Table of Contents

What is ESGF Tools UI?

A tool that is integrated in [?ToolsUI NetCDF Java](#) for including a desktop client for [?ESGF services](#), with features like search of data along multiple nodes, download manager and metadata aggregation for allowing a full exploration through dataset services.

The added functionality to ToolsUI NetCDF is located in the ESGF tab. This tab contains 3 sub-tabs, each one with an specific functionality.

If you find any issue, problem or want to make a comment, go to [Issues](#)

NetCDF (4.3) Tools												_ □ ×
System Modes Debug Help												
Viewer	Writer	NCDump	Iosp	CoordSys	FeatureTypes	THREDDS	Fmrc	GeoTiff	Units	NcML	URLdump	ESGF

Login ■ Remaining time of credentials: 70,78 hours (Expire date: Sat May 31 10:19:37 CEST 2014)

Saved searches & Harvesting | Search | Downloads
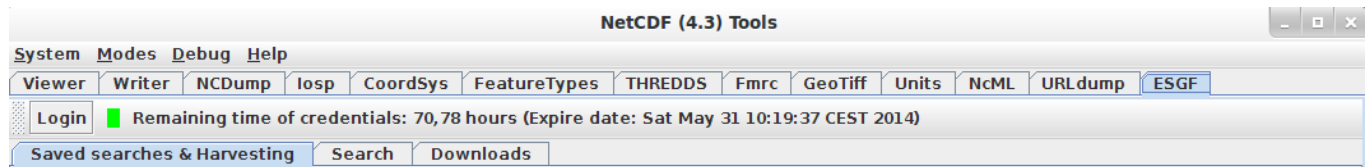
Getting started

Pre-requisites

JDK from Sun / Oracle (1.5 - newer) or OpenJDK 6

OpenJDK 7 is not supported

### Installation

1. Download the jar in [ESGFToolsUI-v0.6.1.jar](#)
   [Other versions..](#)
2. Go to download target directory:
   - In Windows:
     - Open ESGFToolsUI.jar
   - In Linux
     - In shell:

```
java -jar ESGFToolsUI.jar
```

Search of ESGF Climate Data

A search in ESGF returns the records that match the search constraints after querying to an index node. The ESGF search service is always served by a index node and this node is able to do local and distributed searches. In this tool, the distributed capabilities of the ESGF are used so any selected index node will be used to query all nodes in the ESGF system.

The search panel of ESGF climate data is in the "Search" sub-tab (be patient, it takes a while for the panel to load). The search constraints are defined by parameter configuration in the panel or by introduction of a free-text search query. The result is the number of records finding in the federation that satisfy the constraints of search (in the bottom right area of the panel). A record is the physical replica of a climate dataset that is stored on a data node.
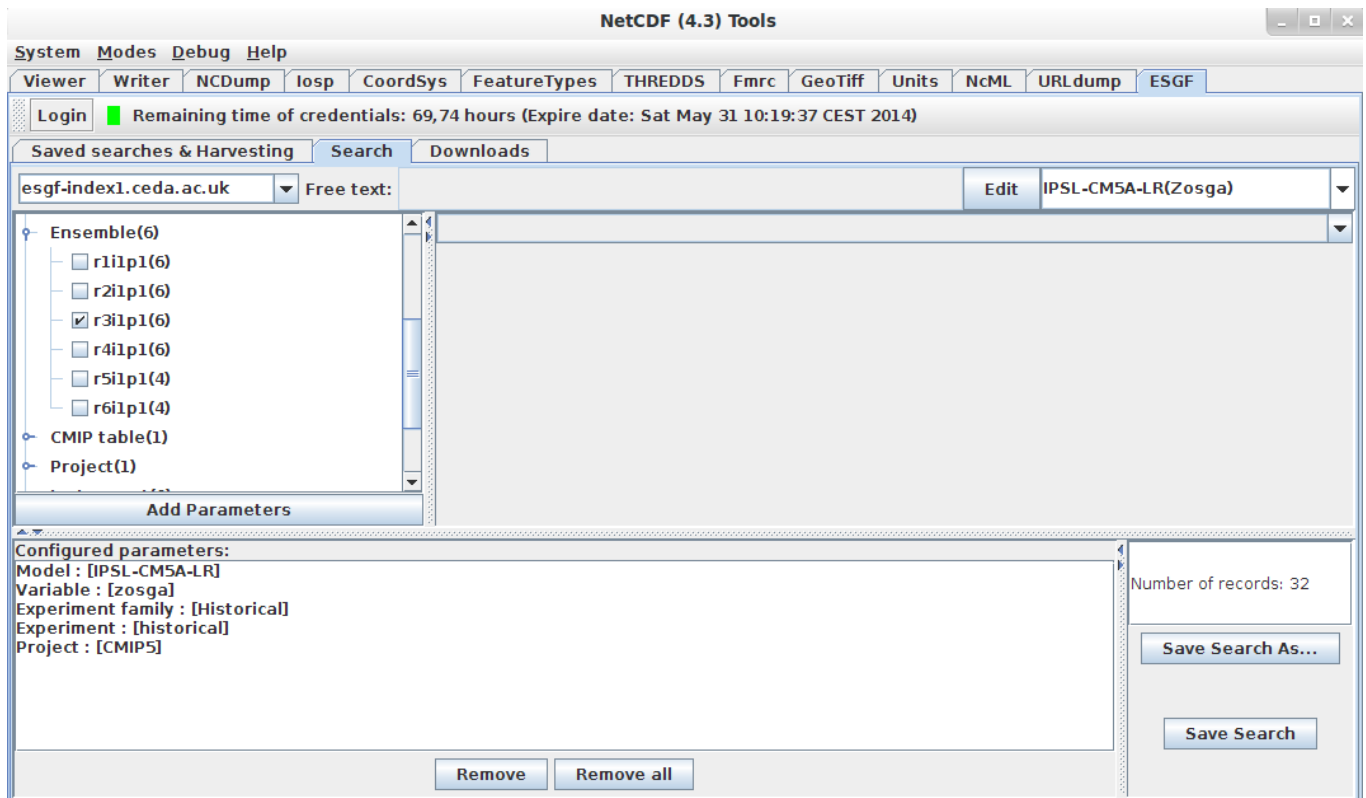
Selecting a index node
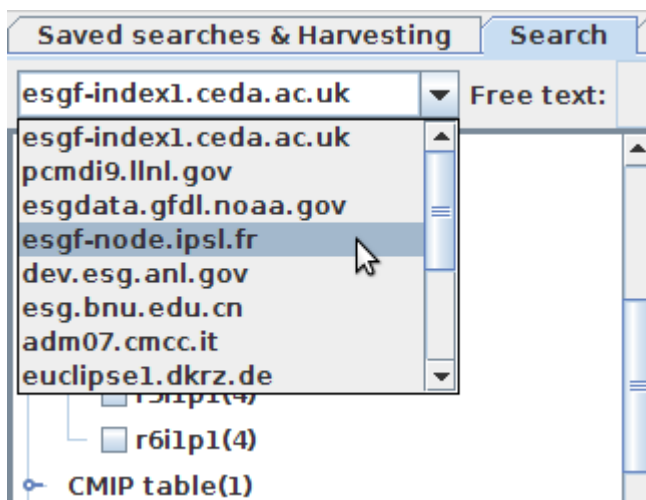
The searches are executed in a index node of the federation. Each node has different processing times for the response. For this reason, the user is allowed to select the index node where he/she prefers to send the search petition. The index node can be configured in the top left drop-down list. After selecting a new index node, the search panel is updated.

