Using the ESGF portal to access…

…PMIP3-CMIP5-<your xMIPx> data

Theory ⇔ and practice!

This presentation is not… (1/2)

…a scientific introduction to PMIP3 and CMIP5

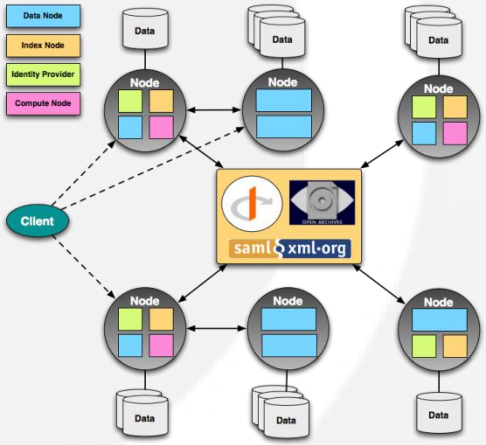
- Paleoclimate Modelling Intercomparison Project, Phase 3
  https://pmip3.lsce.ipsl.fr/

- Coupled Model Intercomparison Project (⇔ IPCC model DB)
  http://cmip-pcmdi.llnl.gov/cmip5/

This presentation is not… (2/2)

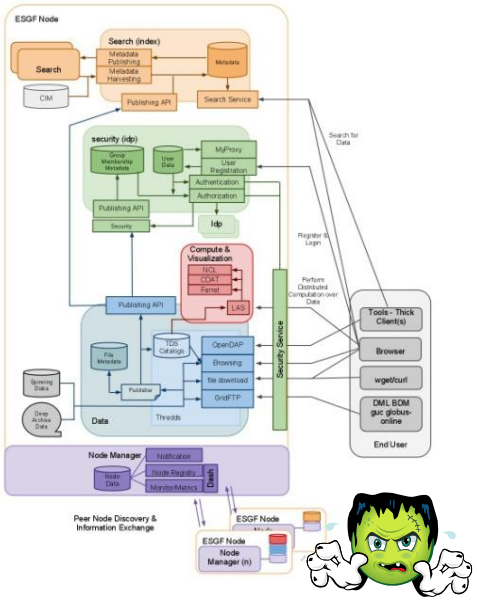
.. very technical

- ESGF = Earth System Grid Federation
  - *"is a spontaneous collaboration of groups, agencies and institutions around the world, that are dedicated to the development and operation of a long-term system for the management, access and analysis of climate data"*
  - ANL, LBNL, LLNL, NASA, NCAR, ORNL, BADC, CMCC, DKRZ, IPSL, …
- http://esgf.org/wiki/ESGF_Overview



What you need to know ☺

ESGF Distributed DB:
…is a collection of
Data nodes
Access portals



Forget he ugly technical details

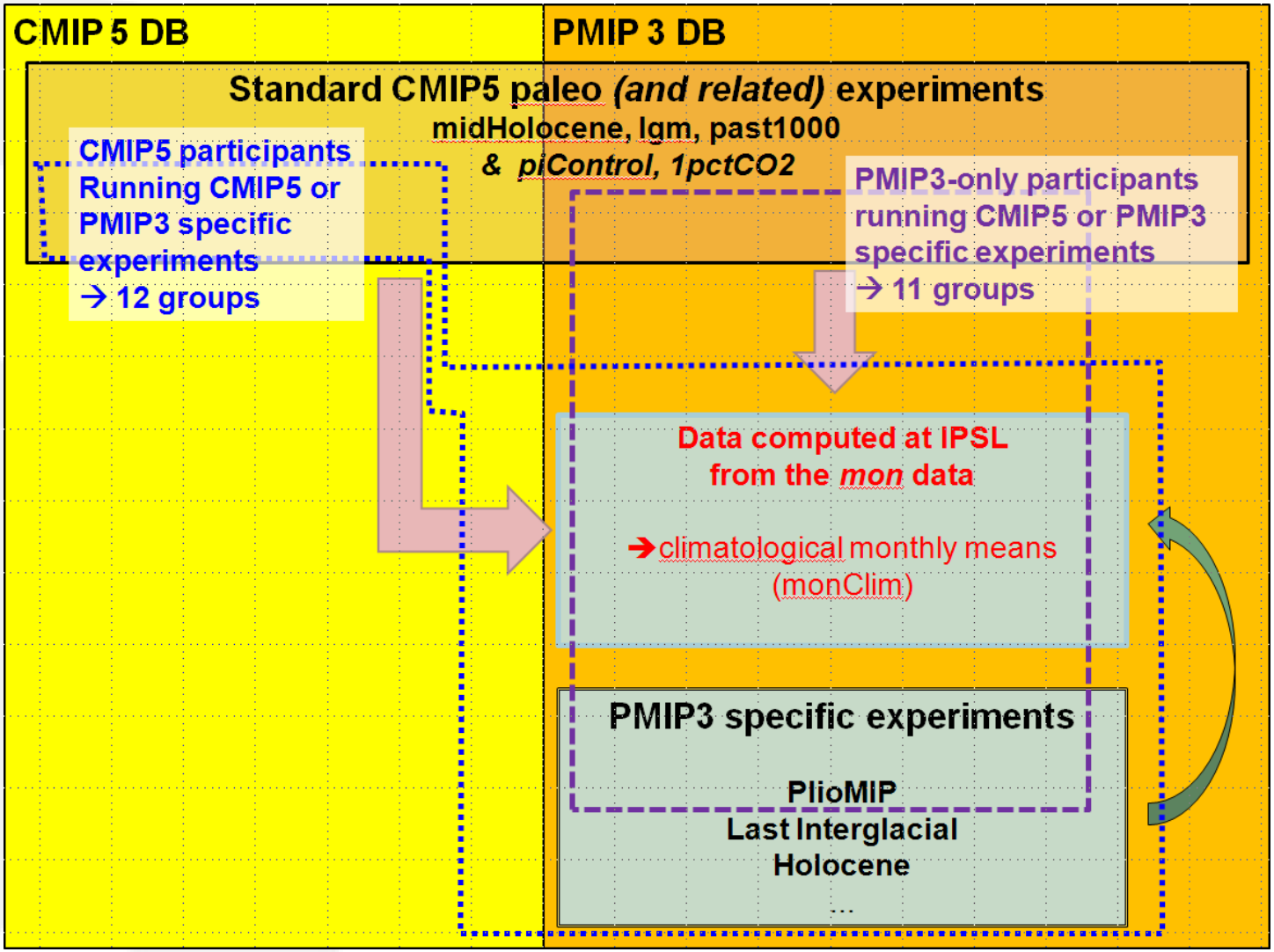
Why use ESGF to distribute PMIP3 data?

