

```
>>> print a.name
a
```

Note that there’s a difference between the variable name (the local name `a`) and its attribute `name`; in this example they are equal, but we could reference our object using any other name:

```
>>> b = a # b now points to a
>>> print b.name
a
```

We can use special characters for the variable names; they will be quoted accordingly:

```
>>> c = BaseType(name='long & complicated')
>>> print c.name
long%20%26%20complicated
```

The second attribute is called `id`. In the examples we've seen so far, `id` and `name` are equal:

```
>>> print a.name, a.id
a a
>>> print c.name, c.id
long%20%26%20complicated long%20%26%20complicated
```

This is because the `id` is used to show the position of the variable in a given dataset, and in these examples the variables do not belong to any datasets. First let's store our variables in a container object called `StructureType`. A `StructureType` is a special type of ordered dictionary that holds other Pydap types:

```
>>> s = StructureType(name='s')
>>> s['a'] = a
>>> s['c'] = c
Traceback (most recent call last):
...
KeyError: 'Key "c" is different from variable name "long%20%26%20complicated"!'
```

Note that the variable name has to be used as its key on the `StructureType`. This can be easily remedied:

```
>>> s[c.name] = c
```

There is a special derivative of the `StructureType` called `DatasetType`, which represent the dataset. The difference between the two is that there should be only one `DatasetType`, but it may contain any number of `StructureType` objects, which can be deeply nested. Let's create our dataset object:

```
>>> dataset = DatasetType(name='example')
>>> dataset['s'] = s
>>> print dataset.id
example
>>> print dataset['s'].id
s
>>> print dataset['s']['a'].id
s.a
```

Note that for objects on the first level of the dataset, like `s`, the `id` is identical to the name. Deeper objects, like `a` which is stored in `s`, have their `id` calculated by joining the names of the variables with a period. One detail is that we can access variables stored in a structure using a “lazy” syntax like this:

```
>>> print dataset.s.a.id
s.a
```

The third common attribute that variables share is called `attributes`, which hold most of its metadata. This attribute is a dictionary of keys and values, and the values themselves can also be dictionaries. For our variable `a` we have:

```
>>> print a.attributes
{'long_name': 'variable a'}
```

These attributes can be accessed lazily directly from the variable:

```
>>> print a.long_name
variable a
```

But if you want to create a new attribute you'll have to insert it directly into `attributes`:

```
>>> a.history = 'Created by me'
>>> print a.attributes
{'long_name': 'variable a'}
>>> a.attributes['history'] = 'Created by me'
>>> print a.attributes
{'long_name': 'variable a', 'history': 'Created by me'}
```

It's always better to use the correct syntax instead of the lazy one when writing code. Use the lazy syntax only when introspecting a dataset on the Python interpreter, to save a few keystrokes.

The fourth attribute is called `data`, and it holds a representation of the actual data. We'll take a detailed look of this attribute in the next subsection.

Finally, all variables have also an attribute called `_nesting_level`. This attribute has value 1 if the variable is inside a `SequenceType` object, 0 if it's outside, and >1 if it's inside a nested sequence. This will become clearer later when we talk about sequential data.

Data

As we saw on the last subsection, all Pydap objects have a `data` attribute that holds a representation of the variable data. This representation will vary depending on the variable type.

BaseType

For the simple `BaseType` objects the `data` attribute is usually a Numpy array, though we can also use a Numpy scalar or Python number:

```
>>> a = BaseType(name='a', data=1)
>>> print a.data
1

>>> import numpy
>>> b = BaseType(name='b', data=numpy.arange(4), shape=(4,))
>>> print b.data
[0 1 2 3]
```

Note that the default type for variables is `Int32`:

```
>>> print a.type, b.type
<class 'pydap.model.Int32'> <class 'pydap.model.Int32'>
```

When you *slice* a `BaseType` array, the slice is simply passed onto the `data` attribute. So we may have:

```
>>> print b[-1]
3
>>> print b[:2]
[0 1]
>>> print a[0]
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
```

```
File "pydap/model.py", line 188, in __getitem__
TypeError: 'int' object is unsubscriptable
```

You can think of a `BaseType` object as a thin layer around Numpy arrays, until you realize that the `data` attribute can be *any* object implementing the array interface! This is how the DAP client works – instead of assigning an array with data directly to the attribute, we assign a special object which behaves like an array and acts as a *proxy* to a remote dataset.

