Import data using OPeNDAP

Function:	Import data from an OPeNDAP server directly into Delft-FEWS
Where to Use?	This can be used for importing data into the Delft-FEWS system.
Why to Use?	The advantage of importing data directly from an OPeNDAP server, as opposed to importing local files, is that the files do not have to be stored locally. Furthermore if only part of a file is needed, then only that part will be downloaded instead of the entire file. This can save a lot of network bandwidth (i.e. time) for large data files.
Preconditions:	The data to import needs to be available on an OPeNDAP server that is accessible by the Delft-FEWS system.
Outcome(s):	The imported data will be stored in the Delft-FEWS dataStore.
Available since:	Delft-FEWS version 2011.02

Contents

Overview

OPeNDAP (Open-source Project for a Network Data Access Protocol) can be used to import NetCDF or GRIB data from an OPeNDAP server directly into Delft-FEWS. For more information on OPeNDAP see http://opendap.org/. Three types of NetCDF data can be imported: grid time series, scalar time series and profile time series. For more information on these specific import types see their individual pages: NETCDF-CF_GRID, NETCDF-CF_TIMESERIES and NETCDF-CF_PROFILE. Also see NetCDF formats that can be imported in Delft-FEWS and Available data types.

How to import data from an OPeNDAP server

Import configuration

Data can be imported into Delft-FEWS directly from an OPeNDAP server. This can be done using the Import Module. The following import types currently support import using OPeNDAP:

import type	usage
NETCDF-CF_GRID	Use this for importing grid time series that are stored in NetCDF format
NETCDF-CF_TIMESERIES	Use this for importing scalar time series that are stored in NetCDF format
NETCDF-CF_PROFILE	Use this for importing profile time series that are stored in NetCDF format
GRIB1	Imports grid time series data from grib1 format used by meteorological institutes.
GRIB2	Imports grid time series data from grib2 format used by meteorological institutes.

To instruct the import to use OPeNDAP instead of importing local files, specify a server URL instead of a local import folder. Below is an example import configuration with a serverUrl element.

```
<?xml version="1.0" encoding="UTF-8"?>
<timeSeriesImportRun xmlns="http://www.wldelft.nl/fews" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.wldelft.nl/fews http://fews.wldelft.nl/schemas/version1.0/timeSeriesImportRun.
xsd">
        <import>
                <general>
                        <importType>NETCDF-CF_GRID</importType>
                        <serverUrl>http://test.opendap.org/opendap/hyrax/data/nc/sst.mnmean.nc.gz</serverUrl>
                        <startDateTime date="2007-07-01" time="00:00:00"/>
                        <endDateTime date="2008-01-01" time="00:00:00"/>
                        <idMapId>OpendapImportIdMap</idMapId>
                        <missingValue>32767</missingValue>
                </general>
                <timeSeriesSet>
                        <moduleInstanceId>OpendapImport</moduleInstanceId>
                        <valueType>grid</valueType>
                        <parameterId>T.obs</parameterId>
                        <locationId>gridLocation1</locationId>
                        <timeSeriesType>external historical</timeSeriesType>
                        <timeStep unit="nonequidistant"/>
                        <readWriteMode>add originals</readWriteMode>
                </timeSeriesSet>
        </import>
</timeSeriesImportRun>
```

Here the serverURL is the URL of a file on an OPeNDAP server. For details on specifying the URL see Import data from a single file or Import data from a catalog below. The time series set(s) define what data should be imported into Delft-FEWS. Only data for the configured time series sets is downloaded and imported, all other data in the import file(s) is ignored. For more details see Import Module configuration options.

Id map configuration

The import also needs an id map configuration file, that contains a mapping between the time series sets in the import configuration and the variables in the file(s) to import. Below is an example id map configuration.

