



ESGF

Earth System Grid Federation

Tutorial



INDEX

- ◉ Introduction
- ◉ Setting up an account
- ◉ Searching and accessing
- ◉ Download
- ◉ Advanced features



Index

- ◉ Introduction
 - ◉ IT Challenges of Climate Change Research
 - ◉ ESGF Introduction
 - ◉ ESGF System Architecture
 - ◉ ESGF Security Infrastructure
- ◉ Setting up an account
- ◉ Searching and accessing
- ◉ Download
- ◉ Advanced features

IT Challenges of Climate Change Research

The study of the climate change is one of the most critical scientific problems of our time: causes, predictions, and effect on Earth ecosystem and human society

- Supporting scientific research on climate change involves non-trivial IT challenges:
 - Multiple data sources
 - Model output (CMIP5, CORDEX)
 - Heterogenous observations (obs4MIP)
 - In-situ, satellites, vertical profiles, ...
 - Reanalysis data (ana4MIPs)
 - Very large, ever increasing data volumes
 - Current archives are in PB (e.g. CMIP5 is planned to have 2.5PB)
Increasing to EB in the next 5 years!!
 - Next generation satellites will collect many **TB/day**

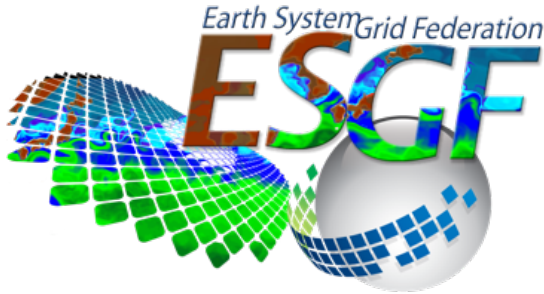


IT Challenges of Climate Change Research

- Data are stored and managed at many sites around the world
- Must serve multiple communities (scientist, politicians, educators, private sector)
- Data must be analyzed at multiple scales (global, regional and local)

ESGF aims at facilitating climate change research by addressing and solving these challenges

ESGF Introduction



The Earth System Federation (ESGF) is a multi-agency, international collaboration of people and institutions working together to build an open source software infrastructure for the management and analysis of Earth Science data on a global scale

- Started with *DoE Earth System Grid project*
 - Expanded to include groups and institutions that are involved in management of climate data around the world
 - Recently evolved into a next generation architecture Peer-to-Peer (P2P) system
 - Recommendation from the WCRP to be supported operationally and as a facto standard for future MIPs

ESGF System Architecture

ESGF is a system of distributed and federated Nodes that interact dynamically through a Peer-to-Peer (P2P) paradigm

- **Distributed**

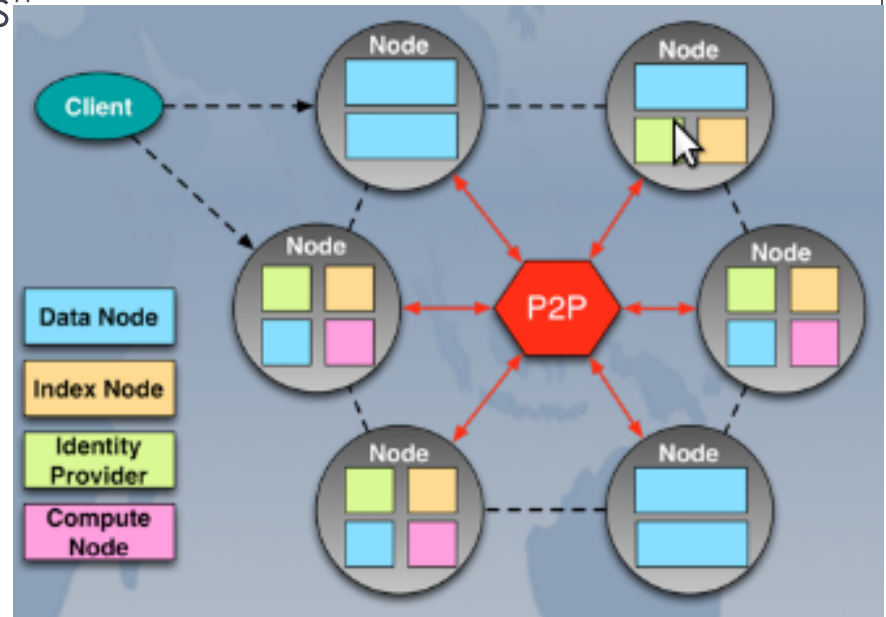
- Data and metadata are published, stored and served from multiple centers “Nodes”

- **Federated**

- Nodes interoperate
 - Common services, protocols and APIs
 - Establishment of mutual trust relationships

- **Dynamic**

- Nodes can join/leave federation dynamically

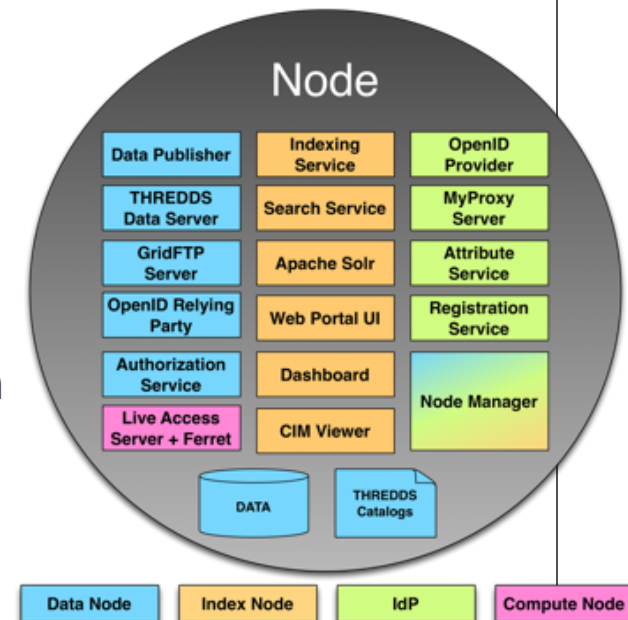


ESGF Node Architecture

Internally each ESGF node is composed of services and applications that collectively enable data and metadata access and user management

Software components are grouped into four areas of functionality (a.k.a. Flavours):

- Data Node:
 - Secure data publication and access
- Index Node:
 - Metadata indexing and searching
 - Web portal UI to drive human interaction
 - Admin dashboard
 - Model metadata viewer plugin
- Identity Provider:
 - User Authentication and Group membership
- Compute Node:
 - Analysis and Visualization





ESGF Security Infrastructure

Purpose: enable access to distributed data resources by implementing local authentication and authorization policies

- ESGF Security Highlights
 - Based on standards:
 - OpenID, SAML, SSL, X509
 - Language neutral
 - Java and Python implementations
 - Non-intrusive
 - Filters and libraries
 - Requires digital trust relationships among Nodes