Here's an example:

```
>>> from pydap.proxy import ArrayProxy
>>> pseudo_array = ArrayProxy(
...     'SST.SST',
...     'http://test.opendap.org/dap/data/nc/coads_climatology.nc',
...     (12, 90, 180))
>>> print pseudo_array[0, 10:14, 10:14] # download the corresponding data
[[ -1.26285708e+00  -9.99999979e+33  -9.99999979e+33  -9.99999979e+33]
 [ -7.69166648e-01  -7.79999971e-01  -6.75454497e-01  -5.95714271e-01]
 [  1.28333330e-01  -5.00000156e-02  -6.36363626e-02  -1.41666666e-01]
 [  6.38000011e-01   8.95384610e-01   7.21666634e-01   8.10000002e-01]]
```

In the example above, the data is only downloaded in the last line, when the pseudo array is sliced. The object will construct the appropriate DAP URL, request the data, unpack it and return a Numpy array.

StructureType

A `StructureType` holds no data; instead, its `data` attribute is a property that collects data from the children variables:

```
>>> s = StructureType(name='s')
>>> s[a.name] = a
>>> s[b.name] = b
>>> print a.data
1
>>> print b.data
[0 1 2 3]
>>> print s.data
(1, array([0, 1, 2, 3]))
```

The opposite is also true; it's possible to specify the structure data and have it propagated to the children:

```
>>> s.data = (1, 2)
>>> print s.a.data
1
>>> print s.b.data
2
```

The same is true for objects of `DatasetType`, since the dataset is simply the root structure.

SequenceType

A `SequenceType` object is a special kind of `StructureType` holding sequential data. Here's an example of a sequence holding the variables `a` and `c` that we created before:

```
>>> s = SequenceType(name='s')
>>> s[a.name] = a
>>> s[c.name] = c
```

Let's add some data to our sequence. This can be done in several ways, the easiest of which is adding the data to its children:

```
>>> s.a.data = [1,2,3]
>>> s.b.data = [10,20,30]
>>> s.data
array([[1, 10],
       [2, 20],
       [3, 30]], dtype=object)
```

Note that the data for the sequence is an aggregation of the children data, similar to Python's `zip()` builtin. This will be more complicated when encountering nested sequences, but for flat sequences they behave the same.

We can also iterate over the `SequenceType`. In this case, it will return a series of `StructureType` objects, each one containing data for the children variables. The `StructureType` will have the same children as the sequence, with each one containing data for a single record:

```
>>> for record in s:
...     print type(record), record.keys(), record.data
<class 'pydap.model.StructureType'> ['a', 'b'] (1, 10)
<class 'pydap.model.StructureType'> ['a', 'b'] (2, 20)
<class 'pydap.model.StructureType'> ['a', 'b'] (3, 30)
```

The second way of defining the data of a `SequenceType` is by setting it directly to the object:

```
>>> s.data = [(4,40), (5,50)]
>>> s['a'].data
array([4, 5], dtype=object)
```

Like in the `StructureType`, the data is propagated to its children. Note that in the two cases the data was defined using Python lists, being automatically converted to Numpy arrays. In fact, the `SequenceType` behaves pretty much like `record arrays` from Numpy, since we can reference them by column (`s['a']`) or by index:

```
>>> s[1].data
array([[5, 50]], dtype=object)
>>> s[s.a < 5].data
array([[4, 40]], dtype=object)
```

Note that these objects are also `SequenceType` themselves. The basic rules when working with sequence data are:

1. When a `SequenceType` is sliced with a string the corresponding children is returned. For example: `s['a']` will return child `a`;
2. When a `SequenceType` is iterated over it will return a series of `StructureType` objects, each one containing the data for a record;
3. When a `SequenceType` is sliced with an integer, a comparison or a `slice()` a new `SequenceType` will be returned;
4. When a `SequenceType` is sliced with a tuple of strings a new `SequenceType` will be returned, containing only the children defined in the tuple in the new order. For example, `s[('c', 'a')]` will return a sequence `s` with the children `c` and `a`, in that order.