- The ESG infrastructure was initially developed to distribute CMIP5 and we adapted it for PMIP3
  - History: PMIPn and CMIPn have always been very close
    - Common variables, file formats & conventions, …
  - Now: PMIP3 and CMIP5 have a lot of common data
  - PMIP3 has benefited A LOT from the PRODIGUER project
    - Distributing IPSL data for CMIP5 (project initiated by S. Denvil)
      - Data servers, ESGF software configuration and updates, manpower

- ESGF is constantly evolving (for the better):
  - Old Gateways portal: slow, frustrating
    - ➔ New Peer-to-peer Front-End (P2P FE): fast, easy
  - More developments for handling non-CMIP5 data

- We still depend on people who know how to make ESGF work reliably at IPSL! And can maintain it…

The PMIP3 and CMIP5 databases overlap



The PMIP3 groups

List of PMIP3/CMIP5 participants

Information as of April 22nd 2013

Up to date info: https://pmip3.lsce.ipsl.fr/wiki/doku.php/pmip3:database:status

Some P2P nodes: PCMDI IPSL DKRZ BADC

CMIP5 Errata: http://pcmdi-cmip.llnl.gov/cmip5/errata/cmip5errata.html

Models' documentation: http://q.cmip5.ceda.ac.uk/cmip5/publisheddocs/

	Institute	Country	0k *piControl*	6k *midHolocene*	21k *lgm*	LM *past1000 (1000 years)*	1 % CO_2 *1pctCO2 (140 years)*	CMIP5	PlioMIP	Last interglacial	Holocene	Carbon cycle	Atm	Ocn	Model i	Term of Use	Data Node	Publish to (index node)	Extra Doc Errata etc...
1	AWI	Germany	Completed	Completed	Completed			No	Yes	Yes	No	Yes	96x48 x L19	120×101 x L40	COSMOS-a		DKRZ ?	ESG-WDCC ?	
2	BCC	China	CMIP5 (500)	CMIP5 (100)		CMIP5	CMIP5	Yes	No	No	No	Yes	128×64 x L26	360x232 x L40	bcc-csm1-1	Unrestricted	BCC	ESG-PCMDI	
3	BCCR	Norway	Completed	Running Summer 2013	Completed	Running Summer 2013	Running Summer 2013	No	Yes	Yes	No	Yes	96×48 x L26	100×116 x L32	NorESM1-L	Unrestricted	DKRZ ?	ESG-WDCC ?	
4	CAU-GEOMAR	Germany	*PMIP3*	*PMIP3*	Starting		*PMIP3*	No	Yes	Yes	?	No	96x48 x L19	182x149 x L31	KCM1-2-1	Non-commercial	DKRZ	ESG-WDCC ?	
5	CNRM-CERFACS	France	CMIP5 (850)	CMIP5 (200)	CMIP5 (200)		CMIP5	Yes	No	No	No	No	256x128 x L31	362x292 x L42	CNRM-CM5	Unrestricted	CNRM	ESG-IPSL	Doc
6	FUB	Germany	PMIP3 (400)		PMIP3 (600)			No	No	No	No	Yes	96×48 x L19	120×101 x L40	COSMOS-ASO	Unrestricted	IPSL (DKRZ later?)	ESG-BADC	
7	NOAA-GFDL	USA	CMIP5 (470)		Starting avail. Fall 2013		CMIP5	Yes	No	No	No	Yes	144x90 x L24	360x200 x L50 360x210 x L64	GFDL-ESM2M GFDL-ESM2G	Unrestricted	GFDL	ESG-PCMDI	
8	NASA-GISS	USA	CMIP5 (7 x)	CMIP5 (100)	CMIP5 (2 x 100)	CMIP5 (8 x 1000)	CMIP5 (151)	Yes	Yes	?	Yes	No	144x90 x L40	288x180 x L32	GISS-E2-R	Unrestricted	NCCS	ESG-PCMDI	Doc
9	IPSL	France	CMIP5 (1000)	CMIP5 (500)	CMIP5 (200)	CMIP5	CMIP5	Yes	Yes	Yes	Yes	Yes	96x95 x L39	182x149 x L31	IPSL-CM5A-LR	Unrestricted	IPSL	ESG-BADC	Doc
10	ICHEC (KNMI)	Netherlands	PMIP3 (40)	PMIP3 (40)				No	Yes	No	No	No	320×160 x L62	362x292 x L42	EC-Earth-2-2	Unrestricted	ICHEC	ESG-BADC	
11a	LASG-CESS	China	CMIP5 (900)	CMIP5 (100)	CMIP5 (100)		CMIP5	Yes	Yes	No	No	No	128x60 x L26	360x180 x L30	FGOALS-g2	Unrestricted	LASG	ESG-PCMDI	
11b	LASG-IAP	China	CMIP5 (501)	CMIP5 (100)		CMIP5	CMIP5	Yes	No	No	No	No	128x108 x L26	360x180 x L30	FGOALS-g2	Unrestricted	LASG	ESG-PCMDI	
11c	LASG-IAP	China	Completed			CMIP5		Yes	No	No	No	No	72x45 x L26	360x180 x L30	FGOALS-gl	Unrestricted	LASG	ESG-PCMDI	Doc
12	LOVECLIM	Belgium France Netherlands	Completed	Completed	Completed	Completed		No	No	Yes	Yes	No	32x64 x L3	122x65 x L20	LOVECLIM1-2	Unrestricted	IPSL	ESG-BADC	
13	MIROC	Japan	CMIP5 (531)	CMIP5 (100)	CMIP5 (100)	CMIP5	CMIP5	Yes	Yes	?	?	Yes	128x64 x L80	256x192 x L44	MIROC-ESM	Non-commercial only	DIAS	ESG-PCMDI	
14	MPI-M	Germany	CMIP5 (1156)	CMIP5 (2x100)	CMIP5 (2 x 100)	CMIP5	CMIP5	Yes	No	?	?	No	196x98 x L47	256x220 x L40	MPI-ESM-P	Unrestricted	DKRZ	ESG-WDCC	
15	MRI	Japan	CMIP5 (500)	CMIP5 (100)	CMIP5 (100)	Running September 2013	CMIP5	Yes	Yes	No	No	No	320x160 x L48	364x368 x L51	MRI-CGCM3	Non-commercial	DIAS	ESG-PCMDI	
16	NCAR	USA	CMIP5 (501)	CMIP5 (1x301 + 1x30)	CMIP5 (1x101 + 1x31)	CMIP5	CMIP5	Yes	Yes	No	No	No	288x192 x L26	320x384 x L60	CCSM4	Unrestricted	NCAR	ESG-NCAR	
17	OSUVic	USA	Completed	Completed (400)	Running May 2013		Starting April 2013	No	No	No	No	No	128x64 x L10	100 x 100 x L19	OSUVic-2.9	Unrestricted	?	?	
18	CSIRO-QCCCE	Australia	CMIP5 (500)	CMIP5 (100)			CMIP5	Yes	No	No	No	No	192x96 x L18	192x192 x L31	CSIRO-Mk3.6-0	Non-commercial	NCI	ESG-NCI	
19a	MOHC (UK groups)	UK	CMIP5 (497)	CMIP5 (102)	Starting	Running Spring 2013	CMIP5	Yes	Yes	Yes	Yes	Yes	192x144 x L38	360x216 x L40	HadGEM2-ES	Unrestricted	BADC	ESG-BADC	Doc
19b	MOHC (UK groups)	UK	CMIP5 (240)	CMIP5 (35)				Yes	No	No	No	Yes	192x144 x L60	360x216 x L40	HadGEM2-CC	Unrestricted	BADC	ESG-BADC	
19c	UOED	UK	PMIP3 (1200)			PMIP3	CMIP3	Yes	Yes	Yes	Yes	No	96x73 x L19	288x144 x L20	HadCM3	Unrestricted	BADC	ESG-BADC	
20	UNSW	Australia	PMIP3 (1000)	PMIP3 (500)	Running June 2012	PMIP3	PMIP3	No	Yes	Yes	Yes	No	64x56 x L18	128x112 x L21	CSIRO-Mk3L-1-2	Non-commercial	IPSL (IDRIS later?)	ESG-BADC	Doc
21	UofT	Canada	Completed	Completed	Completed	Completed	Completed	No	No	No	No	No	256x128 x L26	320x386 x L40	UofT-CCSM3	Unrestricted	?	?	