⚠ The external parameter id is case sensitive.

```
<?xml version="1.0" encoding="UTF-8"?>
<idMap xmlns="http://www.wldelft.nl/fews" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:
schemaLocation="http://www.wldelft.nl/fews http://fews.wldelft.nl/schemas/version1.0/idMap.xsd" version="1.1">
        <parameter internal="T.obs" external="sst"/>
        <location internal="gridLocation1" external="unknown"/>
</idMap>
```

Import data from a single file

- To import data from a single file on an OPeNDAP server, the correct URL needs to be configured in the serverUrl element. To get the correct URL for a single file:
    1. Use a browser to browse to a data file on an OPeNDAP server, e.g. http://test.opendap.org/opendap/hyrax/data/nc/sst.mnmean.nc.gz. html
    2. Copy the URL that is listed on the page after the keyword "Data URL:", e.g. http://test.opendap.org/opendap/hyrax/data/nc/sst.mnmean. nc.gz
    3. Paste this URL in the serverUrl element in the import configuration file.

Import data from a catalog

Instead of specifying the URL of a single file on an OPeNDAP server, it is also possible to specify the URL of a catalog. The files on an OPeNDAP server are usually grouped in folders and for each folder there is a catalog file available. The catalog usually contains a list of files and subfolders, but can also refer to other catalog files.

To filter the catalog for a specific period, a fileNameObservationDateTimePattern should be specified in combination with a relative or absolute period to import. Note that specifying the fileNameObservationDateTimePattern also causes the parser to use the observation date as indicated by the filename instead of the meta data that may or may not be included in the file content.

If the URL of a catalog file is specified for the import without configuring the fileNameObservationDateTimePattern element, then all files that are listed in the catalog will be parsed, and only the content that falls within the specified absolute or relative period will be imported in the database, which can be very inefficient. Other catalogs that are listed in the specified catalog are also imported recursively.

A catalog file is usually called catalog.xml. The URL of a catalog file can be obtained in the following way.

For a THREDDS opendap server:	First browse to a folder on the server. Then copy the current URL from the address line and replace ".html" at the end of the url by ".xml".
For a HYRAX opendap server:	First browse to a folder on the server. Then click on the link "THREDDS Catalog XML" on the bottom of the page. Then copy the current URL from the address line.

For example to import data from the folder http://test.opendap.org/opendap/hyrax/data/nc/ use the catalog URL http://test.opendap.org/opendap/hyrax/data/nc/catalog.xml in the import configuration. For example:

```
<import>
        <general>
                <importType>NETCDF-CF_GRID</importType>
                <serverUrl>http://test.opendap.org/opendap/hyrax/data/nc/catalog.xml</serverUrl>
                <fileNameObservationDateTimePattern>'file_prefix'yyyyMMdd'-S'HHmmss'???'<
/fileNameObservationDateTimePattern>
                <startDateTime date="2007-07-01" time="00:00:00"/>
                <endDateTime date="2008-01-01" time="00:00:00"/>
                <idMapId>OpendapImportIdMap</idMapId>
                <missingValue>32767</missingValue>
        </general>
        <timeSeriesSet>
                <moduleInstanceId>OpendapImport</moduleInstanceId>
                <valueType>grid</valueType>
                <parameterId>T.obs</parameterId>
                <locationId>gridLocation1</locationId>
                <timeSeriesType>external historical</timeSeriesType>
                <timeStep unit="nonequidistant"/>
                <readWriteMode>add originals</readWriteMode>
        </timeSeriesSet>
</import>
```