Selecting search parameters

The search parameters can be configured on the panel in two ways:

- By values selection in tree of parameters `double click`

- By parameter selection in drop-down list
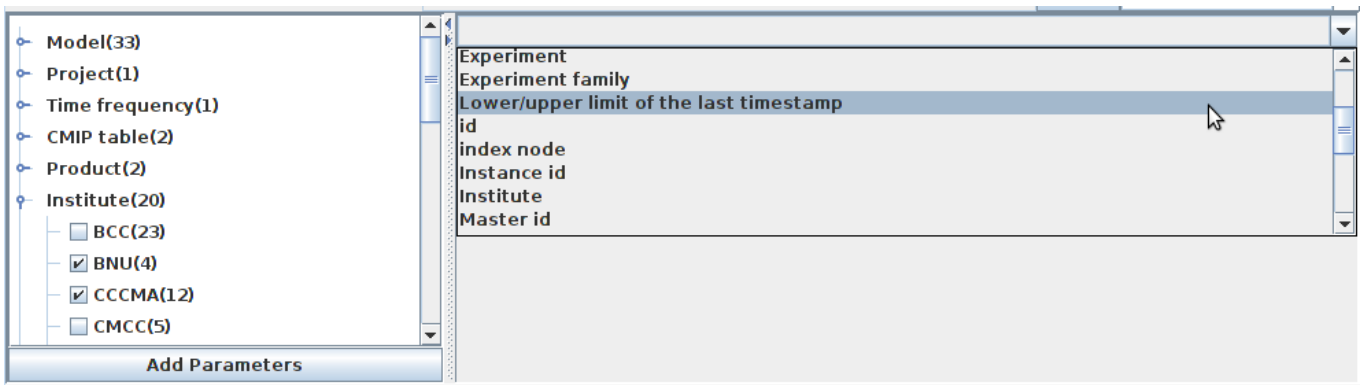
Tree of parameters

The tree of parameters shows the parameters with bounded values which are defined by the federation. In the first level is the name of the parameter and the number of its bounded values. In the second level are the parameter values, each one with the count of records that satisfy the previous configured constraints (restrictions that are listed in the "configured parameters" section) + the new constraint (parameter-value).

