Interfacing with data using the CEDA Web Processing Service

Eduardo D. da Costa (STFC CEDA) &
Ag Stephens (STFC CEDA)

KNMI, Nov. 2014

Overview

1. Introducing the CEDA WPS
2. The built-in User Interface
3. Interfacing the WPS with Climate Data:
    a. Specialised processes for CMIP5
    b. Generic processes for wider use

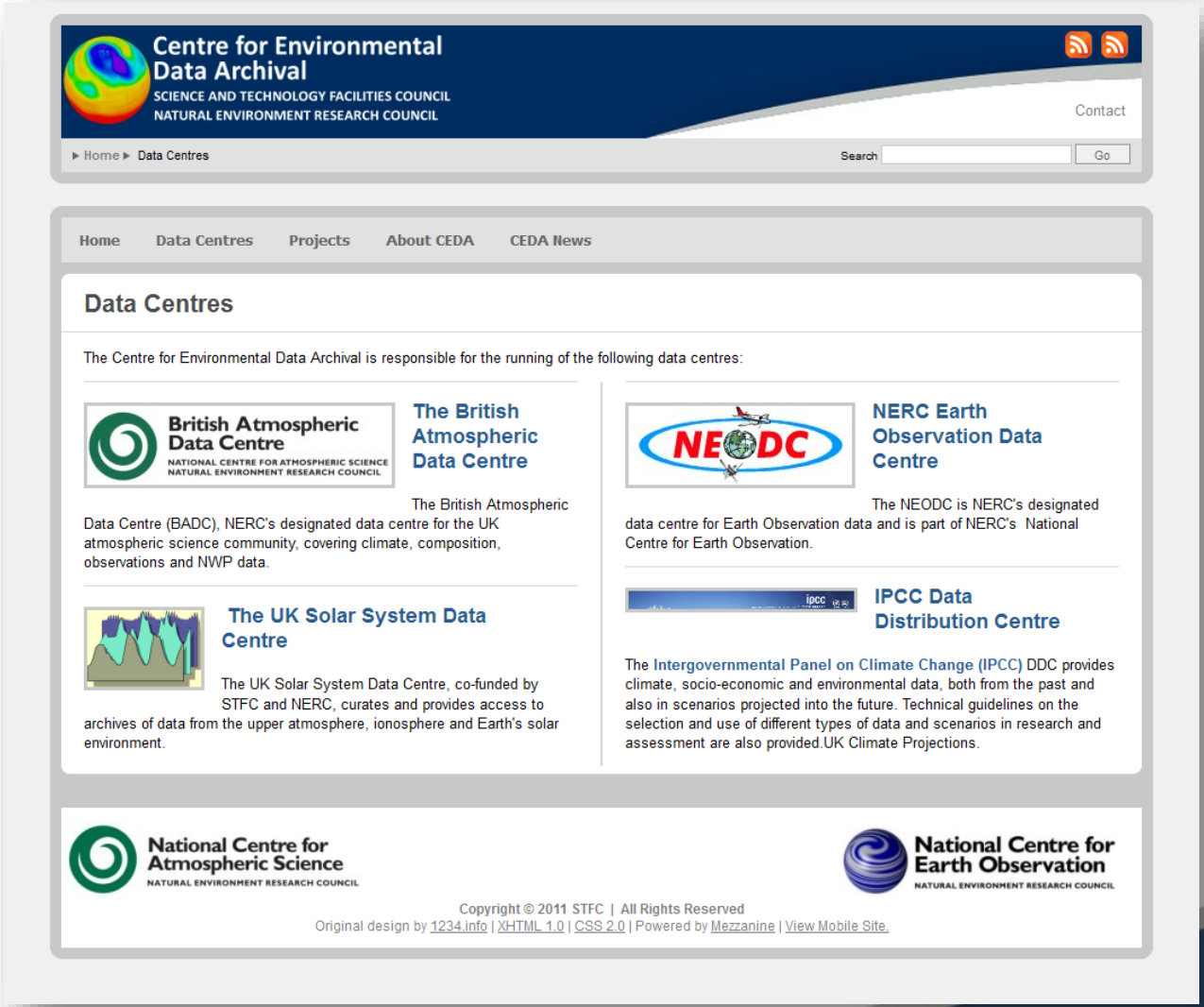
Users looking for environmental data

And they find:

The Centre for Environmental Data Archival (CEDA)