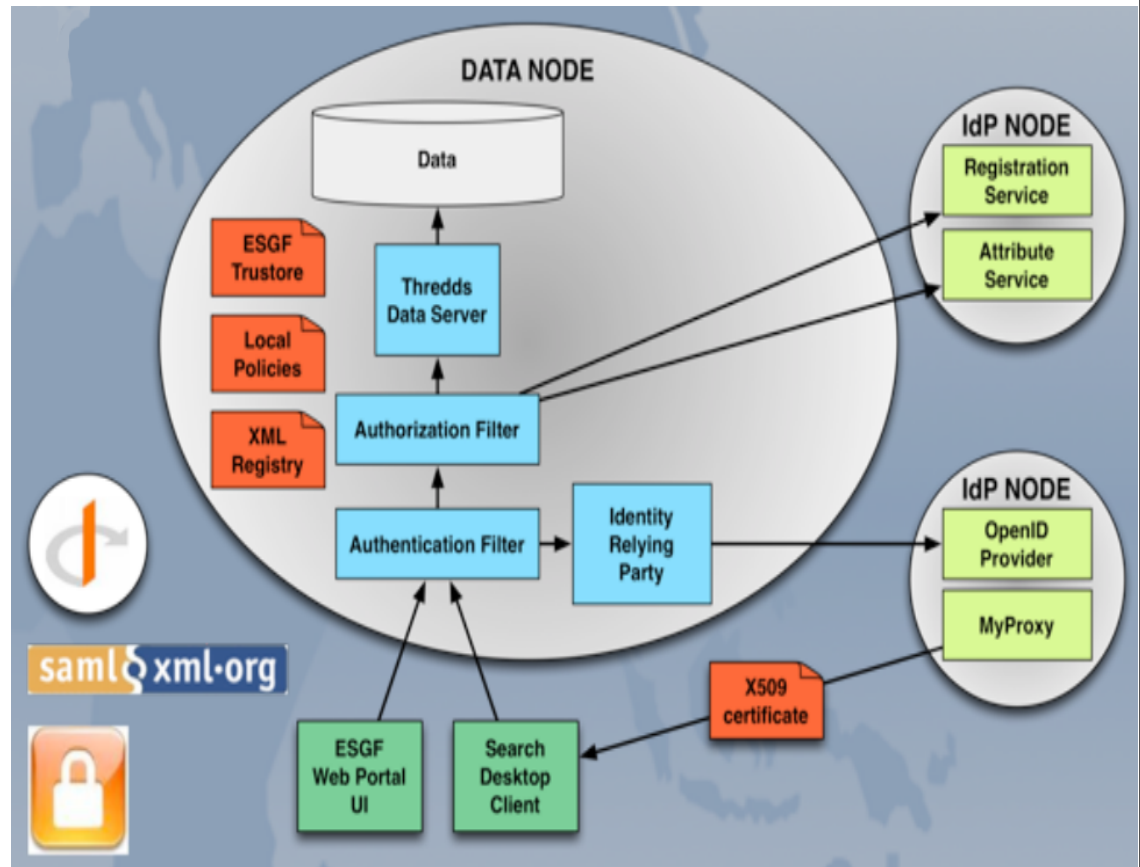
ESGF Security Infrastructure

- **Single-Sign-On the ESGF**

- OpenID for browsers
- X509 certificates for desktop clients (PKI)

- **Distributed access control**

- XML registry tracks security services endpoints
- User attributes are propagated as digital SAML statements





Index

- ◉ Introduction
- ◉ Setting up an account
 - ◉ Browser Support
 - ◉ Register
 - ◉ Enroll in CMIP5 group
- ◉ Searching and accessing
- ◉ Download
- ◉ Advanced features

Browser Support



Firefox 7+



Chrome 16+



Safari 5+

*In some ESGF sites
could fail to display the
login page*

Register

- Primary Nodes

- PCMDI

- <http://pcmdi9.llnl.gov/>

- BADC

- <http://esgf-index1.ceda.ac.uk>

- DKRZ








- <http://esgf-data.dkrz.de>

- NCI

- <http://esg2.nci.org.au>



Peer Nodes

-
-  [ANL Node](#) 
 -  [BADC Node](#) 
 -  [CMCC Node](#) 
 -  [DKRZ Node](#) 
 -  [DKRZ CMIP5 Node](#) 
 -  [IPSL Node](#) 
 -  [NASA-GSFC Node](#) 
 -  [NASA-JPL Node](#) 
 -  [NCI Node](#) 
 -  [ORNL Node](#) 
 -  [PCMDI Node](#) 



Welcome to this ESGF P2P Node

Welcome to the new CMIP5 distributed archive. Our new ESGF peer-to-peer (P2P) enterprise system is now the official site for CMIP5 model output. The old [gateways](#) will remain active and output from all models will continue to be available until the end of July 2012. Please send e-mail to esgf-user@lists.llnl.gov to report bugs and provide feedback.



Quick Search

Keyword:

Search

Advanced Search (Category, Geospatial, Temporal, and more)...



About esgf-pcmdi-9

The PCMDI mission is to develop improved methods and tools for the diagnosis and intercomparison of general circulation models (GCMs) that simulate the global climate. The need for innovative analysis of GCM climate simulations is apparent, as increasingly more



Resources



Quick Links

- [Create Account](#)
- [MyProxyLogin](#)
- [Expert Search \(XML\)](#)



Register

- Name
- Email
- User Name
 - Can contain only digits and numbers
- Password
 - Must contain at least one letter and one number
 - At least 6 characters long.
- **Organization**
- City
- State
- Country

ESGF User Registration

First Name

Middle Name

Last Name

Email

User Name

Username can contain only digits and numbers.

Password

*Password must contain at least one letter and one number,
and be at least 6 characters long.*

Confirm Password

Organization

City

State

Country

Submit

Register

- An OpenID account is assigned
 - Can be used to login through the whole ESGF
 - URL string:
<https://pcmdi9.llnl.gov/esgf-idp/openid/MariaFerGon>
- Registration is effective immediately
 - No confirmation email



Home Search Tools Login

SUCCESS.

Congratulations, your account is active immediately.

You can now [log in](#) with your openid: <https://pcmdi9.llnl.gov/esgf-idp/openid/MariaFerGon> .

Guest User

ESGF P2P Version 1.4.1-brooklyn_college-release-2-g2fe3e51-master [fe-2.2.8]

[Privacy Policy & Legal Notice](#) | [Contact ESGF](#)

Register

- An OpenID account is assigned
 - Can be used to login through the whole ESGF
 - URL string:
<https://pcmdi9.llnl.gov/esgf-idp/openid/MariaFerGon>
- Registration is effective immediately
 - No confirmation email



Home Search Tools Login

SUCCESS.

Congratulations, your account is active immediately.

You can now [log in](https://pcmdi9.llnl.gov/esgf-idp/openid/MariaFerGon) with your openid: <https://pcmdi9.llnl.gov/esgf-idp/openid/MariaFerGon> .

Guest User

ESGF P2P Version 1.4.1-brooklyn_college-release-2-g2fe3e51-master [fe-2.2.8]

[Privacy Policy & Legal Notice](#) | [Contact ESGF](#)

Account Home for user MariaFerGon

About

Username: MariaFerGon

Last Name: Fernández González **Middle Name:** N/A **First Name:** Maria

Email address: mail@mail.com

Organization: UC

City: Santander

State: Cantabria

Country: SPAIN

Openid: <https://pcmdi9.llnl.gov/esgf-idp/openid/MariaFerGon>

Domain Name: N/A

?

Group Administered

Group name	Description	Role
------------	-------------	------

Change password

Old password:

New password:

Confirm password:

Change password

Enroll in CMIP5 Group

- There is no direct way to do it
 - To enroll in a group is necessary to try to download restricted data
 - And then...

To obtain access to these data, please register with one of the following groups:

Status: not registered		
Group: CMIP5 Research	Role: user	<input type="button" value="Register"/>
Group: CMIP5 Commercial	Role: user	<input type="button" value="Register"/>
Group: NASA OBS	Role: user	<input type="button" value="Register"/>

Thank you for your interest in accessing these data.

Your openid: <https://pcmdi9.llnl.gov/esgf-idp/openid/MariaFerGon>

- A menu of groups that are allowed to access the data is shown:

http://esg-datanode.jpl.nasa.gov/thredds/fileServer/esg_dataroot/obs4

[MIPs/observations/atmos/husNobs/mon/grid/NASA-JPL/AIRS/v20110608/husNobs_AIRS_L3_RotStd_v5_200209_201105.nc](#)

Account Home for user MariaFerGon

About

Username: MariaFerGon

Last Name: Fernández González **Middle Name:** N/A **First Name:** Maria

Email address:

Organization:

City: Santander

State: Cantabria

Country: SPAIN

Openid: <https://pcmdi9.llnl.gov/esgf-idp/openid/MariaFerGon>

Domain Name: N/A

Group Administered

MariaFerGon

Group name	Description	Role
CMIP5 Research	Users of CMIP5 data for non-commercial research purposes only	user



Index

- ◉ Introduction
- ◉ Setting up an account
- ◉ Searching and accessing
 - ◉ Free text search
 - ◉ Faceted search
 - ◉ Results selection
- ◉ Download
- ◉ Advanced features



Searching and accessing

- The search returns datasets, no returns files
 - Dataset are divided into files
- There are two search operations (not mutually exclusive)
 - Free text search
 - Entering arbitrary words
 - Rich search syntax: logical ops and wildcards
 - Faceted search
 - Using the pre-defined search categories
 - May select more than one category

Free Text Search

Current Selections

No search criteria selected



Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

[Temporal Search](#)
[Geospatial Search](#)
[Clear search constraints and datacart](#)
[Search Help](#)
[Search Controlled Vocabulary](#)

Faceted Search

Search Categories

Project

Institute

Model

SubModel

Instrument

Experiment Family

Experiment

SubExperiment

Time Frequency

Product

Realm



Search All Sites



Show All Replicas



Show All

Versions

Display datasets per page

[Add All Displayed to Datacart](#)