For each parameter can be selected more than one value `doing double click in the checkbox`. Finally, for adding the selected values on the tree in the search constraints you must click on "Add parameters" button. All values selected in a parameter are linked by logic OR and the selected parameters are linked by logic AND. Generally, the parameter configuration is as follows, where "P" is a parameter and "V" a value.

$$[(P_1-V_1)\vee(P_1-V_2)\vee...\vee(P_1-V_n)]\wedge[(P_2-V_2)\vee(P_2-V_2)\vee...\vee(P_2-V_n)]$$
$$\wedge[(P_n-V_1)\vee(P_n-V_2)\vee...\vee(P_n-V_n)]$$

```
For example, if are selected:

* Institute
 * BNU    ?
 * CCCMA ?
* Model
 * BNU-ESM ?

Then the result will be the number of records that belong to BNU-ESM model and made by BNU or CCCMA institutes i.e.
((Institute,BNU)V(Institute,CCCMA))??((Model,BNU-ESM))
```

After adding new parameters the search panel is updated (?Add parameters? button). Now in the tree, all parameters that already have values configured only shows this values. This happens because the allowed values ??for a parameter are always those that do not result in an empty set if they are selected. To configure a parameter previously configurated you can uncheck the previously configured values and then click on "Add parameters" or you can remove them in "Configured parameters" section.

Drop-down list of parameters

The parameter selection in drop-down list (top right) allows values configuration of some parameters. In this list are contained the parameters without bounded values and that for this reason can't be configured in the tree of parameters. Also contains the parameters with bounded values. After selecting a list parameter, below of the list will be displayed a specific configuration panel for the selected parameter.

In the figure below you can see the panel that is displayed when ?temporal range of data coverage? is selected. Finally, to be added this configuration click in the bottom button "Add parameter" ("Add temporal range parameter" in this case).

`Temporal range` parameter selection has bugs in ESGF search service (still not fixed), and for that, maybe the files could not be properly selected. In this case, the files must be selected manually

Configured parameters

All configured parameters are displayed in a specific panel in bottom left. In this panel are listed the configured parameters with its values assigned.

In the figure below can see that the current search is configured for searching climate data that belong to CMIP5 project, that were made by BNU or CCCMA institute, that have a time frequency of six hours, belongs to "historical" experiment and that have data between 15-01-1920 and 20-01-2000, and finally contains at least one of this variables: "hus", "ps", "psl", "ta".

This panel also allow to select a parameter and delete its configuration doing click on "Remove" and delete all parameters configuration doing click on "Remove all"

Using free text search

The free-text search o full text search allows to do a search rich in syntax, by arbitrary words, logic operations and wild-cards. In this case, the records searched will be records that contains metadata that are related with what is specified in the free text query.

In top center in the search panel can see the text box where can be introduced the query of free text search. The "Edit" button enables input by keyboard in the text box. The "Save" button adds the free text query parameter in the search. The new configured parameter will be displayed in the "configured parameters" section with the name "query".

The free-text query may have:

- Logic operations (`AND` & `OR`) between words
- Wildcard `*` for the words
- Parentheses `( )` as separators
- `metadata_name:word` to specify the metadata name where must match the given word

```
For example, the query:

   institute:CCCMA OR (id:*IPSL* AND (model:IPSL-CM5A-LR OR model:IPSL-CM5B-LR))

Will search all records in ESGF whose institute is CCCMA OR that satisfy the following
points:

1. Have an id that contains the word IPSL (id:*IPSL*)

2. Belong to IPSL-CM5A-LR model OR IPSL-CM5B-LR model
```

Save a search

A configured search can be saved for later for to be able do a harvesting of metadata and services of datasets that satisfy the search constraints. The save section is in right bottom of search panel. You can overwrite the selected search with "Save search" button or save it with a new name clicking on "Save Search As..." button.
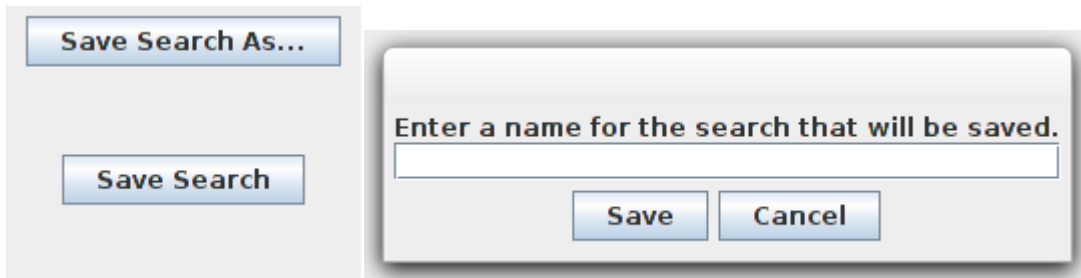
The name of search must be unique. Duplicated names aren't allowed