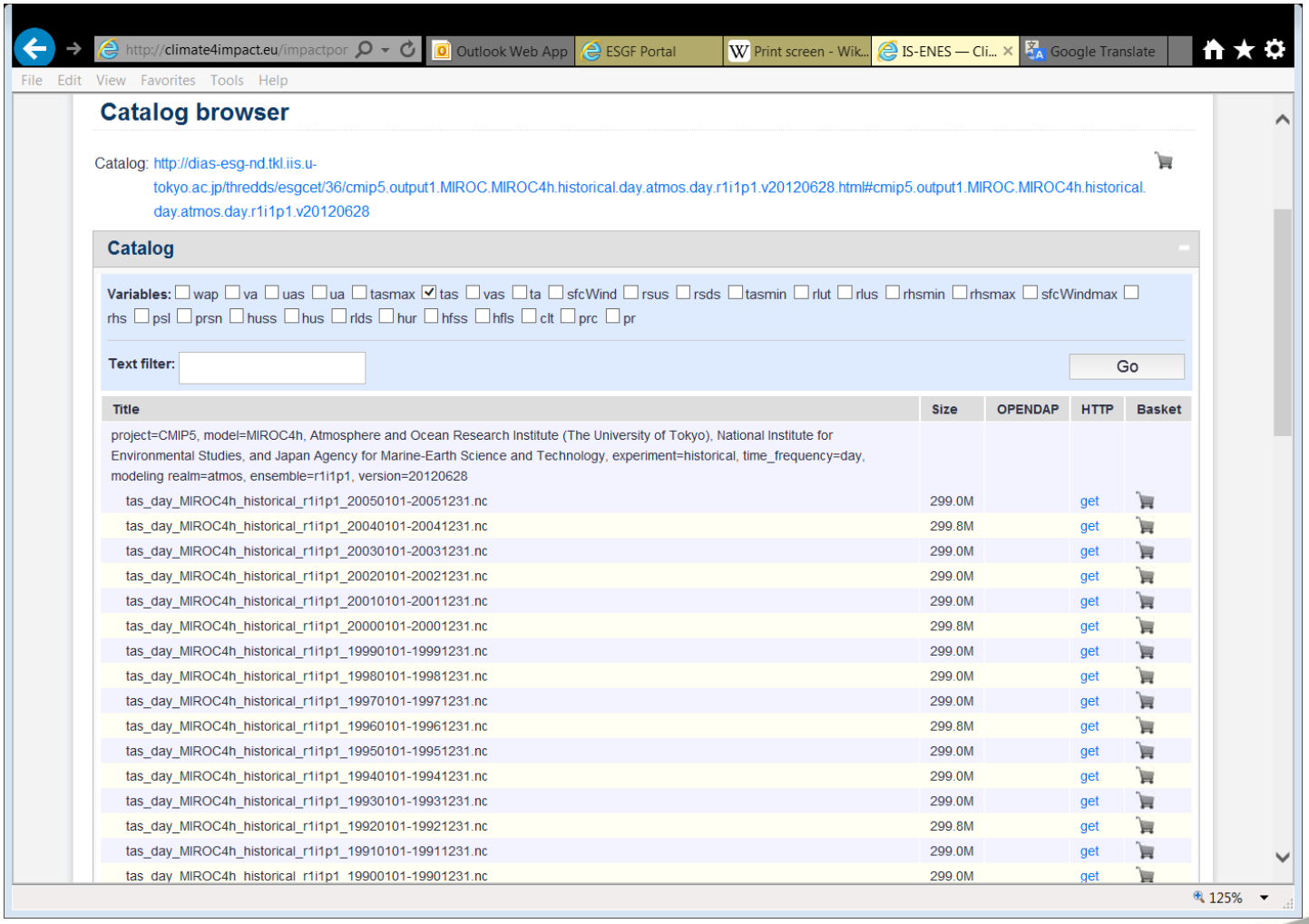
www.ceda.ac.uk

- BADC
- NEODC
- UKSSDC
- IPCC DDC



And we give them:

Files



Usability

Our strength: we are a data centre holding the data

Our problem (one of them): we are not give to the users the data in the best format for their needs:

Are the files ready to be used?
Could we improve their usability?
Aren't they usable enough already?

Our tool : WPS to provide the user with a simple way to modify some characteristics of the format:

- Time span
- Area
- Grid resolution

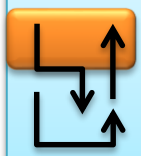
COWS WPS in a nutshell

STANDARDS COMPLIANCE

With OGC WPS 1.0.0 Standard

OGC

ASYNCHRONOUS PROCESSING



Framework handles synchronous and asynchronous processes.

WEB-INTERFACE

Web-client automatically generates job submission and monitoring forms.

BUILT ON COMMON TOOLS

matplotlib

python

NCO

netCDF

Installs with a set of common packages and tools.

EASY TO ADD NEW PROCESSES

Simple interface to plug in your own code.

SCALABLE ARCHITECTURE

For batch processing.

Features of COWS WPS

- COWS WPS adds functionality to the standard spec:
  - Enabled **"cost-only" request** to estimate the job size and duration without executing the process.
  - XML listing of previous **users jobs** and download access
  - **Zipped output files** with XML content description
  - **OpenDap endpoint** to outputs (when NetCDF).
  - **E-mail to user** when a job has completed (or failed).

The COWS WPS User Interface

Clients can talk directly to the WPS server. However, they can also use the User Interface…

- The **User Interface (UI)** has been added to enable users to interact with the WPS through a browser.
- For submission and monitoring the status of jobs: it talks XML to the WPS server (following the 1.0.0 specification).

VIEWING AVAILABLE PROCESSES

CMIP5 Global Averager
View details | Process XML | Submit a request — See USER GUIDE

The "CMIP5 Global Averager" process accepts accepts a set of CMIP5 data files defined in either a file group URL (pointing to an XML document) or a list of input file paths. These files are processed to calculate a global average using the Climate Data Operators (CDO) tools. Outputs are written to NetCDF files. Please see the general disclaimer and the disclaimer specific to CDO-related processes.