[Remove All Displayed from Datacart](#)

Results

Data Cart

To search datasets on the ESGF select "Search All Sites"

Free text search

- Logical operators **AND** and **OR** (by default multiple words are combined as OR)
 >> air AND temperature
- Quoted text matches the exact phrase
 >> "air temperature"
- Search for specific metadata field with **field:word**
 >> project:CMIP5 AND variable:hus
- The wildcard * matches a portion of a word, in all fields or a specific field
 >> id:*IPSL* AND variable:hus*

Current Selections

(x) [text:id:*IPSL* AND variable:hus*](#)

Search Categories

- Project
- Institute
- Model
- SubModel
- Instrument
- Experiment Family
- Experiment
- SubExperiment
- Time Frequency
- Product
- Realm
- Variable
- Variable Long Name

 id:*IPSL* AND variable:hus*

Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

[Temporal Search](#)
[Geospatial Search](#)
[Clear search constraints and datacart](#)
[Search Help](#)
[Search Controlled Vocabulary](#)

☒ Search All Sites
 ☐ Show All Replicas
 ☐ Show All

Versions

< 1 2 3 ... [112](#) [113](#) > displaying 1 to 10 of 1125 search results

Display datasets per page

[Add All Displayed to Datacart](#)
[Remove All Displayed from Datacart](#)

Results Data Cart

project=PMIP3, model=IPSL-CM5A-LR, Institut Pierre-Simon Laplace, experiment=1 percent per year CO2, time frequency=monClim, cmor table=Aclim, modeling realm=atmos, ensemble=r1i1p1, version=20120418

Data Node: [esgf-node.ipsl.fr](#)

Version: 20120418

Description: IPSL-CM5A-LR model output prepared for CMIP5 1 percent per year CO2

Further options: [Add To Cart](#) [Visualize and Analyze](#)

project=PMIP3, model=IPSL-CM5A-LR, Institut Pierre-Simon Laplace, experiment=last glacial maximum, time frequency=monClim, cmor table=Aclim, modeling realm=atmos, ensemble=r1i1p1, version=20120418

Data Node: [esgf-node.ipsl.fr](#)

Version: 20120418



FACETED SEARCH

- Results can be filtered giving values for the following CMIP5 facets in “Search categories”:
 - Project
 - Institute
 - Model
 - Submodel
 - Instrument
 - Experiment Family
 - Experiment
 - SubExperiment
 - Time Frequency
 - Product
 - Realm
 - Variable
 - Variable Long Name
 - CMIP Table
 - CF Standard Name
 - Ensemble
 - Data Node
- Selected facets values are displayed in “Current selections”

Current Selections

[remove all](#)
[\(x\) project:CMIP5](#)
[\(x\) experiment family:Historical](#)
[\(x\) time frequency:6hr](#)
[\(x\) realm:atmos](#)
[\(x\) ensemble:r1i1p1](#)

Search Categories

- Project
- Institute
- Model
- SubModel
- Instrument
- Experiment Family
- Experiment
- SubExperiment
- Time Frequency
- Product

Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
 To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

☒ Search All Sites
 ☐ Show All Replicas
 ☐ Show All

Versions

< 1 2 3 ... 8 9 > displaying 1 to 10 of 86 search results

Display datasets per page

[Add All Displayed to Datacart](#)
[Remove All Displayed from Datacart](#)

Results Data Cart

[project=CMIP5 / IPCC Fifth Assessment Report, model=HadGEM2-CC, Met Office Hadley Centre, experiment=historical, time frequency=6hr, modeling realm=atmos, ensemble=r1i1p1, version=20120116](#)

Data Node: cmip-dn1.badc.rl.ac.uk

Version: 20120116

Description: HadGEM2-CC model output prepared for CMIP5 historical

Further options: [Add To Cart](#) [Model Metadata](#)

[project=CMIP5 / IPCC Fifth Assessment Report, model=HadGEM2-ES, Met Office Hadley Centre, experiment=historical, time frequency=6hr, modeling realm=atmos, ensemble=r1i1p1, version=20110921](#)

Data Node: cmip-dn1.badc.rl.ac.uk

Version: 20110921

Description: HadGEM2-ES model output prepared for CMIP5 historical

[Temporal Search](#)
[Geospatial Search](#)
[Clear search constraints and datacart](#)
[Search Help](#)
[Search Controlled Vocabulary](#)

Results selection

- Clicking on a search result will show dataset metadata

Metadata		
		Show/Hide Properties <input type="text" value="cf_standard_name"/>
	Property	Value
expand	cf_standard_name	air_pressure_at_sea_level ; air_temperature ; eastward_wind ; northward_wind...
	cmor_table	6hrPlev
	data_node	cmip-dn1.badc.rl.ac.uk
expand	dataset_id_template_	cmip5.%(product)s.%(institute)s.%(model)s.%(experiment)s.%(time_frequency)s.%(realm)s.%(cmor_table)s.%(...
expand	description	HadGEM2-CC model output prepared for CMIP5 historical...
expand	drs_id	cmip5.output1.MOHC.HadGEM2-CC.historical.6hr.atmos.6hrPlev.r1i1p1...
	ensemble	r1i1p1
	experiment	historical
	experiment_family	All ; Historical
expand	forcing	GHG, Oz, SA, LU, SI, VI, BC, OC, (GHG = CO2, N2O, CH4, CFCs)...
	format	netCDF, CF-1.4

Results selection

- Datasets to be downloaded must be added to the **Data Cart** clicking on “Add to data cart”

Results

Data Cart

project=CMIP5 / IPCC Fifth Assessment Report, model=HadGEM2-CC, Met Office Hadley Centre,
experiment=historical, time frequency=6hr, modeling realm=atmos, ensemble=r1i1p1, version=20120116
Data Node: cmip-dn1.badc.rl.ac.uk
Version: 20120116
Description: HadGEM2-CC model output prepared for CMIP5 historical
Further options: [Remove From Cart](#) [Model Metadata](#)

project=CMIP5 / IPCC Fifth Assessment Report, model=HadGEM2-ES, Met Office Hadley Centre,
experiment=historical, time frequency=6hr, modeling realm=atmos, ensemble=r1i1p1, version=20110921
Data Node: cmip-dn1.badc.rl.ac.uk
Version: 20110921
Description: HadGEM2-ES model output prepared for CMIP5 historical
Further options: [Add To Cart](#) [Model Metadata](#)

INDEX

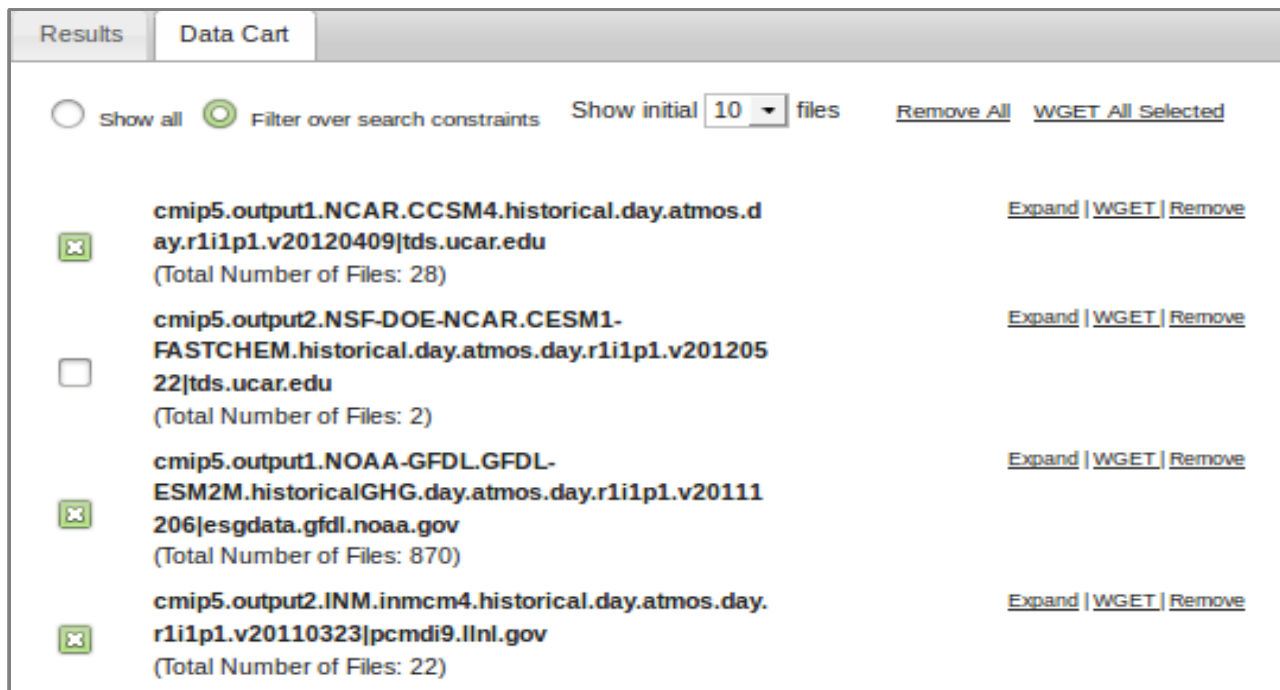
- ◉ Introduction
- ◉ Setting up an account
- ◉ Searching and accessing
- ◉ Download
 - ◉ Data Cart
 - ◉ Data access options
 - ◉ Running wget scripts
- ◉ Advanced features