Select a saved search

You can select a previously saved search from the right top drop-down list. After select a search, the search panel is updated.
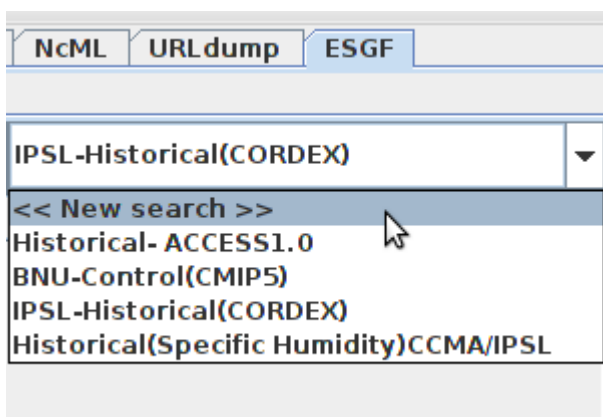
Reset search panel

To restart the search panel must be selected in the drop-down list the option: `<<New search>>`

Metadata Harvesting of ESGF Datasets of a search

The metadata harvesting will allow, after completed, download the datasets from multiple data nodes, and will allow exploring climate datasets without having to download the dataset itself. So we can know its nature in detail before selecting them to download or we can know the services offered from ESGF to access and/or explore the dataset. Some of these services can be explored from ToolsUI NetCDF Java as will be described after.

The harvesting is doing at `dataset` level.
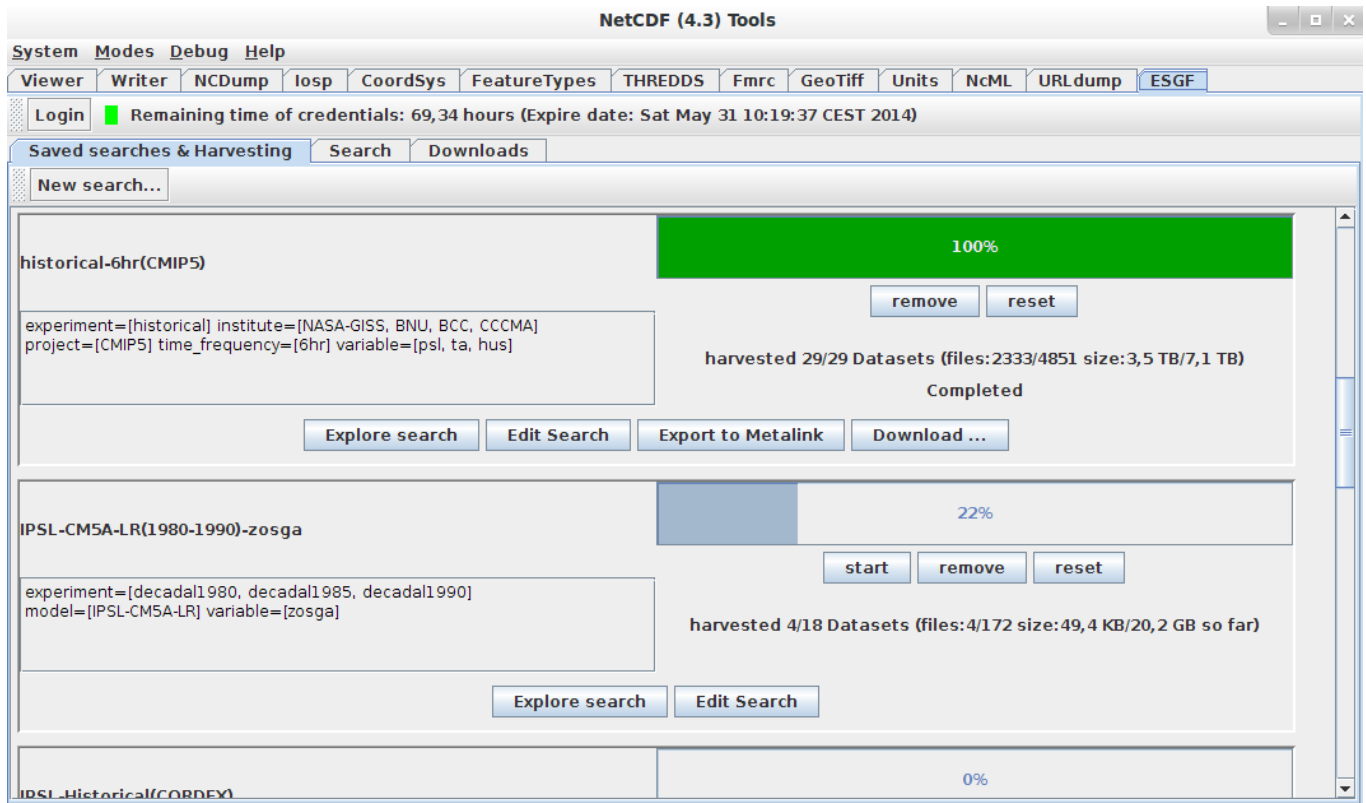
What is a ESGF Dataset?

Dataset is a climate data in a specific version stored by ESGF. One dataset may have several records i.e. the records are the physical replica of a dataset in a data node.

Also, the datasets are formed of files, i.e. dataset are a virtual container of data, so that the information is contained in files and sometimes in aggregations. The versions of datasets are generated when errors are found in the datasets, giving rise to a new corrected version.