CMIP5 Regional Extractor
View details | Process XML | Submit a request — See USER GUIDE

The "CMIP5 Regional Extractor" process accepts accepts a set of CMIP5 data files defined in either a file group URL (pointing to an XML document) or a list of input file paths. These files are processed to extract a regional box from global fields using the Climate Data Operators (CDO) tools. Outputs are written to NetCDF files. Please see the general disclaimer and the disclaimer specific to CDO-related processes.

CMIP5 Regridder With File Selection
View details | Process XML | Submit a request — See USER GUIDE

The "CMIP5 Regridder" process accepts a set of CMIP5 data files defined in either a file group URL (pointing to an XML document) or a list of input file paths. These files are re-gridded to a regular 1 or 2 degree grid using the Climate Data Operators (CDO) tools. Outputs are written to NetCDF files. Please see the general disclaimer and the disclaimer specific to CDO-related processes.

CMIP5 Time Averager
View details | Process XML | Submit a request — See USER GUIDE

The "CMIP5 Time Averager" process accepts a set of CMIP5 data files defined in either a file group URL (pointing to an XML document) or a list of file paths. These files are processed to calculate a time average using the Climate Data Operators (CDO) tools. Outputs are written to NetCDF files. Please see the general disclaimer and the disclaimer specific to CDO-related processes.

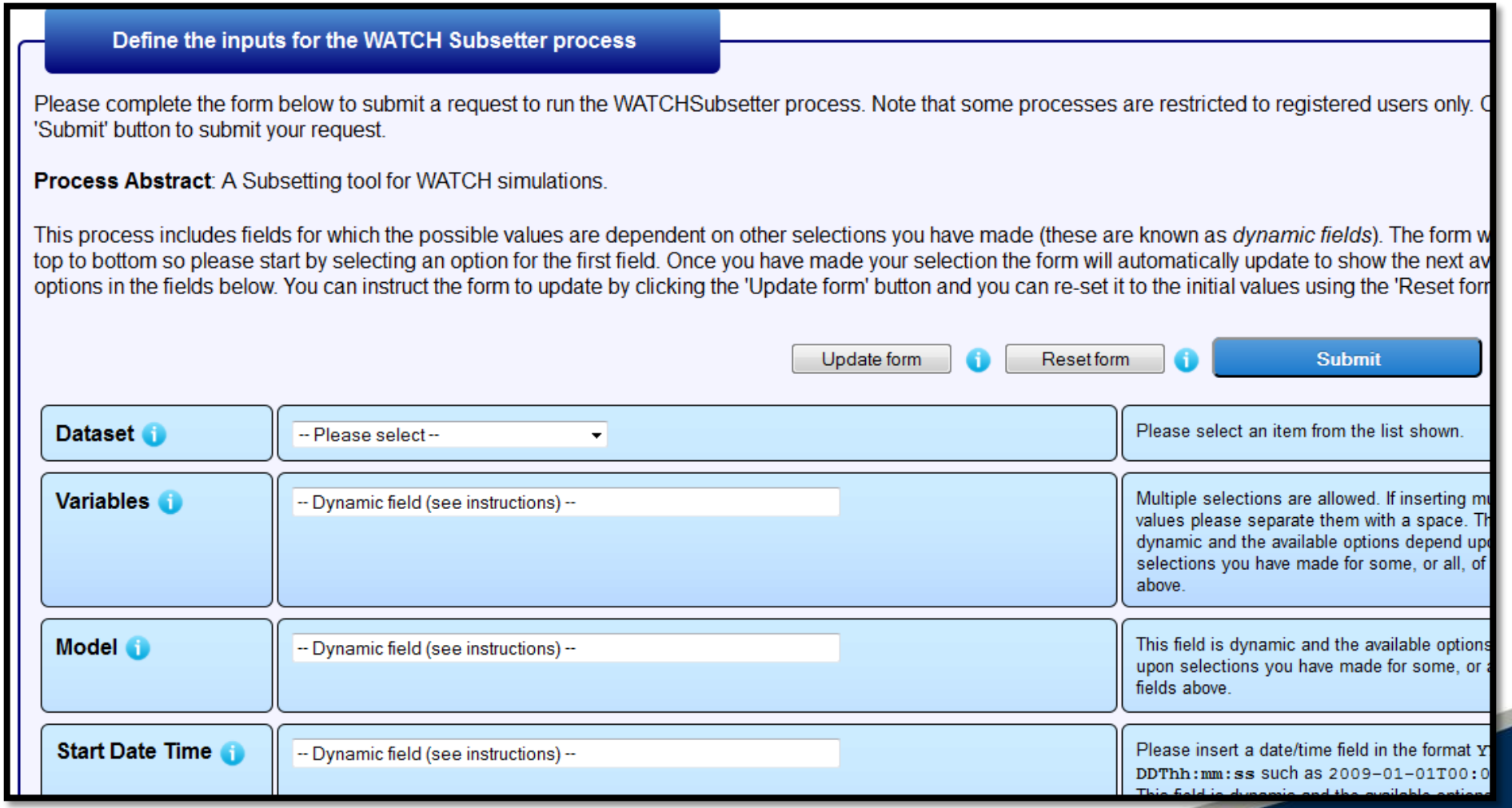
View process details (HTML)

View "Describe Process" (XML)

Submit a request (via UI)

View User Guide (HTML)

SUBMISSION FORM
Automatically generated



Define the inputs for the WATCH Subsetter process

Please complete the form below to submit a request to run the WATCHSubsetter process. Note that some processes are restricted to registered users only. C
'Submit' button to submit your request.