Download

- To download data is available a "Data cart " similar to the e-commerce "shopping-cart"
 - Datasets may be access restricted
- Access data options
 - All datasets can be accessed by:
 - HTTP
 - Some datasets can be accesed by:
 - OPeNDAP
 - LAS
 - GridFTP
 - Tutorial: <https://www.youtube.com/watch?v=5hsRw4GELCE>

Data Cart

- The "Data cart" tab shows dataset selected
 - To view individual files associated with a dataset clicked in "expand" link




The screenshot displays the "Data Cart" tab in a web application. At the top, there are two tabs: "Results" and "Data Cart". Below the tabs, there are controls for filtering and displaying results. A radio button labeled "Show all" is unselected, while a radio button labeled "Filter over search constraints" is selected. To the right of these is a "Show initial" dropdown menu set to "10" and the word "files". Further right are two links: "Remove All" and "WGET All Selected".

The main content area lists four datasets, each with a checkbox on the left and a set of links (Expand, WGET, Remove) on the right. The datasets are:

- ☒ **cmip5.output1.NCAR.CCSM4.historical.day.atmos.d**
ay.r1i1p1.v20120409|tds.ucar.edu
(Total Number of Files: 28) [Expand](#) | [WGET](#) | [Remove](#)
- ☐ **cmip5.output2.NSF-DOE-NCAR.CESM1-**
FASTCHEM.historical.day.atmos.day.r1i1p1.v201205
22|tds.ucar.edu
(Total Number of Files: 2) [Expand](#) | [WGET](#) | [Remove](#)
- ☒ **cmip5.output1.NOAA-GFDL.GFDL-**
ESM2M.historicalGHG.day.atmos.day.r1i1p1.v20111
206|esgdata.gfdl.noaa.gov
(Total Number of Files: 870) [Expand](#) | [WGET](#) | [Remove](#)
- ☒ **cmip5.output2.INM.inmcm4.historical.day.atmos.day.**
r1i1p1.v20110323|pcmdi9.llnl.gov
(Total Number of Files: 22) [Expand](#) | [WGET](#) | [Remove](#)

Data Cart

- To hide individual files clicked in “Collapse”

 **cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409** [Collapse](#) [WGET](#) [Remove](#)
tds.ucar.edu
(Total Number of Files: 28)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409.pr_day_CCSM4_historical_r1i1p1_18500101-18841231.nc_1
tracking_id: 02321604-c9e8-4140-8a21-ea5f5240b5f2
checksum: 1275e3e5fde840700c712856342ad5b2 (MD5)

[HTTP](#)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409.pr_day_CCSM4_historical_r1i1p1_18850101-19191231.nc_1
tracking_id: 69711df8-dfc3-46b9-8def-bf6e2325fe99
checksum: e5515fde1ab322de13c9c41e4d83b31d (MD5)

[HTTP](#)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409.pr_day_CCSM4_historical_r1i1p1_19200101-19541231.nc_1
tracking_id: daa7cd1b-2cfb-4825-9e26-961e87286efb
checksum: 6f03a2b04b0b76f7c09c43d63d1e7c47 (MD5)


[HTTP](#)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409.pr_day_CCSM4_historical_r1i1p1_19550101-19891231.nc_1
tracking_id: 3233c6d0-d669-48af-8b77-7c62707d8e84

[HTTP](#)

Data Cart

- To hide individual files clicked in “Collapse”

 **cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409** [Collapse](#) [WGET](#) [Remove](#)
tds.ucar.edu
(Total Number of Files: 28)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409.pr_day_CCSM4_historical_r1i1p1_18500101-18841231.nc_1
tracking_id: 02321604-c9e8-4140-8a21-ea5f5240b5f2
checksum: 1275e3e5fde840700c712856342ad5b2 (MD5)

[HTTP](#)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409.pr_day_CCSM4_historical_r1i1p1_18850101-19191231.nc_1
tracking_id: 69711df8-dfc3-46b9-8def-bf6e2325fe99
checksum: e5515fde1ab322de13c9c41e4d83b31d (MD5)

[HTTP](#)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409.pr_day_CCSM4_historical_r1i1p1_19200101-19541231.nc_1
tracking_id: daa7cd1b-2cfb-4825-9e26-961e87286efb
checksum: 6f03a2b04b0b76f7c09c43d63d1e7c47 (MD5)

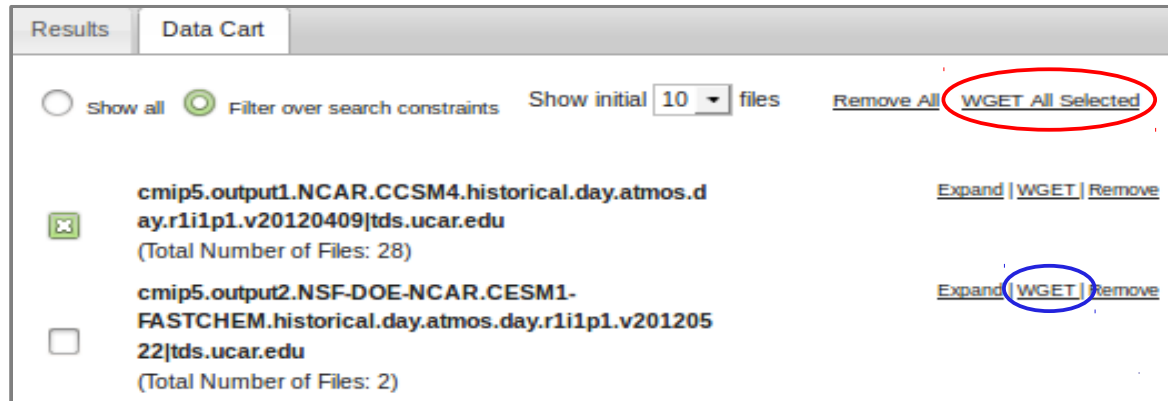
[HTTP](#)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v20120409.pr_day_CCSM4_historical_r1i1p1_19550101-19891231.nc_1
tracking_id: 3233c6d0-d669-48af-8b77-7c62707d8e84

[HTTP](#)

Data access options

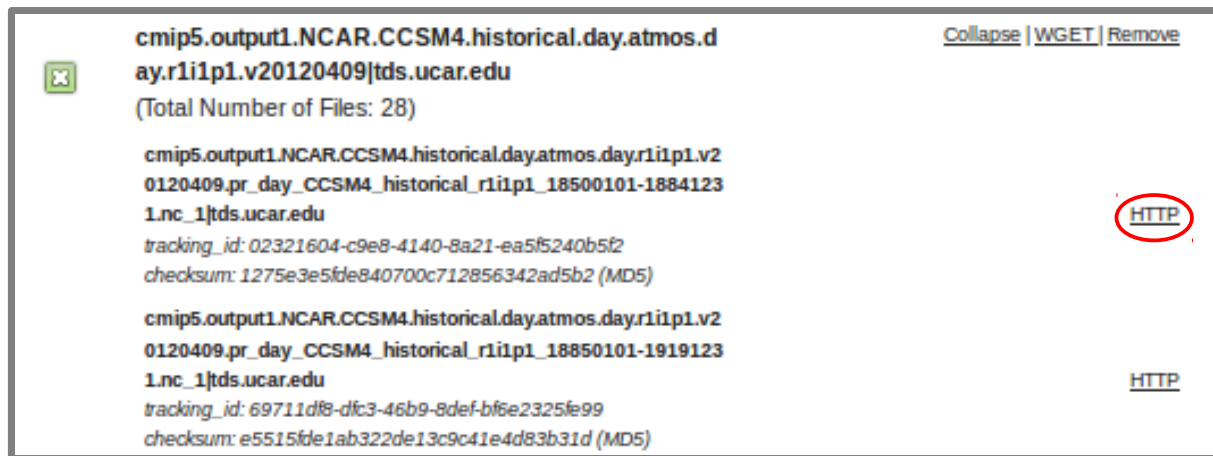
- Download dataset - Wget script (All datasets)
- to download dataset selected clicked in "WGET All Selected"