Datasets, files and aggregations have replicas in data nodes. And the ESGF services (THREDDS, LAS, HTTP, GridFTP and OPeNDAP) are always served at replicas level. That's the reason why the harvesting must be done before the download.

Search Harvesting Panel

The harvesting panel may be selected clicking on the sub-tab "Search Harvesting". In this panel are deployed a list of searches and their harvesting states. Also, provides several options for flow control, complete exploration of harvested datasets and the posibility to do a manual selection of files to put them in download queue.

Each harvesting for a search is an element in the list deployed in the harvesting panel. In the figure below can see one search harvesting. In the left are the search data (name of search and configured parameters). In the right are the state data of the harvesting and the flow options, also shows the number of files that are selected from the total number of files and their sizes in bytes.

One harvesting is always of a completed dataset (with all files). However, a search may include only some files of the total in a dataset (e.g a search with less variables that there is in the dataset or a search in a range of time). That is why, by default, the files selected to future downloads (in the application or by generation of metalink) are always the files that satisfy ALL search constraints.

`Noteworthy` that this application allows to manually select files to download in case that you want to download files that are not selected by default or deselected some of these.
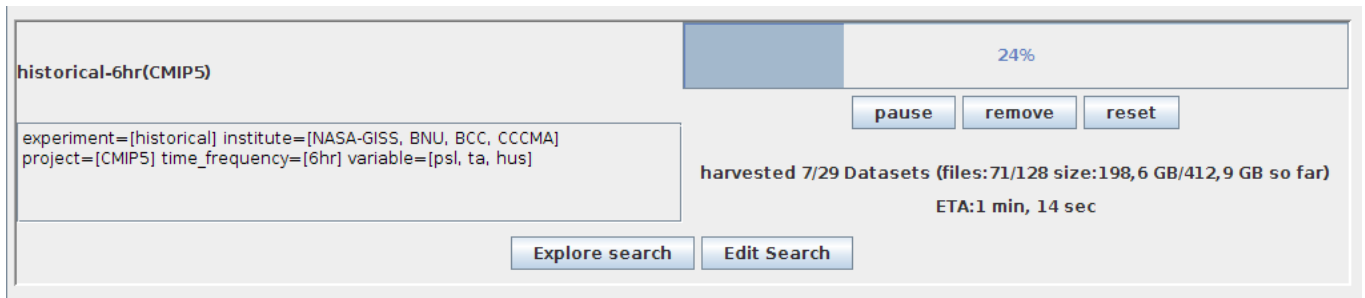
The flow control options are below progress bar:

- start - start the harvesting
- pause - pause the harvesting
- remove - remove the search and its harvesting (but the harvesting remains in file system)
- reset - remove all data of harvesting in file system and put the state of harvesting to zero)

In the center bottom there are the options to explore and download the harvested datasets:

Explore Search	is always visible and allows explore and put to download a individual dataset. Also, view a individual state of harvesting of a dataset
Edit Search	is always visible and allows edit a search in the Search Panel
Export to metalink	is visible when harvesting of search is completed. Allows to generate a metalink file with the files and its resources (replicas in data nodes)
Download ...	is visible when harvesting of search is completed. Allows to put to download all files that satisfy the constraints of the search or to put to download a set of files manually selected of the search

Exploring search harvesting

To exploring the search harvesting of a search you must click on Explore Search option that is always visible in harvesting panel. And then a window pops up with the exploring options.

This window provides a paginated list of harvesting states of datasets that belong to the search. Each dataset are identify by its instance_id. The last value comma separated is the version of dataset. The exploring options in this window are provided at dataset level. Each dataset have options, and these are only provided when the dataset harvesting is completed.

In ESGF the dataset are identify by id, instance_id and master_id. And in this application the "instance_id" is the more important because identify all replicas across federation and is specific to each version.

For more information ?https://github.com/ESGF/esgf.github.io/wiki/ESGF_Search_REST_API#identifiers

In figure below can see the exploring window and its options.

Harvesting view

`Harvesting View` option allows to explore a dataset in a tree view. Here you can explore the metadata for datasets that include files, replicas of datasets, replicas of files and finally, ESGF services offered for explore them (by URL endpoint).

- LAS and THREDDS are services of datasets offered at replica level.
- HTTP, OPeNDAP and GridFTP are services of files offered at replica level.

Open metadata

`Open metadata` option allows to explore an abstract of metadata of a dataset in a table view.