Import data for a given variable

An import file (local or on an OPeNDAP server) can contain multiple variables. For each time series set in the import configuration the import uses the external parameter id from the id map configuration to search for the corresponding variable(s) in the file(s) to import. If a corresponding variable is found, then the data from that variable is imported. Only data for the found variables is downloaded and imported, all other data in the import file(s) is ignored. For NetCDF files the external parameter id is by default matched to the names of the variables in the NetCDF file to find the required variable to import. There also is an option to use the **standard_name** attribute or **long_name** attribute of a variable in the NetCDF file as external parameter id. To use this option add the **variable_identification_method** property to the import configuration, just above the time series set(s). For example:

```
<import>
        <general>
                <importType>NETCDF-CF_GRID</importType>
                <serverUrl>http://test.opendap.org/opendap/hyrax/data/nc/sst.mnmean.nc.gz</serverUrl>
                <startDateTime date="2007-07-01" time="00:00:00"/>
                <endDateTime date="2008-01-01" time="00:00:00"/>
                <idMapId>OpendapImportIdMap</idMapId>
                <missingValue>32767</missingValue>
        </general>
        <properties>
                <string key="variable_identification_method" value="long_name"/>
        </properties>
        <timeSeriesSet>
                <moduleInstanceId>OpendapImport</moduleInstanceId>
                <valueType>grid</valueType>
                <parameterId>T.obs</parameterId>
                <locationId>gridLocation1</locationId>
                <timeSeriesType>external historical</timeSeriesType>
                <timeStep unit="nonequidistant"/>
                <readWriteMode>add originals</readWriteMode>
        </timeSeriesSet>
</import>
```

The **variable_identification_method** property can have the following values:

variable_identification_method	behaviour
standard_name	All external parameter ids are matched to the **standard_name** attributes of the variables in the NetCDF file to find the required variable(s) to import.
long_name	All external parameter ids are matched to the **long_name** attributes of the variables in the NetCDF file to find the required variable(s) to import.
variable_name	All external parameter ids are matched to the names of the variables in the NetCDF file to find the required variable(s) to import.

If the **variable_identification_method** property is not present, then **variable_name** is used by default. The **variable_identification_method** property currently only works for the import types NETCDF-CF_GRID, NETCDF-CF_TIMESERIES and NETCDF-CF_PROFILE.

> ⚠ Currently it is not possible to import data from the same variable in the import file to multiple time series sets in Delft-FEWS. If required, this can be done using a separate import for each time series set.

Import data for a given period of time

To import only data for a given period of time, specify either a relative period or an absolute period in the general section of the import configuration file. See relativeViewPeriod, startDateTime and endDateTime for more information. The import will first search the metadata of each file that needs to be imported from the OPeNDAP server. Then for each file that contains data within the specified period, only the data within the specified period will be imported. The start and end of the period are both inclusive.
This can be used to import only the relevant data if only data for a given period is needed, which can save a lot of time. However, for this to work the import still needs to search through all the metadata of the file(s) to be imported. So for large catalogs that contain a lot of files, it can still take a lot of time for the import to download all the required metadata from the OPeNDAP server.

Example: to import only data within the period from 2007-07-01 00:00:00 to 2008-01-01 00:00:00, add the following lines to the import configuration:

```
        <startDateTime date="2007-07-01" time="00:00:00"/>
        <endDateTime date="2008-01-01" time="00:00:00"/>
```

Alternatively you can use the relativeViewPeriod element so set a period to import relative to the T0. If you do this you can use the manual forecast dialog to set the period to import data from using the Cold/Warm state selection options.

Import data for a given forecast time

It often happens that on an OPeNDAP server there is data available for multiple forecasts with different forecast times. If the time periods of the forecasts overlap, then only one of the forecasts can be imported at a time. If there is a separate file for each forecast, then this can be done by specifying the URL of the required file in the import configuration. However, when such data is imported in an operational system, then the import URL should be changed each time a new forecast becomes available on the OPeNDAP server. If the URLs for the different forecasts contain the forecast time and only differ in forecast time, then the tags TIME_ZERO and/or RELATIVE_TIME_IN_SECONDS can be used to solve this problem. The import will replace any TIME_ZERO tags in the URL with the time zero (forecast time) of the current import run. Any RELATIVE_TIME_IN_SECONDS tags in the URL will be replaced with a time that equals (time0 + relativeTime), where time0 is the time zero (forecast time) of the current import run and relativeTime (specified in the tag) is a time relative to time0 in seconds (can be negative). The time is formatted using the dateFormat that is specified in the tag. This way different forecast data is imported each time the import runs for a different time zero.