- to download one dataset clicked in "WGET"
- The script is downloaded to the end user's local machine

Data access options

- Download files - HTTP (All datasets)
 - to download individual files associated with a dataset clicked in HTTP after to expand dataset



The screenshot shows a file listing interface. At the top, there is a header bar with a green 'x' icon on the left and three links: 'Collapse', 'WGET', and 'Remove'. Below the header, the main content area displays two file entries. Each entry consists of a file path, a tracking ID, and a checksum. The first file entry is 'cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v2.0120409.pr_day_CCSM4_historical_r1i1p1_18500101-1884123.1.nc_1.tds.ucar.edu' with tracking ID '02321604-c9e8-4140-8a21-ea5f5240b5f2' and checksum '1275e3e5fde840700c712856342ad5b2 (MD5)'. The second file entry is 'cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v2.0120409.pr_day_CCSM4_historical_r1i1p1_18850101-1919123.1.nc_1.tds.ucar.edu' with tracking ID '69711df8-dfc3-46b9-8def-bf6e2325fe99' and checksum 'e5515fde1ab322de13c9c41e4d83b31d (MD5)'. To the right of each file entry, there is a red circle containing the text 'HTTP'.

cmip5.output1.NCAR.CCSM4.historical.day.atmos.d
ay.r1i1p1.v20120409|tds.ucar.edu
(Total Number of Files: 28)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v2
0120409.pr_day_CCSM4_historical_r1i1p1_18500101-1884123
1.nc_1|tds.ucar.edu
tracking_id: 02321604-c9e8-4140-8a21-ea5f5240b5f2
checksum: 1275e3e5fde840700c712856342ad5b2 (MD5)

cmip5.output1.NCAR.CCSM4.historical.day.atmos.day.r1i1p1.v2
0120409.pr_day_CCSM4_historical_r1i1p1_18850101-1919123
1.nc_1|tds.ucar.edu
tracking_id: 69711df8-dfc3-46b9-8def-bf6e2325fe99
checksum: e5515fde1ab322de13c9c41e4d83b31d (MD5)

- GridFTP download - Globus Online

Data access options

- OPeNDAP (Some datasets)
- Provides remote access to individual variables within datasets in a form usable by many tools
- After to expand dataset clicked in “OPeNDAP”



```
E2-R.rcp26.mon.atmos.Amon.r1i1p1.v20120516|esg
data1.nccs.nasa.gov
(Total Number of Files: 612)

cmip5.output1.NASA-GISS.GISS-
E2-R.rcp26.mon.atmos.Amon.r1i1p1.v20120516.ccb_Amon_GI
SS-
E2-R_rcp26_r1i1p1_200601-202512.nc|esgdata1.nccs.nasa.go
v
tracking_id: bf5769ef-7e48-476f-8526-949dbfa3afc8
checksum: d3a0e66bfa31bbc37a921fd53df33fb (MD5)

cmip5.output1.NASA-GISS.GISS-
E2-R.rcp26.mon.atmos.Amon.r1i1p1.v20120516.ccb_Amon_GI
SS-
E2-R_rcp26_r1i1p1_202601-205012.nc|esgdata1.nccs.nasa.go
v
tracking_id: 78f6466c-8058-4387-81c5-b5f2fd097e28
checksum: b56452aeee01e5227c99229c58724b58 (MD5)
```

HTTP OPeNDAP

HTTP OPeNDAP

Current Selections

[remove all](#)
[\(x\) project:CMIP5](#)
[\(x\) institute:NASA-GISS](#)
[\(x\) realm:atmos](#)

Search Categories

[Project](#)
[Institute](#)
[Model](#)
[SubModel](#)
[Instrument](#)
[Experiment Family](#)
[Experiment](#)
[SubExperiment](#)
[Time Frequency](#)
[Product](#)
[Realm](#)
[Variable](#)

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
 To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

☒ Search All Sites ☐ Show All Replicas ☐ Show All

Versions

< [1](#) [2](#) [3](#) ... [33](#) [34](#) > displaying 1 to 10 of 332 search results

Display datasets per page

[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

[Results](#) [Data Cart](#)

project=CMIP5, model=NASA Goddard Institute for Space Studies, GISS-E2-H-CC, experiment=historical, time frequency=mon, modeling realm=atmos, ensemble=r1i1p1, version=20120723

Data Node: [esgdata1.nccs.nasa.gov](#)

Version: 20120723

Description: GISS-E2-H-CC model output prepared for CMIP5 historical

Further options: [Add To Cart](#) [Visualize and Analyze](#) [Model Metadata](#)

project=CMIP5, model=NASA Goddard Institute for Space Studies, GISS-E2-H-CC, experiment=pre-industrial control, time frequency=mon, modeling realm=atmos, ensemble=r1i1p1, version=20120828

Data Node: [esgdata1.nccs.nasa.gov](#)

Version: 20120828

Description: GISS-E2-H-CC model output prepared for CMIP5 pre-industrial control

Further options: [Remove From Cart](#) [Visualize and Analyze](#) [Model Metadata](#)

[Temporal Search](#)
[Geospatial Search](#)
[Clear search constraints and datacart](#)
[Search Help](#)
[Search Controlled Vocabulary](#)

ResultsData Cart

☐ Show all ☒ Filter over search constraints Show initial 10 files [Remove All](#) [WGET All Selected](#)

**cmip5.output1.NASA-GISS.GISS-
E2-H-CC.historical.mon.atmos.Amon.r1i1p1.v201207
23|esgdata1.nccs.nasa.gov**
(Total Number of Files: 156)

[Expand](#) [WGET](#) [Remove](#)

ResultsData Cart

☐ Show all ☒ Filter over search constraints Show initial 10 files [Remove All](#) [WGET All Selected](#)

**cmip5.output1.NASA-GISS.GISS-
E2-H-CC.historical.mon.atmos.Amon.r1i1p1.v201207
23|esgdata1.nccs.nasa.gov**
(Total Number of Files: 156)

[Collapse](#) [WGET](#) [Remove](#)

**cmip5.output1.NASA-GISS.GISS-
E2-H-CC.historical.mon.atmos.Amon.r1i1p1.v20120723.ccb_A
mon_GISS-
E2-H-CC_historical_r1i1p1_185001-190012.nc|esgdata1.nccs.n
asa.gov**
tracking_id: e001616a-30b4-4c27-9be2-d66240f2b52b
checksum: 060b6649a4f2fd24f7a858736d61d3f2 (MD5)

[HTTP OPENDAP](#)

**cmip5.output1.NASA-GISS.GISS-
E2-H-CC.historical.mon.atmos.Amon.r1i1p1.v20120723.ccb_A
mon_GISS-
E2-H-CC_historical_r1i1p1_190101-195012.nc|esgdata1.nccs.n
asa.gov**
tracking_id: 8a7f125b-353f-4b19-a7ea-72798bab18af

[HTTP OPENDAP](#)

OPeNDAP Dataset Access Form

Netscape 4.61 and Internet Explorer 5.00.