Process Abstract: A Subsetting tool for WATCH simulations.

This process includes fields for which the possible values are dependent on other selections you have made (these are known as *dynamic fields*). The form w
top to bottom so please start by selecting an option for the first field. Once you have made your selection the form will automatically update to show the next av
options in the fields below. You can instruct the form to update by clicking the 'Update form' button and you can re-set it to the initial values using the 'Reset forr

Update form | Reset form | Submit

Dataset	-- Please select --	Please select an item from the list shown.
Variables	-- Dynamic field (see instructions) --	Multiple selections are allowed. If inserting mu values please separate them with a space. Th dynamic and the available options depend up selections you have made for some, or all, of above.
Model	-- Dynamic field (see instructions) --	This field is dynamic and the available options upon selections you have made for some, or a fields above.
Start Date Time	-- Dynamic field (see instructions) --	Please insert a date/time field in the format Y DDThh:mm:ss such as 2009-01-01T00:0

Science & Technology Facilities Council
Rutherford Appleton Laboratory

SUBMISSION FORM

Supports types such as: float, int, string, date-time and bounding box

Bounding Box

Current Bounding Box: -180.0,-55.8,180.0,85.8 Reset bounding box

North : 85.8

West : -180.0 East : 180.0

South : -55.8

Please select a valid bounding box within the follow
geographical extent: -180., -55.75, 180., 85.75.

Output Format

nc

Please select an item from the list shown.

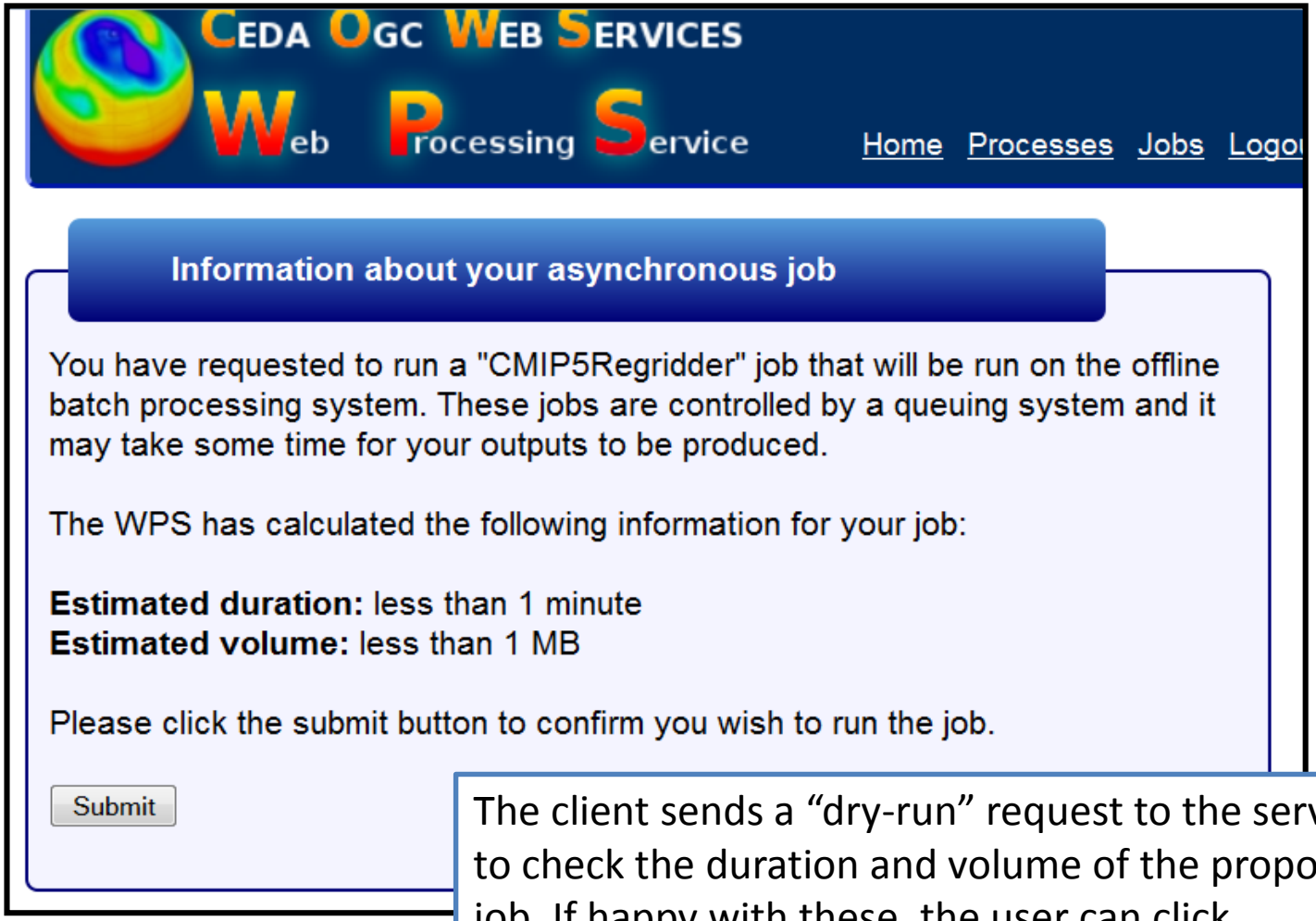
Split Output Files Into Time Chunks

AUTOMATIC

Please select an item from the list shown.