> ⚠ If the import runs for a time zero for which there is no forecast data available on the OPeNDAP server, then the import will fail with an error message. Therefore make sure that the import only runs at the specific forecast times for which there is data available on the OPeNDAP server.

Example of an import URL with TIME_ZERO tags:

```
<serverUrl>http://nomads.ncep.noaa.gov:9090/dods/gfs/gfs%TIME_ZERO(yyyyMMdd)%/gfs_%TIME_ZERO(HH)%z</serverUrl>
```

Example of an import URL with RELATIVE_TIME_IN_SECONDS tags:

```
<serverUrl>http://nomads.ncep.noaa.gov:9090/dods/gfs/gfs%RELATIVE_TIME_IN_SECONDS(yyyyMMdd, -18000 )%/gfs_%
RELATIVE_TIME_IN_SECONDS(HH,-18000)%z</serverUrl>
```

For more information on how to use the TIME_ZERO and RELATIVE_TIME_IN_SECONDS tags see server url.

Import data for a given subgrid

⚠

⚠️ Importing data for a subgrid currently only works for regular grids from CF compliant NetCDF files using the NETCDF-CF_GRID import type as in the above example.

For NetCDF grids that are not fully compliant with the CF conventions, the NetcdfGridDataset import type can be used, which requires that the complete grid extent is imported.

This section only applies to the import of grid data. For data with a regular grid that is imported from a NetCDF file, it is in most cases not required to have a grid definition in the grids.xml configuration file. Because for regular grids the import reads the grid definition from the NetCDF file and stores the grid definition directly in the datastore of Delft-FEWS. If for the imported data there is no grid definition present in the grids.xml configuration file, then data for the entire grid is imported.
To import data for only part of the original grid, it is required to specify a grid definition in the grids.xml configuration file. The grid definition defines the part of the grid that needs to be imported. In other words the grid definition defines a subgrid of the original grid. In this case only data for the configured subgrid is downloaded and imported, the data for the rest of the original grid is ignored. The following restrictions apply:

- The subgrid must be fully contained within the original grid.
- The subgrid must have the same geodatum as the original grid.
- The cellwidth of the subgrid must be the same as the cellwidth of the original grid within a margin of 10 percent.
- The cellheight of the subgrid must be the same as the cellheight of the original grid within a margin of 10 percent.
- All cell centers in the subgrid must coincide with cell centers in the original grid within a certain margin.

For example to import data for a sub grid from the URL http://test.opendap.org/opendap/hyrax/data/nc/sst.mnmean.nc.gz use e.g. the following grid definition in the grids.xml file. In this example a subgrid of 5x5 cells is imported, where the cell center longitude coordinates range from 0 to 8 degrees and the cell center latitude coordinates range from 50 to 58 degrees.

```
<regular locationId="gridLocation1">
        <rows>5</rows>
        <columns>5</columns>
        <geoDatum>WGS 1984</geoDatum>
        <firstCellCenter>
                <x>0</x>
                <y>58</y>
        </firstCellCenter>
        <xCellSize>2</xCellSize>
        <yCellSize>2</yCellSize>
</regular>
```

For more information about the configuration of grid definitions in Delft-FEWS see Grids.

Import data from a password protected server

For importing data from a password protected OPeNDAP server, it is required to configure a valid username and password for accessing the server. This can be done by adding the user and password elements (see Import Module configuration options#user) to the import configuration, just after the serverUrl element.

