



The CEDA Web Processing Service for rapid deployment of earth system data services

Stephen Pascoe

Ag Stephens

Phil Kershaw

Centre of Environmental Data Archival



Overview of CEDA-WPS

- History – first implementation and deployment
- Architecture
- Generalising first implementation
- Test Beds
- Operational Services

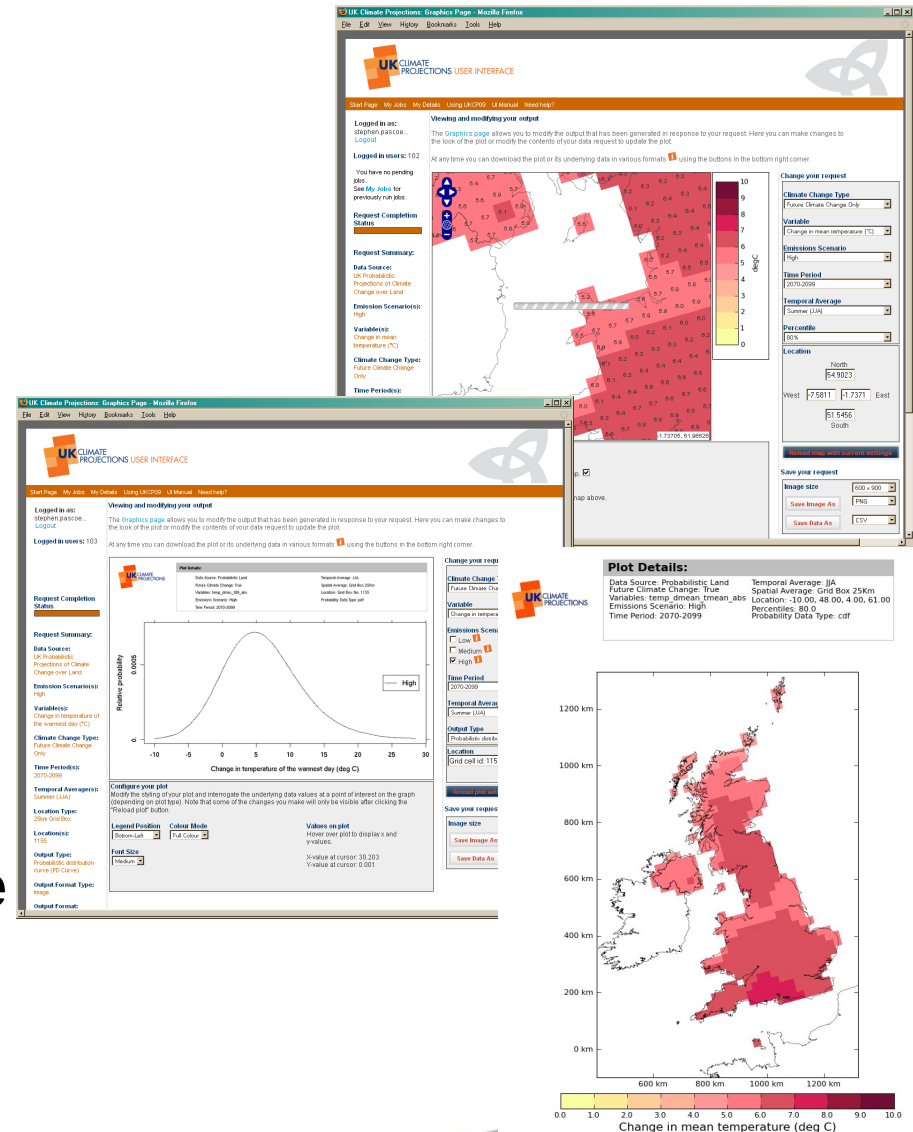


History UKCIP09 User Interface

Requirements

- Map and non-map plots
- Interactive and configurable
- Long-running tasks
 - Large data extractions
 - Secondary simulations (Weather Generator)
- Resilient to variable load