Metadata	Values
CF_STANDARD_NAME	[surface_air_pressure, specific_humidity, air_temperature, eastward_wind, northward_wind]
CMOR_TABLE	[6hrLev]
DATASET_ID_TEMPLATE_	[cmip5.%(product)s.%(institute)s.%(model)s.%(experiment)s.%(time_frequency)s.%(realm)s.%(cmor_table)s.%(ensemble)s]
DATETIME_START	1850-01-01T03:00:00Z
DATETIME_STOP	2005-12-31T21:00:00Z
DESCRIPTION	[IPSL-CM5A-LR model output prepared for CMIP5 historical]
DRS_ID	[cmip5.output1.IPSL.IPSL-CM5A-LR.historical.6hr.atmos.6hrLev.r5i1p1]
EAST_DEGREES	356.25
ENSEMBLE	[r5i1p1]
EXPERIMENT	[historical]
EXPERIMENT_FAMILY	[All, Historical]
FORCING	[Nat,Ant,GHG,SA,Oz,LU,SS,Ds,BC,MD,OC,AA]
FORMAT	[netCDF, CF-1.4]
INSTANCE_ID	cmip5.output1.IPSL.IPSL-CM5A-LR.historical.6hr.atmos.6hrLev.r5i1p1.v20111119
INSTITUTE	[IPSL]
LATEST	true
MASTER_ID	cmip5.output1.IPSL.IPSL-CM5A-LR.historical.6hr.atmos.6hrLev.r5i1p1
METADATA_FORMAT	THREDDS
MODEL	[IPSL-CM5A-LR]
NORTH_DEGREES	90.0
NUMBER_OF_AGGREGATIONS	8
NUMBER_OF_FILES	68
PRODUCT	[output1]
PROJECT	[CMIP5]
REALM	[atmos]

Close

Download Dataset - File selection window

`Download Dataset...` allows to manually select the files that will be put to download. In the top of window can see the id of dataset, the number of files selected, the total number of files of the dataset, the size in bytes of the sum of selected files and the total size of the whole dataset.

This windows displays a list of files and its size in bytes. You can select manually the files that you want download doing click in the check-box associated in each file in the list.

cmip5.output1.CCCma.CanESM2.historical.6hr.atmos.6hrPlev.r4i1p1.v20120618
files: (14/28) size: (11,6 GB/29,0 GB)

- ☑ psl_6hrPlev_CanESM2_historical_r4i1p1_194101010000-195012311800.nc (456,6 MB)
- ☑ psl_6hrPlev_CanESM2_historical_r4i1p1_195101010000-196012311800.nc (456,6 MB)
- ☑ psl_6hrPlev_CanESM2_historical_r4i1p1_196101010000-197012311800.nc (456,6 MB)
- ☑ psl_6hrPlev_CanESM2_historical_r4i1p1_197101010000-198012311800.nc (456,6 MB)
- ☑ psl_6hrPlev_CanESM2_historical_r4i1p1_198101010000-199012311800.nc (456,6 MB)
- ☑ psl_6hrPlev_CanESM2_historical_r4i1p1_199101010000-200012311800.nc (456,6 MB)
- ☑ psl_6hrPlev_CanESM2_historical_r4i1p1_200101010000-200512311800.nc (228,3 MB)
- ☑ ta_6hrPlev_CanESM2_historical_r4i1p1_194101010000-195012311800.nc (1,3 GB)

Deselect all | Select all | Filter by constraints of search | Download selected | Close

This window also provides the next options:

- Deselect all
- Select all
- Filter by constraints of search - select only files that satisfy the constraints of search
- Download selected - put to download the files selected in this window

When you click on `Download selected` a window asks for destination path.
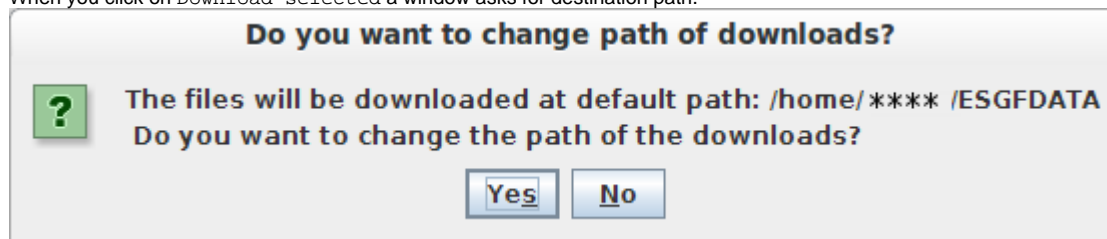
Export to Metalink

To export the search to Metalink you must click on "Export to Metalink". This option is is visible when harvesting of search is completed. And then a window pops up with the save options.

The files that will be include in the Metalink file are only the files that satisfy the constraints of search.

Metalink is an extensible metadata file format that describes one or more computer files available for download. It specifies files appropriate for the user's language and operating system; facilitates file verification and recovery from data corruption; and lists alternate download sources (mirror URIs). For more info: ?Metalink, ?Metalink-Wikipedia

Exist many clients that supports Metalink: ?Aria2, ?GetRight, ?DownloadThemAll, ?Orbit Downloader, etc.

Note that if you want use a external client to download ESGF files you must configure the client for using the `ESGF certificates`.

Download a Search

To put the files that belongs to a search in downloads queue you must click on `Download ....` This option is visible when harvesting of search is completed. And then a window pops up with the download options.

In this window (that you can see below) are displayed the list of datasets that satisfy the constraints of search. Each dataset have: id, description, the number of files that satisfy the search constraints with respect the total number of files and the size in disk of the selected files with respect the total size in disk of dataset.