Update form Reset form Submit

SUBMITTING A JOB: The "dry-run"



The client sends a "dry-run" request to the server to check the duration and volume of the proposed job. If happy with these, the user can click "Submit" to run the job asynchronously.

THE JOB IS RUNNING: polling the WPS server

The client polls the server until the job is completed. However, if you close the client the job still runs and the user receives an e-mail regardless to say it has completed.

Normal View

Your job is currently running. The job is 0% complete....Polling again in 5 seconds...
Cancel this job | stop polling

DETAILS

Job ID	be86a1532b360abd34e4510fb4bfef67
Process ID	CMIP5Regridder
Completion Time	Undefined
Job Request URL	http://ceda-wps2.badc.rl.ac.uk/wps?Request=Execute&Identifier=CMIP5Regridder&Format=text/xml&Inf Status=false&DataInputs=FileGroupURL=http%3A//ceda-wps2.badc.rl.ac.uk/dl/0/49e564cbb14add7fa2f /cmip5_file_group.xml;RegridTo=2%20degree%20lat/lon
Request Description	The "CMIP5 Regridde document) or a list of Operators (CDO) tool CDO-related process
Request Type	data sync
Job Capabilities	download
Estimated Job Duration	Less than 1 minute

CEDA WPS Job Complete - Message (Plain Text)

Message Adobe PDF

Extra line breaks in this message were removed.

From: no-reply@ceda.ac.uk
Sent: Mon 21/01/2013 12:39
To: Stephens, Ag (STFC,RAL,RALSP)
Cc:
Subject: CEDA WPS Job Complete

Your job has successfully completed.

The outputs from your job are now available from the Jobs page at:

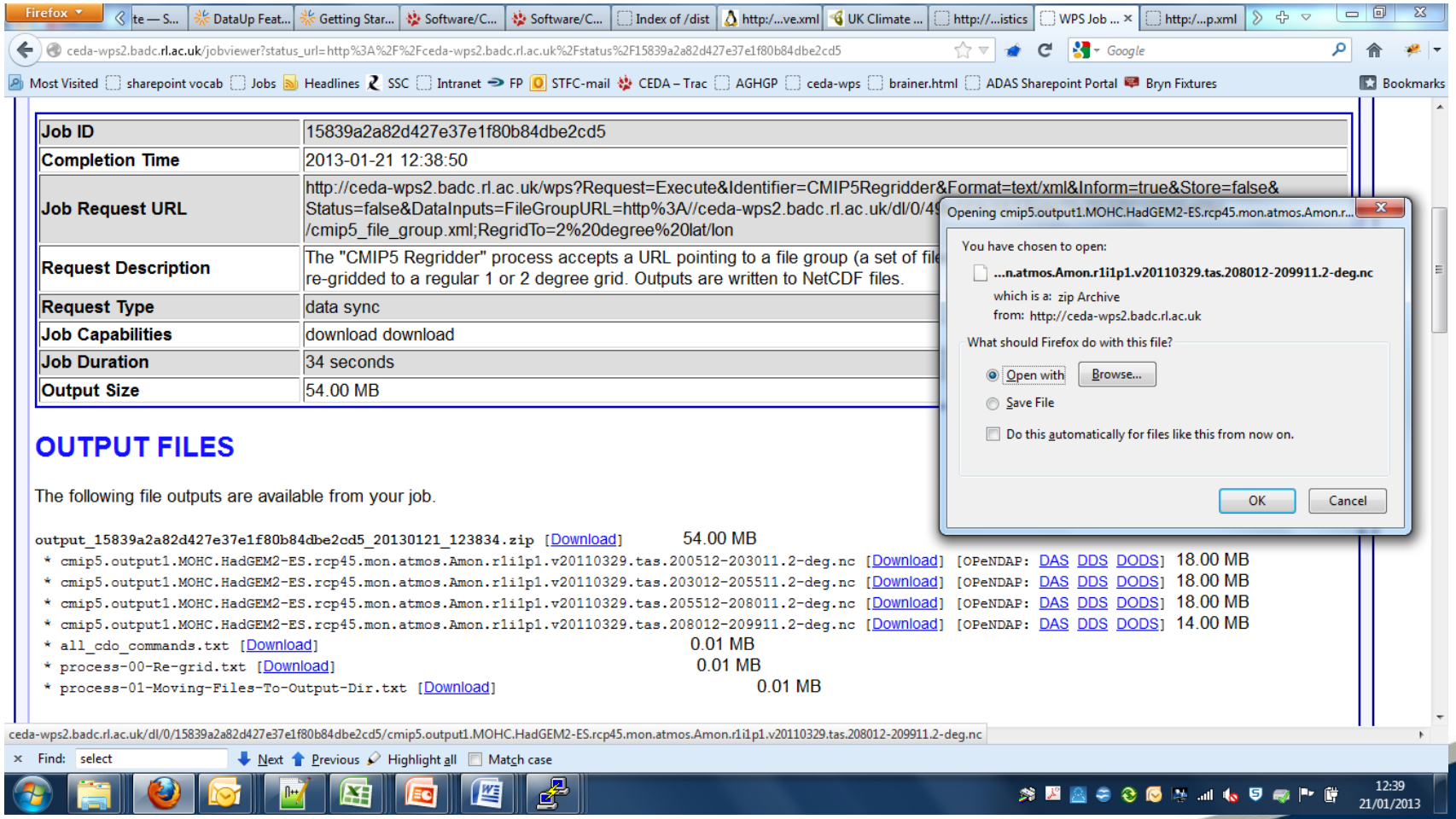
http://ceda-wps2.badc.rl.ac.uk/jobviewer?status_url=http%3A%2F%2Fceda-wps2.badc.rl.ac.uk%2Fstatus%2F15839a2a82d427e37e1f80b84dbe2cd5