Click for up-to-date exp check: piControl ? | midHolocene ? | lgm ? | past1000 ? | 1pctCO2 ?

Legend

Status expected completion date	Available in CMIP5 DB (nb ens X nb years in CMIP5 DB)	Available in CMIP3 DB
No	Available in PMIP3 DB (nb years in PMIP3 DB)	
Yes		

PMIP3 + PlioMIP groups: Note that the models used for PlioMIP are often not (exactly) the same as the ones used for PMIP3/CMIP5

PlioMIP-only (not in the table): LPAP, UoM

PMIP3 community

- 21 institutes (25 models)
- pmip-announce list 600 registered users

Note: models have ***Terms of use***

Unrestricted

Non-commercial

The data life cycle of distributed data (ultra brief overview)

Model (or observation) data

CMORization

Use project-specific CMOR tables and scripts
(CMOR = Climate Model Output Rewriter)

Standardized data

Tech note: *the publication step currently depends **strictly** on correctly defined data files. Some projects may require something easier to use*

Data *publication*

Configure an existing node for hosting, referencing and distributing the project data

Standardized data available on a node of the distributed DB

Data search portal

Happy user who can start doing some comparison science

Standard file names and path

Standardization document for a model intercomparison project:

➔ the **DRS** (Data Reference Syntax) document

Using **CMOR library** + project specific CMOR **MIP tables** ➔ you get:

- Standard file content
  - CF-compliant metadata in the NetCDF file
  - Standard variable names, axis names, …
- Standard file names

  <variable name>_<MIP table>_<model>_<experiment>_
  <ensemble member>[_<*temporal subset*>].nc
  with <*temporal subset*> = 'yyyy[mm[dd[hh][mm]]][-clim]'
- Eg. PMIP3/output/IPSL/IPSL-CM5A-LR/lgm/
  monClim/atmos/Aclim/r1i1p1/v20120418/
  tas/tas_Aclim_IPSL-CM5A-LR_lgm_r1i1p1_260101-280012-clim.nc

Using ESGF to download some data (at last!)

We are just a few steps away from doing some science!