Example of an import configuration with user and password elements:

```
<import>
        <general>
                <importType>NETCDF-CF_GRID</importType>
                <serverUrl>http://test.opendap.org/opendap/hyrax/data/nc/sst.mnmean.nc.gz</serverUrl>
                <user>kermit</user>
                <password>gr33n</password>
                <startDateTime date="2007-07-01" time="00:00:00"/>
                <endDateTime date="2008-01-01" time="00:00:00"/>
                <idMapId>OpendapImportIdMap</idMapId>
                <missingValue>32767</missingValue>
        </general>
        <timeSeriesSet>
                <moduleInstanceId>OpendapImport</moduleInstanceId>
                <valueType>grid</valueType>
                <parameterId>T.obs</parameterId>
                <locationId>gridLocation1</locationId>
                <timeSeriesType>external historical</timeSeriesType>
                <timeStep unit="nonequidistant"/>
                <readWriteMode>add originals</readWriteMode>
        </timeSeriesSet>
</import>
```

Import data from a server that uses SSL

For importing data from an OPeNDAP server that communicates using SSL, the certificate of the server either has to be validated by a known certificate authority (preferred) or it has to be added and trusted in the truststore of your local Delft-FEWS installation.

To add a certificate to the local Delft-FEWS truststore, first export the certificate file from the server using a browser, then import the certificate file into the truststore.

- To export the certificate of a server using Firefox:
    1. Browse to the server URL.
    2. Left click on the certificate icon.
    3. Choose More Information -> Show Certificate -> Details -> Export
    4. Follow the on screen instructions.

- To export the certificate of a server using Internet Explorer:
    1. Browse to the server URL.
    2. Left click on the lock icon.
    3. Choose View Certificates -> Details -> Copy to File
    4. Follow the on screen instructions.

To import the certificate file into the truststore use e.g. the following command on the command line:

```
 G:\java\jre6\bin\keytool.exe –import –v –alias aliasName –keystore G:\FEWS\client.truststore –storepass
d3lftf3ws –file fileName
```

where *fileName* is the pathname of the certificate file, *aliasName* is the alias to use for the certificate, *G:\java\jre6\bin\keytool.exe* is the pathname of the Java keytool.exe file (depends on your Java installation) and *G:\FEWS* is the path of the Delft-FEWS region home directory (depends on your Delft-FEWS installation). If the file client.truststore does not exist, then the above command will create it. After entering this command, the keytool will display details of the server certificate, type 'yes' to trust the certificate. If the above procedure was successful, then the keytool will display "Certificate was added to keystore". The truststore file called "client.truststore" in the Delft-FEWS region home directory is automatically read each time when Delft-FEWS starts, so Delft-FEWS may need to be restarted after the certificate has been added.

In the above command *d3lftf3ws* is the default password for the client.truststore file. By default Delft-FEWS expects the truststore to have this default password. If another password is used for the truststore, then that password must be set as the value of the java system property "javax.net.ssl.trustStorePassword". This is needed so that Delft-FEWS is able to get the password to access the truststore.

See also: ClientConfig XML File for Operator Client and Forecasting Shell Servers - 2018.02 and later#08RootConfigurationFilesforOperatorClientandForecastingShellServers-truststore

Known issues

Export of data

- It is not possible to export data directly using the OPeNDAP protocol, since the OPeNDAP protocol only supports reading data from the server. If it is required to export data from Delft-FEWS and make it available on an OPeNDAP server, then this can be done in two steps:
    1. setup a separate OPeNDAP server that points to a given storage location. For instance a THREDDS server, which is relatively easy to install. The OPeNDAP server picks up any (NetCDF) files that are stored in the storage location and makes these available for download using OPeNDAP.
    2. export the data to a NetCDF file using a Delft-FEWS export run. Export of grid time series, scalar time series and profile time series is supported (respectively export types NETCDF-CF_GRID, NETCDF-CF_TIMESERIES and NETCDF-CF_PROFILE). Set the output folder for the export run to the given storage location. That way the exported data will automatically be picked up by the OPeNDAP server.

Related modules and documentation

Internal

- Import Module
- Import Module configuration options
- NetCDF formats that can be imported in Delft-FEWS
- Available data types
- NETCDF-CF_GRID
- NETCDF-CF_TIMESERIES
- NETCDF-CF_PROFILE

External

- http://cfconventions.org/1.6.html
- OPeNDAP
- THREDDS