For download all files that satisfy the constraints (without manually selection) you must click on "Download" button that is in the bottom of window. Then a dialog box will be showed to confirm the download and select the destination path.

For manually selection you may click on "Select file to Download", that displays a file selection window as explained in [File selection window](#)

Download of ESGF Datasets

To download ESGF Datasets you must done previously a harvesting of a search. The downloads panel is in "Download" tab.

You can add to the download queue:

- A complete search - all files of datasets that satisfy the constraints of the search (view [Download a Search](#))
- A personalized file selection of the set of datasets that satisfy the constraints of search (view [File selection window](#))

  The most of data in ESGF require credentials to access them. To access the user must have:

  1. A ESGF Account (view [ESGF Login](#))

  2. This account have to be authorized to access the desired data. This authorization is done by control groups (each account can belong a many groups). Each group is authorized for download a set of data.

To read more about register in ESGF go to ([?How to register and download data from esgf](#))

The downloads panel displays a list of downloads. Each element of the list is the grouping of files by dataset. The files of each dataset can be showed or hided. Each element have a progress bar and file elements displays the data node from where the download is acceded.

The files can be in five states:

- No completed: When the download isn't complete. It is displayed with a blue progress bar and a percentage
  - CREATED: File has just been added
  - PAUSED: File has paused
  - WAITING/READY: File is in waiting for be downloaded
  - DOWNLOADING: File is downloading
  - UNAUTHORIZED: When the user hasn't been authorized to download a file. It is displayed with a yellow progress bar with the message "UNAUTHORIZED".
- Completed : When the download is completed.
  - FINISHED: Download is successful. It is displayed with a green progress bar.

- FAILED : When happens some error in the current data node. It is displayed with a red progress bar.
- CHECKSUM_FAILED: When the validity algorithm of file failed. It is displayed with a gray progress bar with the message "CHECKSUM_FAILED".

If you is logged in ESGF and there are files with "Unauthorized state" is because the groups they are associated with the user account in the federation do not have permissions for that data.

To join a group with the needed permissions for a unauthorized file, view Join a group
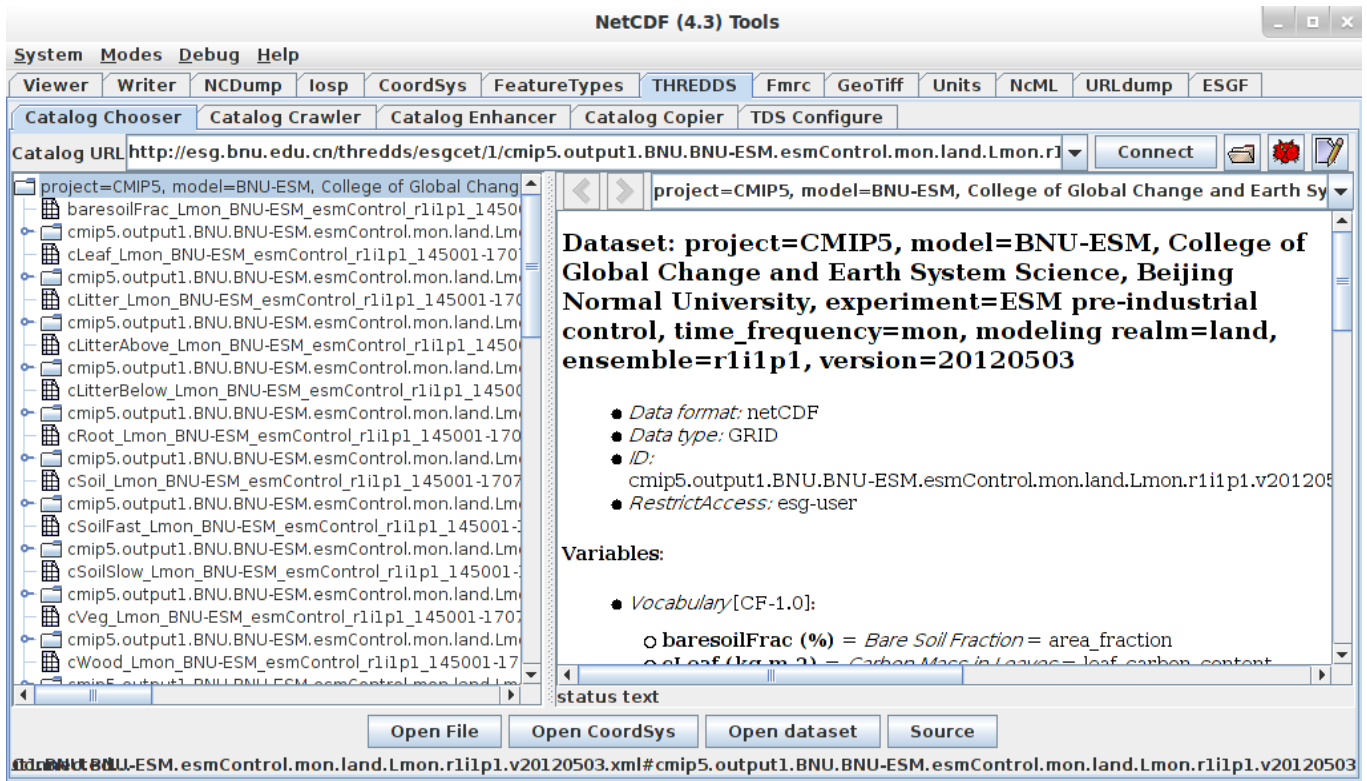
Context menu of downloads