- Find the name of the required variable(s)
- Use the portal to
  - Determine if the variable(s) are available
  - Create a *wget script* to download the variable(s)
- Create an ESGF account to download the data
- Execute the wget script on a Linux machine to download the data

There are other (non standard) ways to access the data…

- @IPSL: CMIP5 mirror on ciclad
- @IPSL: custom portals (for accessing the CMIP5 mirrored data)
  - https://prodiguer.ipsl.fr/
  - http://ocmip5.ipsl.fr/FileFinderAR5/
- Synchro-data (tool used for filling the IPSL CMIP5 mirror)
  - http://dods.ipsl.jussieu.fr/jripsl/synchro_data/README
- …

Variables requested for CMIP5 (and PMIP3)

- It's useful to know the standard name of the variable(s) you are looking for ☺
  - … and the MIP table (CMOR table) where it is defined
- The available variables can be
  - project specific
  - similar to CMIP5

CMIP5(-like) variables defined in CMIP5 *Standard Output* document

http://cmip-pcmdi.llnl.gov/cmip5/output_req.html

Everything (names, units, axes, conventions, …) is defined in the giant xls file

Most tabs of the file are associated with a CMOR table

fx → *CMIP5_fx* table ⇔ Time-Invariant Fields

Eg: *sftlf* → land_area_fraction, %, *atmos* realm

Omon ⇔ Monthly Mean Ocean Fields, Including Biogechemical Fields

Eg: *tos* → sea_surface_temperature, K, *ocean* realm, *mon* frequency

The ESGF P2P FE: **Home** page

http://esgf-node.ipsl.fr/esgf-web-fe/

Access to the **Search interface**

Recommended node for **IPSL** users

Other P2P nodes

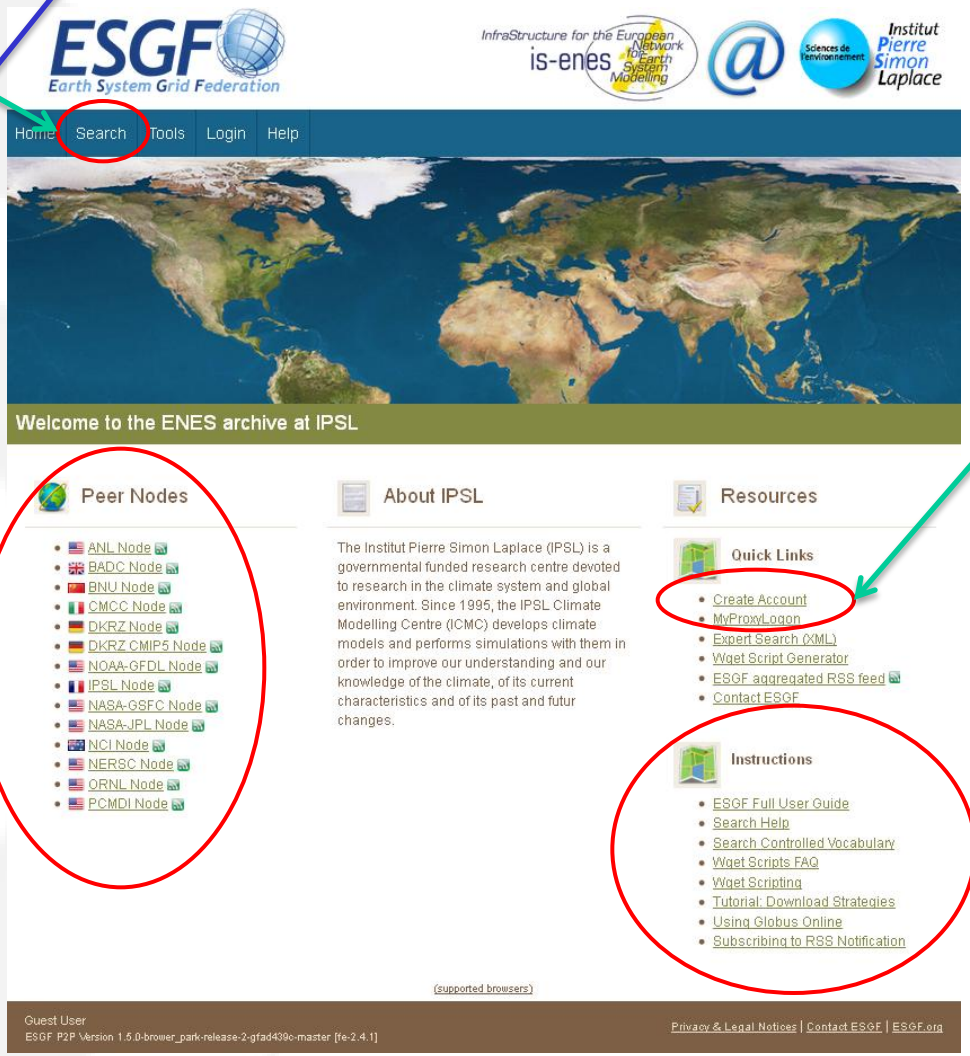
You should use the one closest to your current location !