Note that except for rule 4 `SequenceType` mimics the behavior of Numpy record arrays.

Now imagine that we want to add to a `SequenceType` data pulled from a relational database. The easy way would be to fetch the data in the correct column order, and insert it into the sequence. But what if we don't want to store the data in memory, and instead we would like to stream it directly from the database? In this case we can create an object that behaves like a record array, similar to the proxy object that implements the array interface. Pydap defines a "protocol" called `SequenceData`, which is simply any object that:

1. Returns data when iterated over.
2. Returns a new `SequenceData` when sliced such that:
 - (a) if the slice is a string the new `SequenceData` contains data only for that children;
 - (b) if the slice is a tuple of strings the object contains only those children, in that order;
 - (c) if the slice is an integer, a `slice()` or a comparison, the data is filter accordingly.

The base implementation works by wrapping data from a basic Numpy array:

```
class SequenceData(object):
    """
    An extended Numpy record array.

    The so-called ``SequenceData`` protocol extends the behavior of record
    arrays from Numpy so that tuples passed to ``_getitem`` return a new
    object with only those children.

    """
    def __init__(self, data, keys):
        self.data = data
        self.keys = keys

    def __iter__(self):
        return iter(self.data)

    def __len__(self):
        return len(self.data)

    def __getitem__(self, key):
        if isinstance(key, basestring):
            col = self.keys.index(key)
            return SequenceData(self.data[:,col], ())
        elif isinstance(key, tuple):
            return SequenceData(
                numpy.dstack([self.data[:, self.keys.index(k)] for k in key]),
                key)
        else:
            return SequenceData(self.data[key], self.keys)

    # comparison are passed to the data object
    def __eq__(self, other): return self.data == other
    def __ne__(self, other): return self.data != other
    def __ge__(self, other): return self.data >= other
    def __le__(self, other): return self.data <= other
    def __gt__(self, other): return self.data > other
    def __lt__(self, other): return self.data < other
```

And here is an example of how we would use it:

```
>>> from pydap.model import SequenceData
>>> s.data = SequenceData(numpy.array([(1,2), (10,20)]), ('a', 'b'))
```

```
>>> s2 = s.data[ s['a'] > 1 ]
>>> s2.data
array([[10, 20]])
```

There are many implementations of classes derived from `SequenceData`: `pydap.proxy.SequenceProxy` is a proxy to sequential data on Opendap servers, `pydap.handlers.csv.CSVProxy` wraps a CSV file, and `pydap.handlers.sql.SQLProxy` works as a stream to a relational database.

GridType

A `GridType` is a special kind of object that behaves like an array and a `StructureType`. The class is derived from `StructureType`; the major difference is that the first defined variable is a multidimensional array, while subsequent children are vector maps that define the axes of the array. This way, the `data` attribute on a `GridType` returns the data of all its children: the n -dimensional array followed by n maps.

Here is a simple example:

```
>>> g = GridType(name='g')
>>> data = numpy.arange(6.)
>>> data.shape = (2, 3)
>>> g['a'] = BaseType(name='a', data=data, shape=data.shape, type=Float32,
↳ dimensions=('x', 'y'))
>>> g['x'] = BaseType(name='x', data=numpy.arange(2.), shape=(2,), type=Float64)
>>> g['y'] = BaseType(name='y', data=numpy.arange(3.), shape=(3,), type=Float64)
>>> g.data
(array([[ 0.,  1.,  2.],
        [ 3.,  4.,  5.]]), array([ 0.,  1.]), array([ 0.,  1.,  2.]))
```