Please contact the CEDA Helpdesk if you have any queries regarding this message:

badc@rl.ac.uk

Science & Technology Facilities Council
Rutherford Appleton Laborat

JOB COMPLETION:
Access to outputs and informed by e-mail



Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

JOB VIEWER
Showing XML response

```
<DataInputs>
        <Input>
                <ows:Identifier>BoundingBox</ows:Identifier>
                <ows:Title>BoundingBox</ows:Title>
                <LiteralValue>-42.9,20.0,81.6,68.9</LiteralValue>
        </Input><Input>
                <ows:Identifier>Dataset</ows:Identifier>
                <ows:Title>Dataset</ows:Title>
                <LiteralValue>WATCH 21st Century Driving Data</LiteralValue>
        </Input><Input>
                <ows:Identifier>EndDateTime</ows:Identifier>
                <ows:Title>EndDateTime</ows:Title>
                <LiteralValue>2002-12-31T12:00:00</LiteralValue>
        </Input><Input>
                <ows:Identifier>Model</ows:Identifier>
                <ows:Title>Model</ows:Title>
                <LiteralValue>CNCM3 (SRES A2)</LiteralValue>
        </Input><Input>
                <ows:Identifier>OutputFormat</ows:
                <ows:Title>OutputFormat</ows:Title
                <LiteralValue>nc</LiteralValue>
```

This is the WPS Standard "Execute Response" XML Document

JOBS PAGE
Search previous jobs

Welcome to the Jobs Page

The Jobs Page allows you to view current and previous jobs. Select the filter options to modify the records matched.

Filter the Job List

Please select the filtering options and press the 'Filter jobs' button to submit the new filter settings.
Hide filter options

Job ID matches (regular expression)

Job Status -- Please select --

Process ID -- Please select --

Job Type Asynchronous

Job submitted before (format: YYYY-MM-DDThh:mm:ss)

Job submitted after (format: YYYY-MM-DDThh:mm:ss)

User ID matches (regular expression)

Click the button to re-load the page with these settings Filter jobs

JOBS PAGE
View previous jobs

Jobs

Total jobs matched: 123

Request ID	Process ID	Job type	User	Submission time	Job status	Action
14dca62db650fb621d26d2337e56083a	WATCHSubsetter	Asynchronous	anon	2013-06-28 21:43:43	COMPLETED	View job info/outputs
51f25304e91e1d528ee1950fdf75b54d	WATCHSubsetter	Asynchronous	anon	2013-06-28 21:42:52	COMPLETED	View job info/outputs
6a6dde0308cd85cc7fe4271ef15f31b8	WATCHSubsetter	Asynchronous	anon	2013-06-28 21:42:23	COMPLETED	View job info/outputs
f77b645ef6cb6060c243cf7a94a4bd97	WATCHSubsetter	Asynchronous	anon	2013-06-28 14:37:12	COMPLETED	View job info/outputs
40a24f7a07917c93d84b3fbf7b7905e4	WATCHSubsetter	Asynchronous	anon	2013-06-24 17:02:32	COMPLETED	View job info/outputs
1eaf75f0fde0d5549094cdadd602b126	WATCHSubsetter	Asynchronous	anon	2013-06-24 16:57:45	FAILED or CANCELLED	View job info/outputs
7a71a1e565c1e3c91e7ea80fe7ca68b5	WATCHSubsetter	Asynchronous	anon	2013-06-17 14:08:40	COMPLETED	View job info/outputs
dbb606669d4e8182ba57d6dee35e3fa2	WATCHSubsetter	Asynchronous	anon	2013-06-17 08:42:22	COMPLETED	View job info/outputs

Our Advantage : we have some processing capability (JASMIN/CEMS "super-data-cluster" that we can share with some users

Interfacing the WPS with climate data sets

To give to the users the possibility to manipulate our holdings

Climate Data Operators (CDO) for climate data processing

- The Climate Data Operators (CDO) are an established tool for processing large climate data sets.

- We are basing much of the processing inside our WPS processes on CDO.

https://code.zmaw.de/projects/cdo

Two approaches to WPS-CDO interactions

1. Specialised processes for climate datasets (e.g. CMIP5) – using a single CDO operator.
2. Grouping large numbers of CDO operators and presenting them under a single WPS Process – more or less following the CDO Manual chapters.