ESGF **account creation** page

An account created on any ESGF node will work on **all** the nodes! ☺

No need for an account for searching the data, only for dowloading it

Help pages



The ESGF P2P FE: account creation page

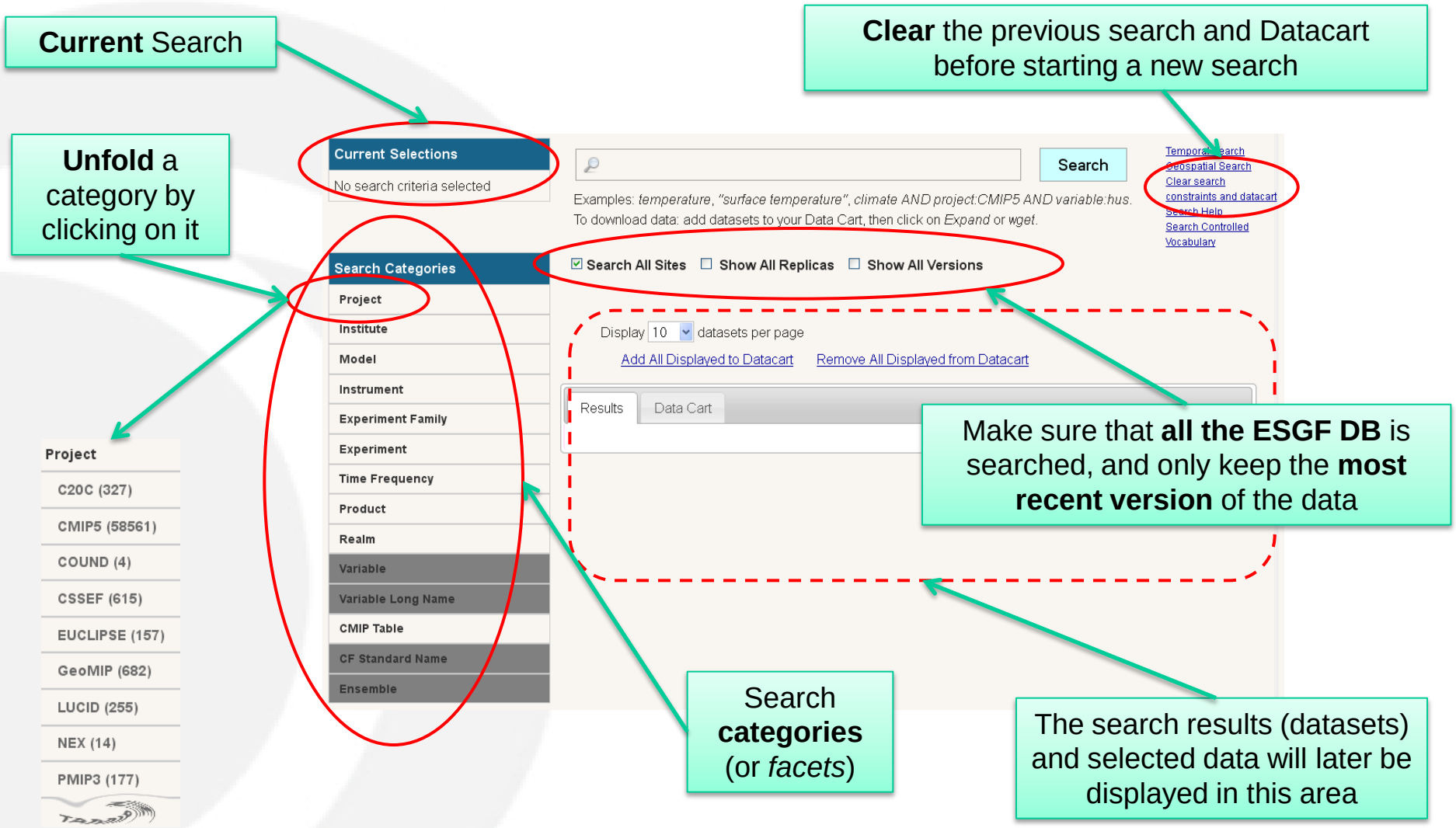
The ESGF account:

- is only required for downloading data

  No need for an account for checking the availability of data or searching!

- can be created on any P2P Front-End
- is active as soon as the request form is filled
- Is the same on the whole ESG Federation

The ESGF P2P FE: **Search** page



Finding which models have midHolocene data

Restrict the results by selecting the ***Paleo*** exps

Add the ***midHolocene*** filter

Unfold the ***Model*** facet to display the models that have midHolocene data

Current Selections

(x) experiment family:Paleo

Search Categories

Project

CMIP5 (442)

PMIP3 (73)

Institute

Model

Instrument

Experiment Family

Paleo (515)

Experiment

lgm (128)

midHolocene (245)

past1000 (142)

Data is available for 2 projects !

Current Selections

remove all

(x) experiment family:Paleo

(x) experiment:midHolocene

Search Categories

Project

CMIP5 (196)

PMIP3 (49)

Institute

Model

Instrument

Experiment Family

Paleo (245)

Experiment

midHolocene (245)

Model

BCC-CSM1.1 (17)

CCSM4 (18)

CNRM-CM5 (19)

CSIRO-Mk3.6.0 (13)

CSIRO-Mk3L-1-2 (12)

EC-EARTH-2-2 (5)

FGOALS-g2 (19)

FGOALS-s2 (6)

GISS-E2-R (9)

HadGEM2-CC (17)

HadGEM2-ES (15)

IPSL-CM5A-LR (21)

KCM1-2-2 (22)