Grid behave like arrays in that they can be sliced. When this happens, a new `GridType` is returned with the proper data and axes:

```
>>> print g
<class 'pydap.model.GridType'>
  with data
[[ 0.  1.  2.]
 [ 3.  4.  5.]]
  and axes
[ 0.  1.]
[ 0.  1.  2.]
>>> print g[0]
<class 'pydap.model.GridType'>
  with data
[[ 0.  1.  2.]]
  and axes
[ 0.]
[ 0.  1.  2.]
```

Handlers

Now that we saw the Pydap data model we can understand handlers: handlers are simply classes that convert data into the Pydap data model. The NetCDF handler, for example, reads a NetCDF file and builds a `DatasetType` object. The SQL handler reads a file describing the variables and maps them to a given table on a relational database. Pydap uses [entry points](#) in order to find handlers that are installed in the system. This means that handlers can be developed

and installed separately from Pydap. Handlers are mapped to files by a regular expression that usually matches the extension of the file.

Here is the minimal configuration for a handler that serves `.npz` files from Numpy:

```
import os
import re

import numpy

from pydap.model import *
from pydap.handlers.lib import BaseHandler
from pydap.handlers.helper import constrain

class Handler(BaseHandler):

    extensions = re.compile(r'^.*\.npz$', re.IGNORECASE)

    def __init__(self, filepath):
        self.filename = os.path.split(filepath)[1]
        self.f = numpy.load(filepath)

    def parse_constraints(self, environ):
        dataset = DatasetType(name=self.filename)

        for name in self.f.files:
            data = self.f[name][:]
            dataset[name] = BaseType(name=name, data=data, shape=data.shape,
→type=data.dtype.char)

        return constrain(dataset, environ.get('QUERY_STRING', ''))

if __name__ == '__main__':
    import sys
    from paste.httpserver import serve

    application = Handler(sys.argv[1])
    serve(application, port=8001)
```

So let's go over the code. Our handler has a single class called `Handler` that should be configured as an entry point in the `setup.py` file for the handler:

```
[pydap.handler]
npz = path.to.my.handler:Handler
```

Here the name of our handler (“npz”) can be anything, as long as it points to the correct class. In order for Pydap to be able to find the handler, it must be installed in the system with either a `python setup.py install` or, even better, `python setup.py develop`.

The class-level attribute `extensions` defines a regular expression that matches the files supported by the handler. In this case, the handler will match all files ending with the `.npz` extension.

When the handler is instantiated the complete filepath to the data file is passed in `__init__`. With this information, our handler extracts the filename of the data file and opens it using the `load()` function from Numpy. The handler will be initialized for every request, and immediately its `parse_constraints` method is called.

The `parse_constraints` method is responsible for building a dataset object based on information for the request available on `environ`. In this simple handler we simply built a `DatasetType` object with the entirety of our dataset, i.e., we added *all data from all variables* that were available on the `.npz` file. Some requests will ask for

only a few variables, and only a subset of their data. The easy way parsing the request is simply passing the complete dataset together with the `QUERY_STRING` to the `constrain()` function:

```
return constrain(dataset, environ.get('QUERY_STRING', ''))
```

This will take care of filtering our dataset according to the request, although it may not be very efficient. For this reason, handlers usually implement their own parsing of the Opendap constraint expression. The SQL handler, for example, will translate the query string into an SQL expression that filters the data on the database, and not on Python.

Finally, note that the handler is a WSGI application: we can initialize it with a filepath and pass it to a WSGI server. This enables us to quickly test the handler, by checking the different responses at `http://localhost:8001/.dds`, for example. It also means that it is very easy to dynamically serve datasets by plugging them to a route dispatcher.

Responses

If handlers are responsible for converting data into the Pydap data model, responses to the opposite: the convert from the data model to different representations. The Opendap specification defines a series of standard responses, that allow clients to introspect a dataset by downloading metadata, and later download data for the subset of interest. These standard responses are the DDS (Dataset Descriptor Structure), the DAS (Dataset Attribute Structure) and the DODS response.

Apart from these, there are additional non-standard responses that add functionality to the server. The ASCII response, for example, formats the data as ASCII for quick visualization on the browser; the HTML response builds a form from which the user can select a subset of the data.