Option 1: Specialised process for climate data processing using single CDO operator

- Define simple processes to work with known data sets (e.g. CMIP5).
- Each process just wraps one CDO operation.
- Provide the capability to chain one process to another (using our User Interface).
- Integrate dynamic file search into the dataset-specific process…

Science & Technology Facilities Council
Rutherford Appleton Laboratory

CMIP5 Regridder With File Selection

Institute and Model	BCC/bcc-csm1-1 BCC/bcc-csm1-1-m BNU/BNU-ESM
Experiment Family	All Atmos-only Control
Experiment	-- Dynamic field (see instructions) --
Time Frequency	-- Dynamic field (see instructions) --
Realm	-- Dynamic field (see instructions) --
Ensemble Member	-- Dynamic field (see instructions) --
Variable	-- Dynamic field (see instructions) --

LOADING...

Select files

Science & Technology Facilities Council
Rutherford Appleton Laboratory

CMIP5 Regridder With File Selection

Institute and Model

BCC/bcc-csm1-1
BCC/bcc-csm1-1-m
BNU/BNU-ESM

Select files

Experiment Family

All
Atmos-only
Control

Experiment

esmControl
piControl

Time Frequency

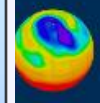
day
mon
monClim

Realm

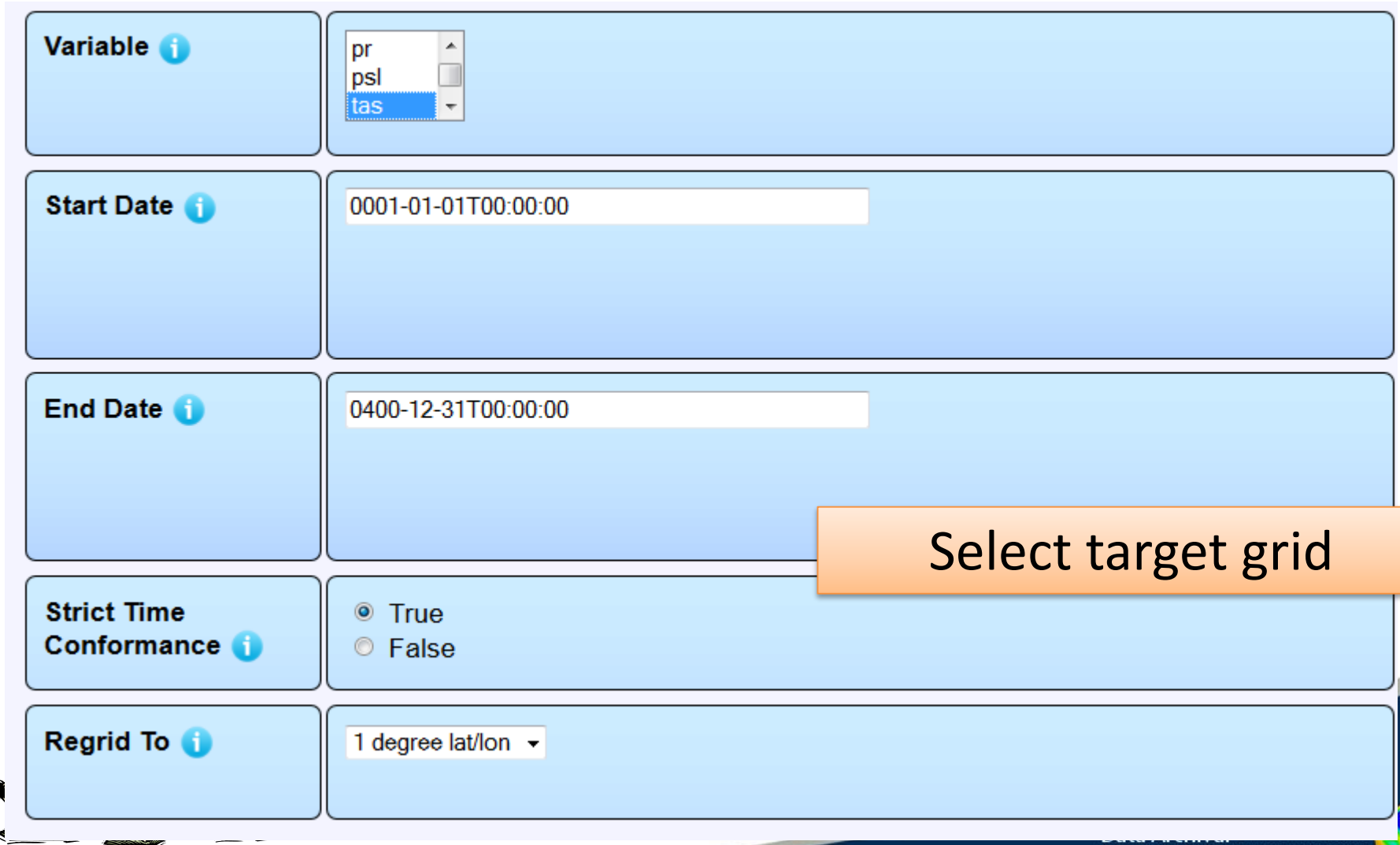
atmos
ocean

Ensemble Member

r1i1p1



CMIP5 Regridder With File Selection



Option 2: Semi-automated mapping of CDO operators to WPS Processes

- Mapping WPS processes to CDO at the command-line is simple to implement
- By defining a set of mappings between CDO operators and WPS processes it is possible to automate the generation of WPS process code, and configuration files, based on a set of common patterns.