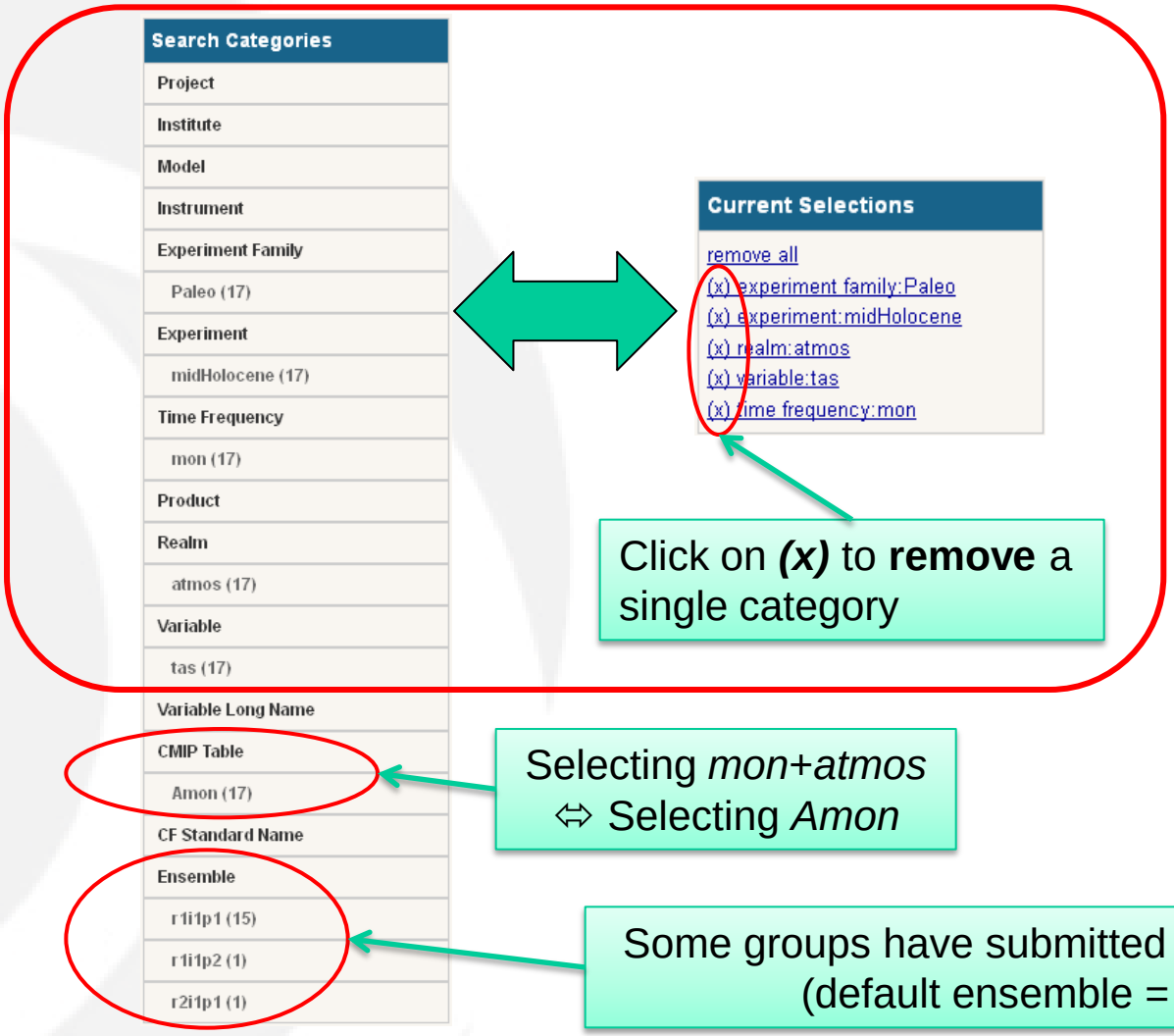
MIROC-ESM (16)

MPI-ESM-P (23)

MRI-CGCM3 (13)

Finding midHolocene surface temperature (*mon* freq)

Just keep sub-selecting new categories !



The search results (datasets)

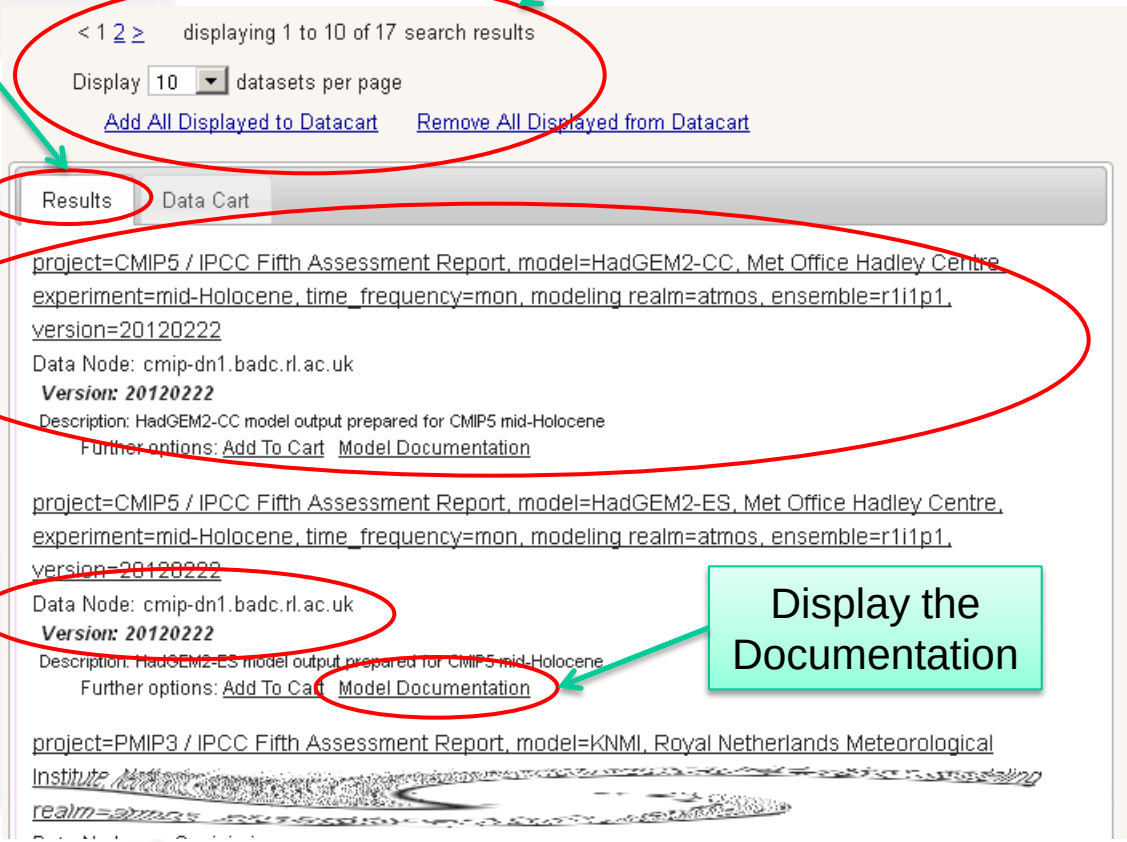
WARNING! Select a number of displayed datasets large enough to **display all the datasets**, check that the **datacart is empty**, then click on ***Add All Displayed to Datacart***