The datasets and the files have a context menu. This menu are displayed after right-clicking in a dataset or file element. Depending on the state of download are enabled different options:

| Option	Description
`Start download/Start all file downloads`	Start the download
`Pause download/Pause all file downloads`	Pause the download
`Reset`	To reset all download. Remove file/files of file system (disk) and reset its/their status
`Retry/Retry download in failed files`	To retry download. In a file element, you must select a data node and then the download is reinitialized. In a dataset element, the download is retry in all files with failed state
`Remove`	To remove a file or a dataset and their files of the downloads queue. This option doesn't delete the files from disk
`Open download URL in browser` (only file)	Open the url of the HTTP service from which the file is being downloaded.
`File info/Dataset info`	To display info of file/dataset
`Services`	To access services of files/datasets (the services will be described in the next table)

Services of context menu of downloads

ESGF Login : Obtaining credentials with ESGF Account

In the top of ESGF panel there is a login toolbar. This toolbar shows a red icon if you aren't logged or a green icon if you are logged. Login allows retrieve the credentials needed to download the most of data in the federation. To login you must select the "Login" button.
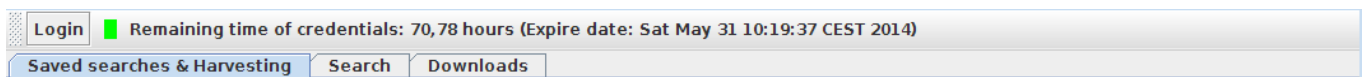
After clicking in "Login", a window is displayed, in this window if you are logged in the bottom of panel is displayed `You are logged` along with information validity time remaining for the credentials that are being used by the application. Else if you still aren't logged or you want to switch accounts, you must select the identity provider node from the dropdown list and after type in the box the account user name.
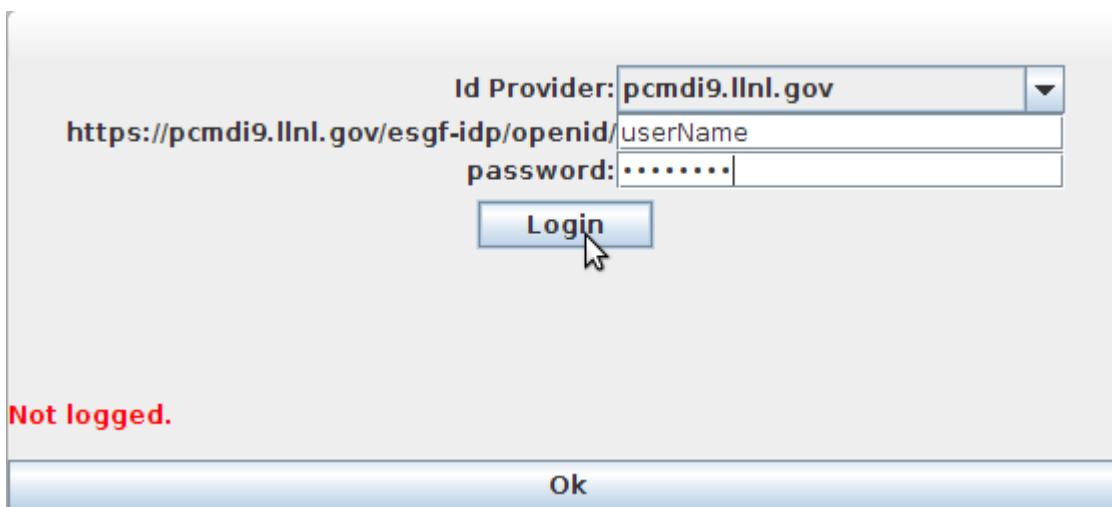
You can also paste the `OpenID URL` directly if selected in the drop down list `<< Custom OpenID URL >>`.

The session starts after pressing the "Login" button. If the login failed then It will be notified by a message on the bottom of the window.

Join a group with the needed permissions for a unauthorized file

Check that you are logged in. If you are logged and the `Unauthorized` state remains then you must do the follow steps

1. Deploy the context menu of the file in state "Unauthorized?
2. Select "Open download URL in browser" option
3. A ESGF site will be loaded in the browser. Select join to the group.
4. Go to "Login" Panel in Tools UI
5. Login with you account again (to retrieve new credentials)

    `If you are logged and you have the specific group permissions.` Then, probably is an error in data node.

    Reset the download and try in another data node