Action:

Data URL:

Global

Attributes:

```
institution: NASA/GISS (Goddard Institute for Space Studies) New
York, NY
institute_id: NASA-GISS
experiment_id: rcp26
source: GISS-E2-R-E135RCP30aF40oQ32 Atmosphere: GISS-E2; Ocean: R
model_id: GISS-E2-R
```

Variables:

☐ **time: Array of 64 bit Reals [time = 0..239]**

time:

```
bounds: time_bnds
units: days since 2006-1-1
calendar: 365_day
axis: T
long_name: time
standard_name: time
```

☐ **time_bnds: Array of 64 bit Reals [time = 0..239][bnds = 0..1]**

time: bnds:

☐ **lat: Array of 64 bit Reals [lat = 0..89]**

lat:

```
bounds: lat_bnds
units: degrees_north
axis: Y
long_name: latitude
standard_name: latitude
```

☐ **lat_bnds: Array of 64 bit Reals [lat = 0..89][bnds = 0..1]**

lat: bnds:

☐ **lon: Array of 64 bit Reals [lon = 0..143]**

lon:

```
bounds: lon_bnds
units: degrees_east
axis: X
long_name: longitude
standard_name: longitude
```




Data access options

- LAS - Data Analysis Integration
 - Highly configurable web server designed to provide flexible access to geo-referenced scientific data.
 - Ferret is the default visualization application used by LAS
 - LAS enables the web user to
 - Visualize data with on-the-fly graphics
Request custom subsets of variables in a choice of file formats
 - Access background reference material about the data (metadata)
 - Compare (difference) variables from distributed locations

Current Selections

[\(x\) institute:COLA-CFS](#)

Search Categories

Project

Institute

Model

SubModel

Instrument

Experiment Family

Experiment

SubExperiment

Time Frequency

Product

Realm

Variable

Variable Long Name

Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

[Temporal Search](#)
[Geospatial Search](#)
[Clear search constraints and datacart](#)
[Search Help](#)
[Search Controlled Vocabulary](#)

☒ Search All Sites ☐ Show All Replicas ☐ Show All

Versions

< 1 2 3 ... 30 31 > displaying 1 to 10 of 305 search results

Display datasets per page

[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

Results

Data Cart

project=CMIP5, model=National Centers for Environmental Prediction, CFSv2-2011, experiment=10- or 30-year run initialized in year 1960, time frequency=mon, modeling realm=atmos, ensemble=r1i1p1, version=20120511

Data Node: esgdata1.nccs.nasa.gov


Version: 20120511


Description: CFSv2-2011 model output prepared for CMIP5 10- or 30-year run initialized in year 1960



Further options: [Add To Cart](#) [Visualize and Analyze](#) [Model Metadata](#)

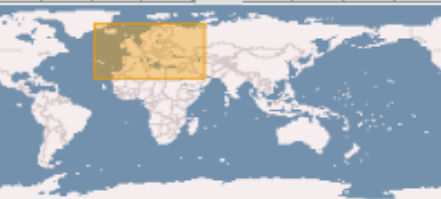
project=CMIP5, model=National Centers for Environmental Prediction, CFSv2-2011, experiment=10- or 30-year run initialized in year 1960, time frequency=mon, modeling realm=atmos, ensemble=r1i1p2, version=20120511

ESGF LAS

Choose dataset Update Plot  Set plot options Animate Compare Google Earth Show Values Export to Desktop Application Save As ... Link To ... Print

 project=CMIP5, model=National Centers for Environmental Prediction, CFSv2-2011, experiment=10- or 30-year run initialized in year 1960, time_frequency=mon, modeling realm=atmos, ensemble=r1i1p1, version=20120511

+ hus  



73.13 N

32.64 W



54.55 E

23.91 N

MAPS

☒ Latitude-Longitude

DEPTH PROFILES

☐ Longitude-Depth

☐ Latitude-Depth

HOVMOLLER PLOTS

☐ Longitude-Time

☐ Latitude-Time

☐ Depth-Time

LINE PLOTS

☐ Time Series

☐ Longitude

☐ Latitude

☐ Depth

SCATTER PLOTS

☐ Property-Property

Date:

Jul 1974

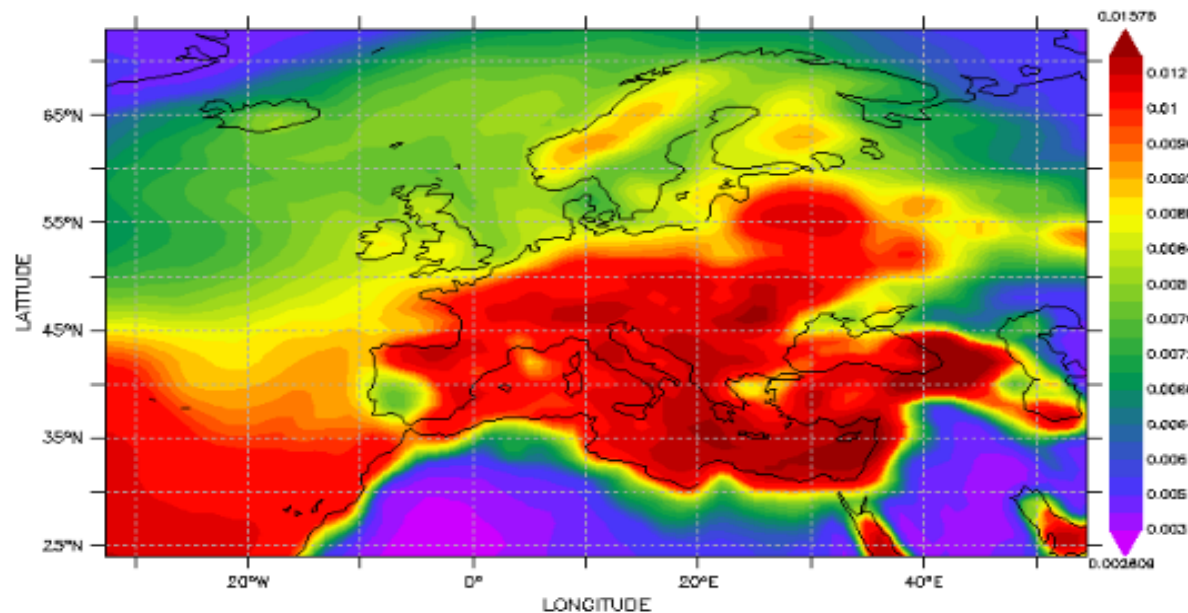
Depth (Pa): 100000

[Apply analysis](#)

LAS 7./Ferret 6.72 NOAA/PMEL

Z (Pa) : 100000
TIME : 16-JUL-1974 12:00

DATA SET: project=CMIP5, model=National Centers for Environmental Prediction, CFSv2-2011, experiment=10- or 30-year run initialized in year 1960, time_frequency=mon, modeling realm=atmos, ensemble=r1i1p1, version=20120511



hus (1)



Downloading with wget scripts

- Pre-requisites
 - WGET (version 1.12 or later)
 - compiled with the OpenSSL libraries (normally it is) and java 1.5+, and available in their execution PATH
 - Be member of a group authorized to access the data that will be downloaded



Running wget scripts

- Files that are created and managed
 - `$HOME/.MyProxyLogon`

Shared with the MyProxy gui stores MyProxy server access data
 - `$HOME/.esg/getcert.jar`

Java program to retrieve credentials from a given OpenID and without a gui
 - `$HOME/.esg/certificates`