Matching datasets are displayed in the Results tab

One of the datasets

Physical **location** and **version** of a dataset

Display the Documentation

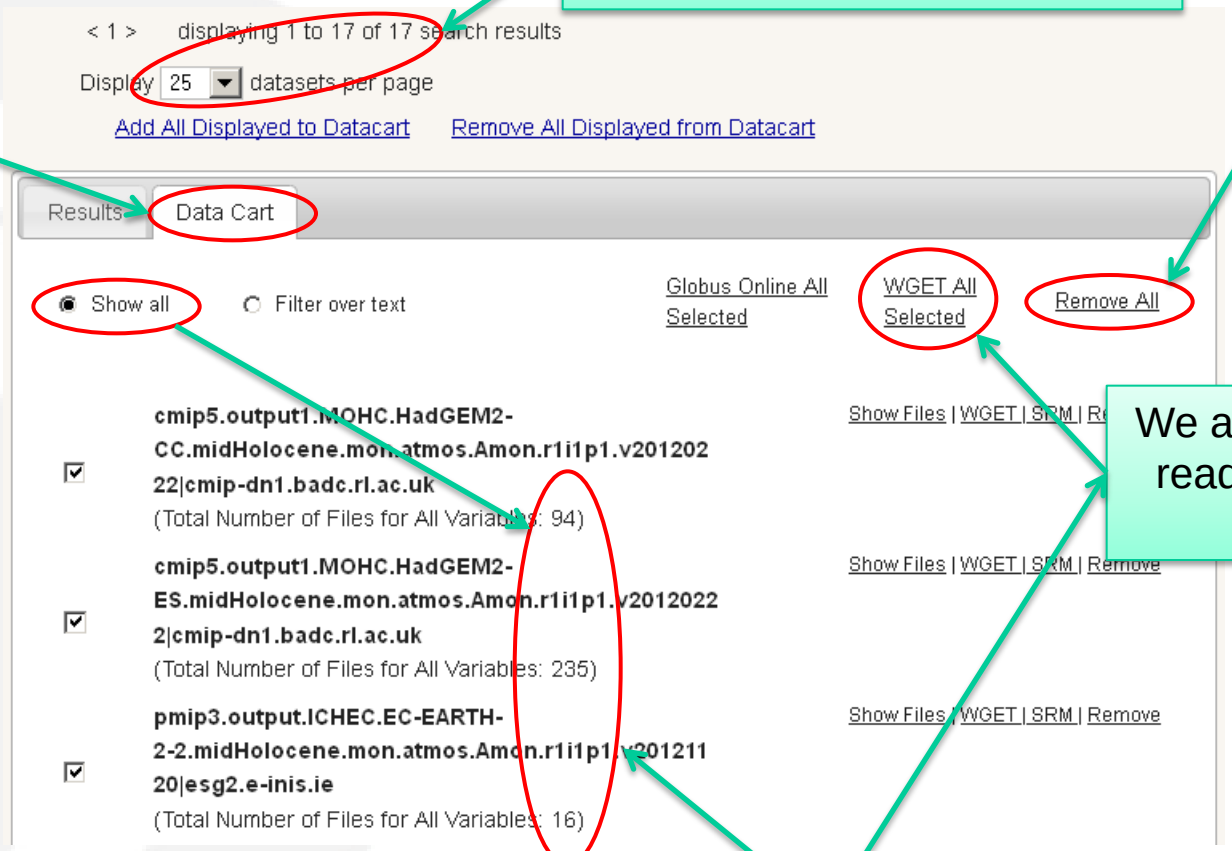


The datacart (filled with full datasets)