Here is an example of a minimal Pydap response that returns the attributes of the dataset as JSON:

```
from simplejson import dumps

from pydap.responses.lib import BaseResponse
from pydap.lib import walk

class JSONResponse(BaseResponse):
    def __init__(self, dataset):
        BaseResponse.__init__(self, dataset)
        self.headers.append(('Content-type', 'application/json'))

    @staticmethod
    def serialize(dataset):
        attributes = {}
        for child in walk(dataset):
            attributes[child.id] = child.attributes

        if hasattr(dataset, 'close'):
            dataset.close()

        return [dumps(attributes)]
```

This response is mapped to a specific extension defined in its entry point:

```
[pydap.response]
json = path.to.my.response:JSONResponse
```

In this example the response will be called when the `.json` extension is appended to any dataset.

The most important method in the response is the `serialize` method, which is responsible for serializing the dataset into the external format. The method should be a generator or return a list of strings, like in this example. Note that

the method is responsible for calling `dataset.close()`, if available, since some handlers use this for closing file handlers or database connections.

One important thing about the responses is that, like handlers, they are also WSGI applications. WSGI applications should return an iterable when called; usually this is a list of strings corresponding to the output from the application. The `BaseResponse` application, however, returns a special iterable that contains both the `DatasetType` object and its serialization function. This means that WSGI middleware that manipulate the response have direct access to the dataset, avoiding the need for deserialization/serialization of the dataset in order to change it.

Here is a simple example of a middleware that adds an attribute to a given dataset:

```
from webob import Request

from pydap.model import *
from pydap.handlers.lib import BaseHandler

class AttributeMiddleware(object):

    def __init__(self, app):
        self.app = app

    def __call__(self, environ, start_response):
        # say that we want the parsed response
        environ['x-wsgiorg.want_parsed_response'] = True

        # call the app with the original request and get the response
        req = Request(environ)
        res = req.get_response(self.app)

        # get the dataset from the response and set attribute
        dataset = res.app_iter.x_wsgiorg_parsed_response(DatasetType)
        dataset.attributes['foo'] = 'bar'

        # return original response
        response = BaseHandler.response_map[ environ['pydap.response'] ]
        responder = response(dataset)
        return responder(environ, start_response)
```

The code should actually do more bookkeeping, like checking if the dataset can be retrieved from the response or updating the `Content-Length` header, but I wanted to keep it simple. Pydap comes with a WSGI middleware for handling server-side functions (`pydap.wsgi.ssf`) that makes heavy use of this feature. It works by removing function calls from the request, fetching the dataset from the modified request, applying the function calls and returning a new dataset.

Templating

Pydap uses an [experimental backend-agnostic templating API](#) for rendering HTML and XML by responses. Since the API is neutral Pydap can use any templating engine, like [Mako](#) or [Genshi](#), and templates can be loaded from disk, memory or a database. The server that comes with Pydap, for example, defines a `renderer` object that loads Genshi templates from a directory `templatedir`:

```
from pydap.util.template import FileLoader, GenshiRenderer

class FileServer(object):
    def __init__(self, templatedir, ...):
        loader = FileLoader(templatedir)
        self.renderer = GenshiRenderer(options={}, loader=loader)
```

```
...  
  
def __call__(self, environ, start_response):  
    environ.setdefault('pydap.renderer', self.renderer)  
    ...
```

And here is how the HTML response uses the renderer: the response requests a template called `html.html`, that is loaded from the directory by the `FileLoader` object, and renders it by passing the context:

```
def serialize(dataset):  
    ...  
    renderer = environ['pydap.renderer']  
    template = renderer.loader('html.html')  
    output = renderer.render(template, context, output_format='text/html')  
    return [output]
```

(This is actually a simplification; if you look at the code you'll notice that there's also code to fallback to a default renderer if one is not found in the `environ`.)

License

Copyright (c) 2003–2010 Roberto De Almeida

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED “AS IS”, WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

CHAPTER 4

Indices and tables

- `genindex`
- `modindex`
- `search`