Retrieved federation CAs
 - `$HOME/.esg/credentials.pem`

Retrieved credential



Running wget scripts

- On a Linux System:
 - Place the script in the folder where you want to download dataset
 - Give execute permissions to the script
`chmod +x wget_script.sh`
 - Run the script
`./wget_script.sh`
 - If the credentials are invalid
 - Authentication will be requested by the terminal


```
terryk@Pixsy:~/Documentos/pruebas$ ./wget-20121023042746.sh
```

```
Running wget-20121023042746.sh version: 1.2.2
```

```
Use wget-20121023042746.sh -h for help.
```

```
Script created for 953 file(s)
```

```
(The count won't match if you manually edit this file!)
```

```
OpenSSL 1.0.1 14 Mar 2012
```

```
** WARNING: ESGF Host certificate checking might not be compatible with OpenSSL 1.0+
```

```
The certificate expires in less than 8 hour(s). Renewing...
```

```
Please give your OpenID (hit ENTER to accept default) [https://myserver/example/username]? https://pcmdi9.llnl.gov/esgf-idp/openid/terryk  
MyProxy Password?
```

```
Retrieving Credentials...done!
```

```
hus_6hrLev_HadGEM2-ES_historical_r1i1p1_195109010600-195112010000.nc ...Already downloaded and verified
```

```
hus_6hrLev_HadGEM2-ES_historical_r1i1p1_195206010600-195209010000.nc ...Already downloaded and verified
```

```
hus_6hrLev_HadGEM2-ES_historical_r1i1p1_195409010600-195412010000.nc ...Already downloaded and verified
```

```
hus_6hrLev_HadGEM2-ES_historical_r1i1p1_195809010600-195812010000.nc ...Downloading
```

```
--2012-11-12 13:29:21-- http://cmip-dn1.badc.rl.ac.uk/thredds/fileServer/esg_dataroot/cmip5/output1/MOHC/HadGEM2-ES/historical/6hr/atmos/6  
orical_r1i1p1_195809010600-195812010000.nc
```

```
Resolviendo cmip-dn1.badc.rl.ac.uk (cmip-dn1.badc.rl.ac.uk)... 130.246.191.71
```

```
Conectando con cmip-dn1.badc.rl.ac.uk (cmip-dn1.badc.rl.ac.uk)[130.246.191.71]:80... conectado.
```

```
Petición HTTP enviada, esperando respuesta... 302 Moved Temporarily
```

```
Ubicación: https://cmip-dn1.badc.rl.ac.uk/esg-orp/home.htm?redirect=http%3A%2F%2Fcmip-dn1.badc.rl.ac.uk%2Fthredds%2FfileServer%2Fesg_dataro  
%2F6hr%2Fatmos%2F6hrLev%2Fr1i1p1%2Fv20101208%2Fhus%2Fhus_6hrLev_HadGEM2-ES_historical_r1i1p1_195809010600-195812010000.nc [siguiente]
```

```
--2012-11-12 13:29:22-- https://cmip-dn1.badc.rl.ac.uk/esg-orp/home.htm?redirect=http%3A%2F%2Fcmip-dn1.badc.rl.ac.uk%2Fthredds%2FfileServe  
S%2Fhistorical%2F6hr%2Fatmos%2F6hrLev%2Fr1i1p1%2Fv20101208%2Fhus%2Fhus_6hrLev_HadGEM2-ES_historical_r1i1p1_195809010600-195812010000.nc
```

```
Conectando con cmip-dn1.badc.rl.ac.uk (cmip-dn1.badc.rl.ac.uk)[130.246.191.71]:443... conectado.
```

```
Petición HTTP enviada, esperando respuesta... 302 Moved Temporarily
```

```
Ubicación: http://cmip-dn1.badc.rl.ac.uk/thredds/fileServer/esg_dataroot/cmip5/output1/MOHC/HadGEM2-ES/historical/6hr/atmos/6hrLev/r1i1p1/v  
195809010600-195812010000.nc [siguiente]
```

```
--2012-11-12 13:29:23-- http://cmip-dn1.badc.rl.ac.uk/thredds/fileServer/esg_dataroot/cmip5/output1/MOHC/HadGEM2-ES/historical/6hr/atmos/6  
orical_r1i1p1_195809010600-195812010000.nc
```

```
Conectando con cmip-dn1.badc.rl.ac.uk (cmip-dn1.badc.rl.ac.uk)[130.246.191.71]:80... conectado.
```

```
Petición HTTP enviada, esperando respuesta... 206 Partial Content
```

```
Longitud: 1523538424 (1,4G), quedan 940448248 (897M) [application/x-netcdf]
```

```
Grabando a: "hus_6hrLev_HadGEM2-ES_historical_r1i1p1_195809010600-195812010000.nc"
```

```
38% [+++++>
```




Running wget scripts

- Can be interrupted
 - The download will restart at the point where it stopped in next running
- When download is completed
 - The script will verify the **checksum** of each file to avoid errors
- To find the start date and the end date of the validity of credential

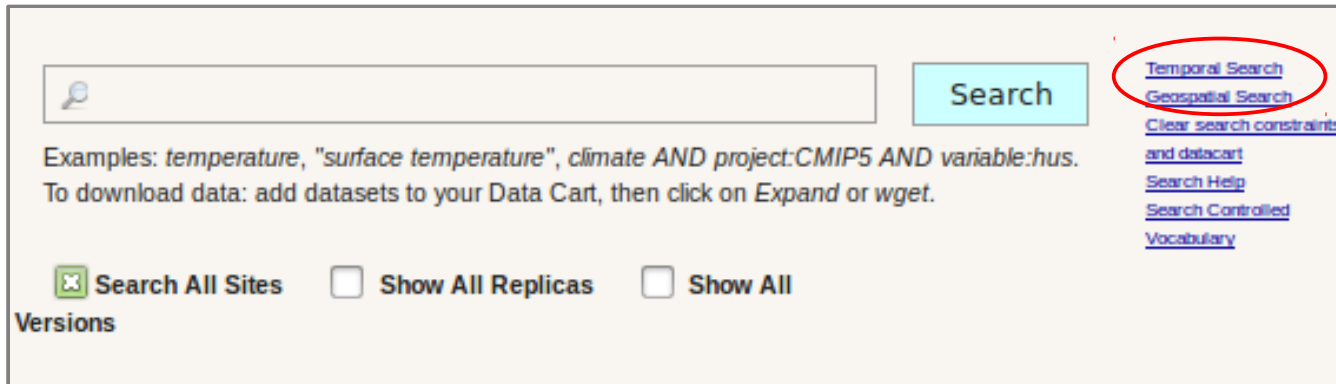
```
>> openssl x509 -in $HOME/.esg/credentials.pem -dates
```




Index

- ◉ Introduction
- ◉ Setting up an account
- ◉ Searching and accesing
- ◉ Download
- ◉ Advanced features
 - ◉ Advanced search in browser
 - ◉ RESTful API
 - ◉ Wget scripting
 - ◉ RSS feed

Advanced search in browser

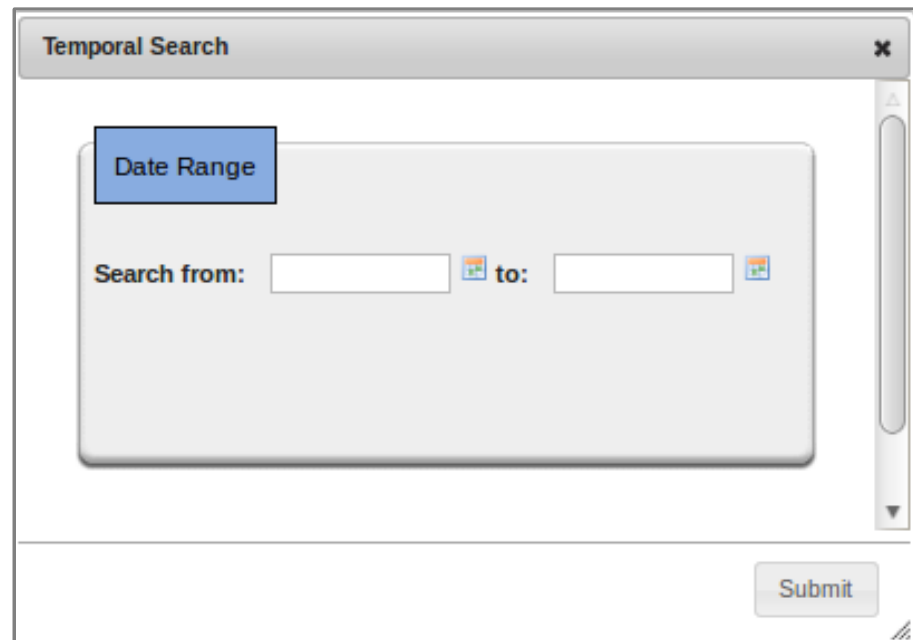


Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

☒ Search All Sites ☐ Show All Replicas ☐ Show All Versions

[Temporal Search](#)
[Geospatial Search](#)
[Clear search constraints and datacart](#)
[Search Help](#)
[Search Controlled Vocabulary](#)