Good! All the results were displayed in the Results tab

Datacart view

When you need to clear the datacart…

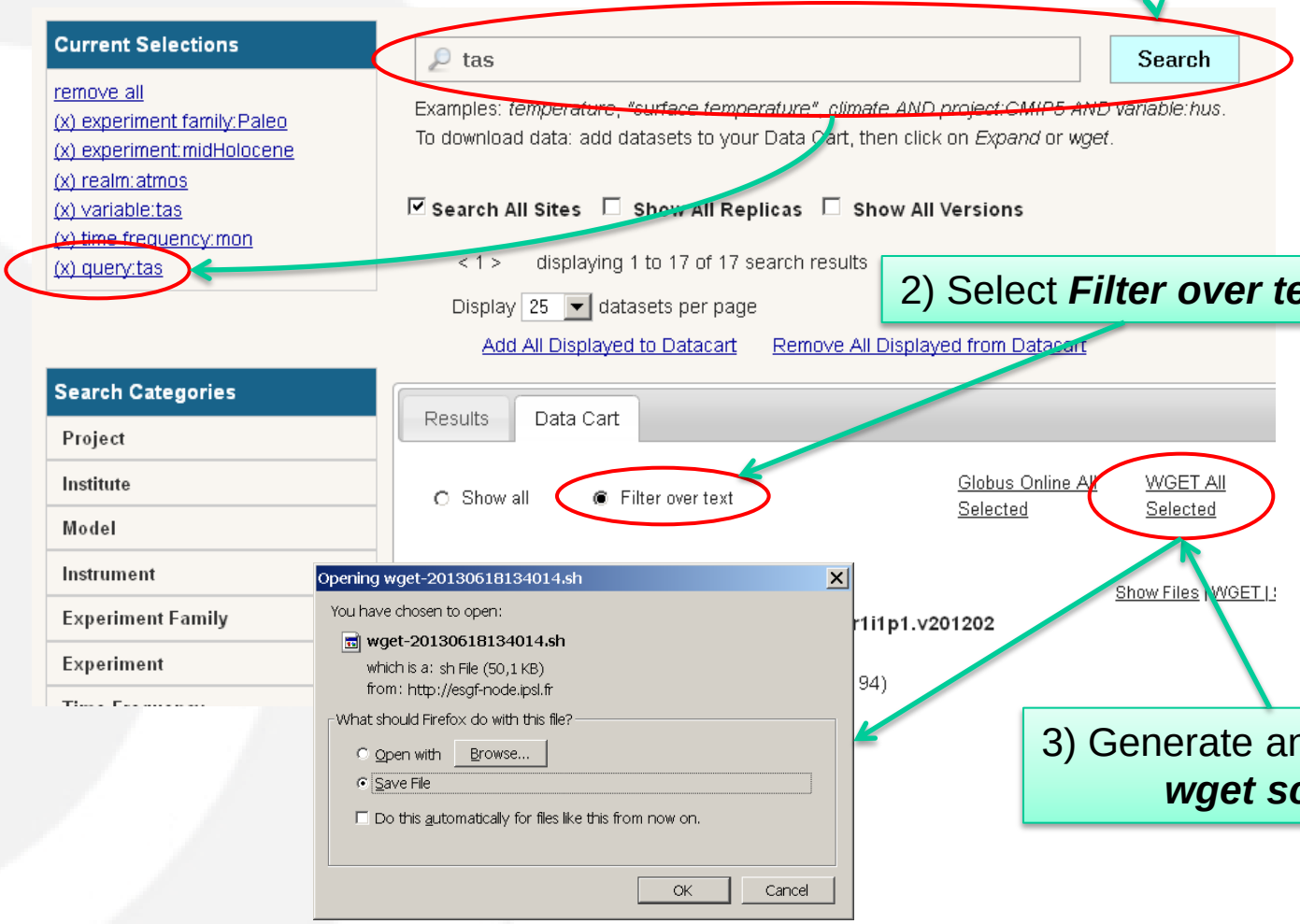


We are not yet ready to use this

Warning! By default, **ALL** the files of the datasets are selected, not just the **tas** variable we are looking for…

The *filtered* datacart

1) Type ***tas*** in the text field, and click Search
→ This creates an extra ***query:tas*** field



2) Select ***Filter over text***

3) Generate and save the ***wget script***

The ESGF wget script

- The wget script generated from a filtered datacart
  - Can be executed to download the selected variable
  - Will check the integrity of the downloaded files
  - Can be run again to resume a failed file transfer
  - Has lots of options (security, etc…): **wget-<date>.sh –h**

- Using the script
  - Save the script in a new directory, because all the files will be downloaded in the same directory (no subdirs created)
  - Edit the script and check the content of the *download_files* var
  - chmod +x wget-<date>.sh

    ./ wget-<date>.sh
  - WARNING! The script requires an ESGF account

    Eg **https://ipcc-ar5.dkrz.de/myopenid/jypmip**

Download the data and start working ☺

```
jypeter@asterix2 - ...P2P_demo - 50 >./wget-20130618134014.sh
Running wget-20130618134014.sh version: 1.3.2
Use wget-20130618134014.sh -h for help.

There were files with the same name which were requested to be download to the same directory. To avoid ov
erwriting the previous downloaded one they were skipped.
Please use the parameter 'download_structure' to set up unique directories for them.
Script created for 96 file(s)
(The count won't match if you manually edit this file!)



OpenSSL 1.0.0-fips 29 Mar 2010
** 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://ipcc-ar5.dkrz.de/myopenid/jypmip)?
MyProxy Password?
Retrieving Credentials...done!
tas_Amon_CCSM4_midHolocene_r1i1p1_100001-130012.nc ...Downloading
--2013-06-18 14:13:29--  http://tds.ucar.edu/thredds/fileServer/datazone/cmip5_data/cmip5/output1/NCAR/CCS
M4/midHolocene/mon/atmos/Amon/r1i1p1/v20120604/tas/tas_Amon_CCSM4_midHolocene_r1i1p1_100001-130012.nc
Resolving tds.ucar.edu... 128.117.181.47
Connecting to tds.ucar.edu|128.117.181.47|:80... connected.
HTTP request sent, awaiting response... 302 Moved Temporarily
Location: https://tds.ucar.edu/OpenidRelyingParty/home.htm?redirect=http%3A%2F%2Ftds.ucar.edu%2Fthredds%2F
fileServer%2Fdatazone%2Fcmip5_data%2Fcmip5%2Foutput1%2FNCAR%2FCCSM4%2FmidHolocene%2Fmon%2Fatmos%2FAmon%2Fr
1i1p1%2Fv20120604%2Ftas%2Ftas_Amon_CCSM4_midHolocene_r1i1p1_100001-130012.nc [following]
--2013-06-18 14:13:30--  https://tds.ucar.edu/OpenidRelyingParty/home.htm?redirect=http%3A%2F%2Ftds.ucar.e
du%2Fthredds%2FfileServer%2Fdatazone%2Fcmip5_data%2Fcmip5%2Foutput1%2FNCAR%2FCCSM4%2FmidHolocene%2Fmon%2Fa
tmos%2FAmon%2Fr1i1p1%2Fv20120604%2Ftas%2Ftas_Amon_CCSM4_midHolocene_r1i1p1_100001-130012.nc
Connecting to tds.ucar.edu|128.117.181.47|:443... connected.
HTTP request sent, awaiting response... 302 Moved Temporarily
Location: http://tds.ucar.edu/thredds/fileServer/datazone/cmip5_data/cmip5/output1/NCAR/CCSM4/midHolocene/
mon/atmos/Amon/r1i1p1/v20120604/tas/tas_Amon_CCSM4_midHolocene_r1i1p1_100001-130012.nc [following]
--2013-06-18 14:13:31--  http://tds.ucar.edu/thredds/fileServer/datazone/cmip5_data/cmip5/output1/NCAR/CCS
M4/midHolocene/mon/atmos/Amon/r1i1p1/v20120604/tas/tas_Amon_CCSM4_midHolocene_r1i1p1_100001-130012.nc
Connecting to tds.ucar.edu|128.117.181.47|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 799025508 (762M) [application/x-netcdf]
Saving to: "tas_Amon_CCSM4_midHolocene_r1i1p1_100001-130012.nc"

 0% [                                                                       ] 17,143      58.9K/s
```