Examples of option 2

CDO Calculate Single File Climate Indices

View details | Process XML | Submit a requ

Calls the Climate Data Operators (CDO) tool with the single file path provided and uses the chosen operator to calculate climate
process wraps the following CDO operators: ['eca_cdd', 'eca_cfd', 'eca_cwd', 'eca_fd', 'eca_id', 'eca_r10mm', 'eca_r20mm', 'ec

CDO Calculate Single File Climate Indices With Argument

View details

Calls the Climate Data Operators (CDO) tool with the file paths and argument provided and uses the chosen operator to calcula
written to a NetCDF file. This process wraps the following CDO operators: ['eca_csu', 'eca_rx1day', 'eca_rx5day', 'eca_su', 'eca

CDO Calculate Single File Statistics

View details | Process XML | Submit a request

Calls the Climate Data Operators (CDO) tool with the single file path provided and uses the chosen operator to calculate statistic
wraps the following CDO operators: ['fldmin', 'fldmax', 'fldsum', 'fldmean', 'fldavg', 'fldvar', 'fldstd', 'zonmin', 'zonmax', 'zonsum', 'zo
'mermin', 'mermax', 'mersum', 'mermean', 'meravg', 'mervar', 'merstd', 'vertmin', 'vertmax', 'vertsum', 'vertmean', 'vertavg', 'vertvar',
'timmean', 'timavg', 'timvar', 'timstd', 'hourmin', 'hourmax', 'hoursum', 'hourmean', 'houravg', 'hourvar', 'hourstd', 'daymin', 'daymax
'daystd', 'monmin', 'monmax', 'monsum', 'monmean', 'monavg', 'monvar', 'monstd', 'seasmin', 'seasmax', 'seassum', 'seasmean', 's
'yearmax', 'yearsum', 'yearmean', 'yearavg', 'yearvar', 'yearstd', 'yhourmin', 'yhourmax', 'yhoursum', 'yhourmean', 'yhouravg', 'yh
'ydaysum', 'ydaymean', 'ydayavg', 'ydayvar', 'ydaystd', 'ymonmin', 'ymonmax', 'ymonsum', 'ymonmean', 'ymonavg', 'ymonvar', 'y
'yseassum', 'yseasmean', 'yseasavg', 'yseasvar', 'yseasstd']

CDO Calculate Single File Statistics With Argument

View details | Process XM

Calls the Climate Data Operators (CDO) tool with the single file path and argument provided and uses the chosen operator to ca
written to a NetCDF file. This process wraps the following CDO operators: ['runmin', 'runmax', 'runsum', 'runmean', 'runavg', 'run

Auto-generated documentation

Process: CDOGetSingleFileInfo

Long Name: CDO Get Single File Info

Short Description: Outputs information about the contents of a single data file.

Long Description: Calls the Climate Data Operators (CDO) tool with the single file pat
information that is written to a text file.

Operators: npar, nlevel, nyear, nmon, ndate, ntime, showformat, showcode, showna
showmon, showdate, showtime, showtimestamp, pardes, griddes, zaxisdes, vct

Name	Synopsis	Parameters	Short Description	Extra
nlevel	nlevel ifile		Number of levels	Prints t

See:
http://cows.ceda.ac.uk/cows_wps/supported_procs/cdo_processes.html#process-cdogetsinglefileinfo

Current limitations of CDO-related WPS processes

Processes provided under the CEDA WPS are currently limited as follows:

1. They can **only access local data** files.
2. The security layer only interacts with the local CEDA user database.

Both of these limitations are being investigated as we seek to improve the service.

Status of CDO-related processes (1)

1. Operational CMIP5 processes:
   - CMIP5 Global Averager
   - CMIP5 Regional Extractor
   - CMIP5 Regridder
   - CMIP5 Time Averager
   - CMIP5 Time Concatenator
   - CMIP5 Time Extractor
   - Select CMIP5 Files

   http://ceda-wps2.badc.rl.ac.uk/submit/choose

Status of CDO-related processes (2)

2. Operational CDO-WPS processes:

- CDO Get Single File Info
- CDO Get Multiple File Info
- CDO Calculate Single File Climate Indices
- CDO Calculate Two File Climate Indices With Argument
- CDO Calculate Single File Statistics
- CDO Calculate Single File Statistics With Argument
- CDO Calculate Multiple File Statistics
- CDO Calculate Single File Climate Indices With Argument

http://ceda-wps2.badc.rl.ac.uk/submit/choose

Further information

- Live CEDA WPS: http://ceda-wps2.badc.rl.ac.uk/ui/home
- COWS WPS documentation: http://ceda-wps2.badc.rl.ac.uk/docs/index.html OR http://cows.badc.rl.ac.uk/cows_wps.html

- EGU Poster on the COWS WPS http://presentations.copernicus.org/EGU2011-7777_presentation.pdf

Thanks