Temporal search



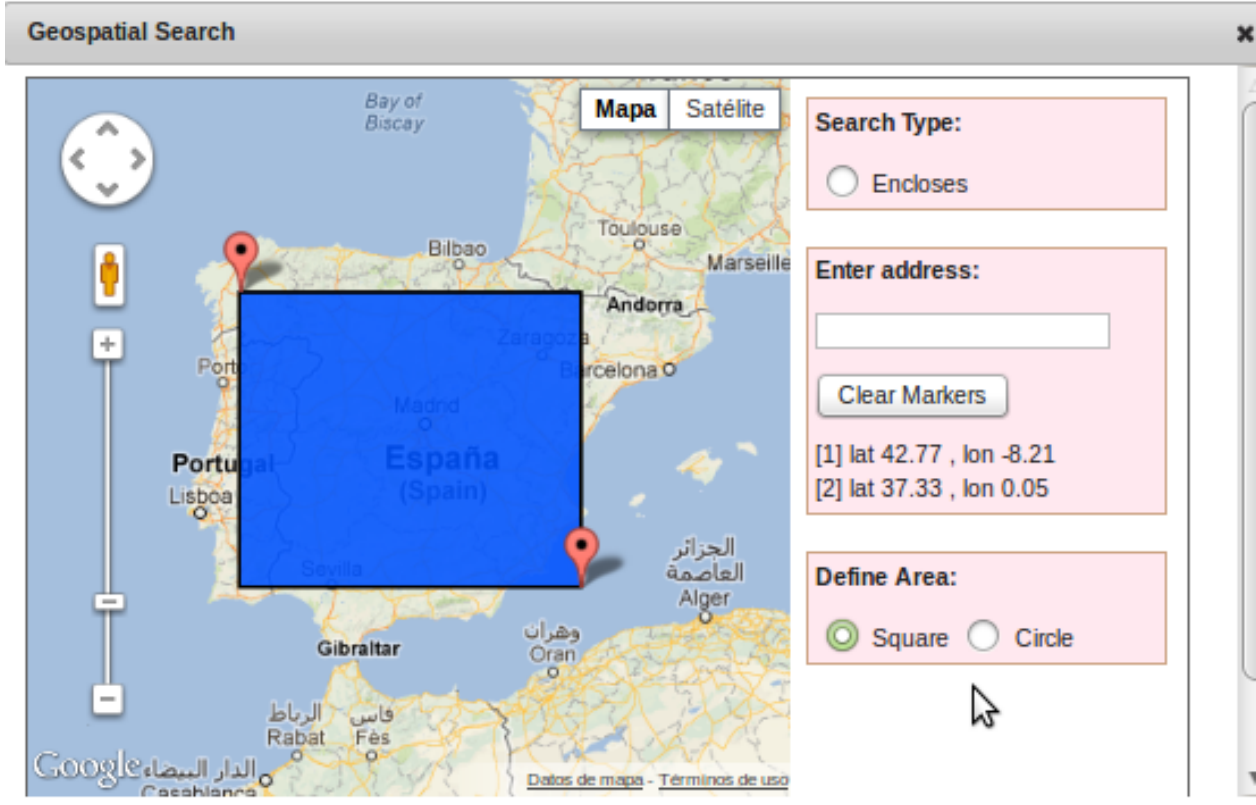
Temporal Search [X]

Date Range

Search from: [] to: []

Submit

Geospatial search



The screenshot shows a web application titled "Geospatial Search". On the left is a map of the Iberian Peninsula. A blue square is drawn over central Spain, with red pins at its corners. The map includes labels for "España (Spain)", "Portugal", "Lisboa", "Madrid", "Barcelona", "Alger", and "Oran". Navigation controls like a compass and zoom in/out buttons are on the left. On the right is a control panel with three sections: "Search Type:" with a radio button for "Encloses"; "Enter address:" with a text input field and a "Clear Markers" button, followed by two lines of coordinates: "[1] lat 42.77 , lon -8.21" and "[2] lat 37.33 , lon 0.05"; and "Define Area:" with radio buttons for "Square" (selected) and "Circle". A mouse cursor is pointing at the "Square" option. At the bottom right is a "Submit" button.

- This geospatial or temporal constraints are added to “Current Selections” section



RESTful API

- ESGF Search Services API is meant to be used by the following actors:
 - Humans using a web browser interface, interacting with the interface in real time
 - Humans using a rich desktop client, or a command line utility, interacting with the client in real time
 - Batch Jobs executing data discovery at regular intervals

RESTful API

- Syntax
 - Where <base_search_URL> is the base URL of the search service
 - All parameters (keyword and facet) are optional
 - The value of all parameters must be **URL-encoded**

http://<base_search_URL>/search?

[keyword parameters as (name, value) pairs] [facet
parameters as (name,value) pairs]

- More explicity

http://<base search url>/? **Keywords parameters**

```
[query=...]  
[offset=...]  
[limit=...]  
[type=...]  
[format=...]  
[facets=...]  
[fields=...]  
[lat,lon,radius,polygon,location=...]  
[start,end=...]  
[from,to=...]  
[facet1=value1][facet2=value2][...]
```

CMIP5 Facet parameters

RESTful API

- Default Query
 - If no parameters at all are specified, the search service will execute a query using all the default values, specifically:

query=* (query all records)

distrib=true (execute a distributed search)

type=Dataset (return results of type "Dataset")

- Example:

<http://esg-datanode.jpl.nasa.gov/esg-search/search>

Base search URL

RESTful API

Determine all the allowed facet names and values at a specific site: ***esg-datanode.jpl.nasa.gov***

http://esg-datanode.jpl.nasa.gov/esg-search/search?facets=*&limit=0&distrib=false

search?facets=*&limit=0&distrib=false



search all facet values and their count

By default, a query to the search service will return the first 10 records matching the given constraints.

Limit and offset changes this default value

In this case limit=0 -> shows all results

Execute local query

RESTful API

Free Text and Facet Query

http://esg-datanode.jpl.nasa.gov/esg-search/search?query=airs%20AND%20humidity&cf_standard_name=air_temperature&project=obs4MIPs

search?

query=airs%20AND%20humidity

&cf_standard_name=air_temperature

&project=obs4MIPs

Project= obs4MIPs

CF Standard Name = air temperature

Free text search datasets that contains "air" and "humidity"

In Free Text upper case "OR" is the same as using simply a blank space



RESTful API

- Output Format

- The keyword **output** can be used to request results in a specific output format
 - Solr/XML (default)
 - Solr/JSON.

- Example:

Default query in JSON format

<http://esg-datanode.jpl.nasa.gov/esg-search/search?format=application%2Fsolr%2Bjson>
[?format=application%2Fsolr%2Bjson](http://esg-datanode.jpl.nasa.gov/esg-search/search?format=application%2Fsolr%2Bjson)

For more: http://www.esgf.org/wiki/ESGF_Search_REST_API



WGET SCRIPTING

- The same RESTful API that is used to query the ESGF search services can also be used to generate a wget script to download all files matching the given constraints.
- `http://<base_search_URL>/wget?`
[keyword parameters as (name, value) pairs]
[facet parameters as (name,value) pairs]
- where <base_search_URL> is the base URL of the search service at a given Index Node. corresponding shell scripts for bulk download of files.

Wget Scripting

- Syntax differences
 - Keyword **type is not allowed**, as the wget URL always assumes type=File.
 - Keyword **format is not allowed**, as the wget URL always returns a shell script as response document.
 - Keyword limit is assigned a **default value of limit=1000** (and must still be limit < 10,000).

A typical workflow pattern consists in first identifying all datasets or files matching some scientific criteria, then **changing the request URL from "/search?" to "/wget?"** to generate the corresponding shell scripts for bulk download of files.

Wget Scripting

- Example:

Download all observational files with variable hus

<http://esg-datanode.jpl.nasa.gov/esg-search/wget/?variable=hus&project=obs4MIPs&distrib=false>

`/wget/?variable=hus&project=obs4MIPs &distrib=false`



Search variable hus

Obs4MIPS project

Search on local node

For more: http://www.esgf.org/wiki/ESGF_Search_REST_API



RSS feed

- ESGF provides extremely powerful notification services based on the popular RSS technology
 - By subscribing to an ESGF RSS feed, a user can keep track of newly published datasets, or of new versions of already published datasets
 - RSS feeds are simply web sites (a.k.a. URLs) that deliver their content in XML format
 - Many applications exist that are capable of displaying the content of an RSS feed in a human-readable fashion

For more: http://www.esgf.org/wiki/ESGF_Search_REST_API

RSS feed

- This feed includes all most recent datasets in the entire ESGF federation.

<http://<esgf.host.name>/esg-search/feed/nodes.rss>

<http://esg-datanode.jpl.nasa.gov/esg-search/feed/nodes.rss>

- This feed is node specific, and includes all the most recent datasets from that node only

<http://<esgf.host.name>/esg-search/feed/node.rss>

<http://esg-datanode.jpl.nasa.gov/esg-search/feed/node.rss>

RSS feed

- This syntax feed includes all the most recent datasets that match a given search category (facet), name and value, across the whole federation

```
http://<esgf.host.name>/esgsearch/feed/<facet_name>/  
<facet_value>.rss
```

Includes all most recent datasets that contain Air Temperature, across all sites

http://esg-datanode.jpl.nasa.gov/esg-search/feed/cf_standard_name/air_temperature.rss

Includes all most recent observational datasets for CMIP5, across all sites

<http://esg-datanode.jpl.nasa.gov/esg-search/feed/project/CMIP5.rss>

For more: http://www.esgf.org/wiki/ESGF_Search_REST_API