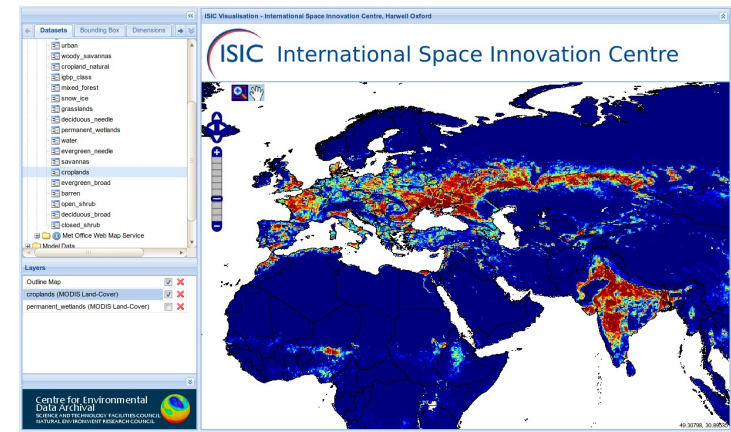
Service Orientated Architecture



History

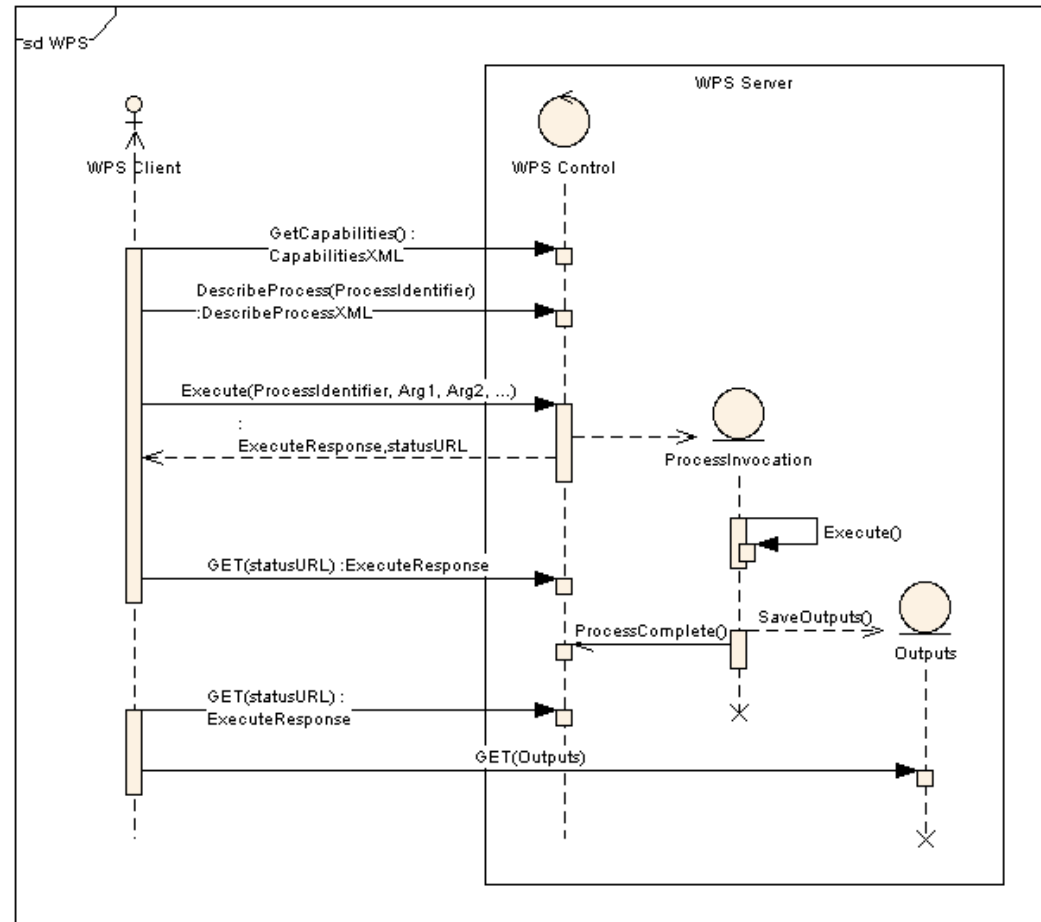
OGC Web Services at CEDA

- CEDA has developed web portals based on OGC Web Services since 2006
 - WMS clients and servers
 - WCS/WFS for complex GML features (CSML)
- NERC Data Grid
- NERC Portals Project
- ISIC Visualisation



Architecture: OGC Web Processing Service

- GET DescribeProcess resource to discover process arguments
- POST to create a process execution resource (unique URL)
- GET to poll status of process execution
- Navigate to outputs when available



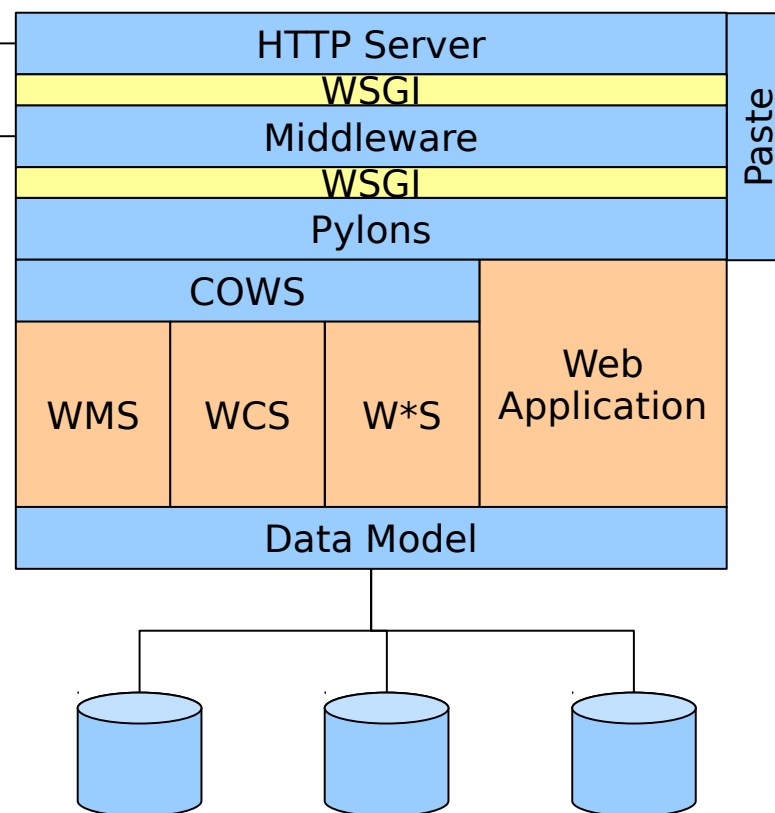
Architecture

CEDA OGC Web Services Framework: COWS

<http://cows.badc.rl.ac.uk/>

Apache + mod_python,
FastCGI, Python HTTPD

e.g. Authentication



- Library
- Standard Interface
- Application code



Architecture

CEDA OGC Web Services Framework: COWS <http://cows.badc.rl.ac.uk/>

COWS Server Implementations

cowsserver

WMS/WCS server component

cowsclient

WMS/WCS web-application

cows-wps

WPS server

COWS library

COWS

OWS operation controllers
KVP encoding and exception handling

cows.pylons

XML templates
Stub server

cows.model

OWS-Common model
implementation

cows.service

Stub service
implementations

**WMS, WCS,
WFS, W*S**

cows.csmlbackend

Data back-end for Climate
Science Markup Language
(CSML)





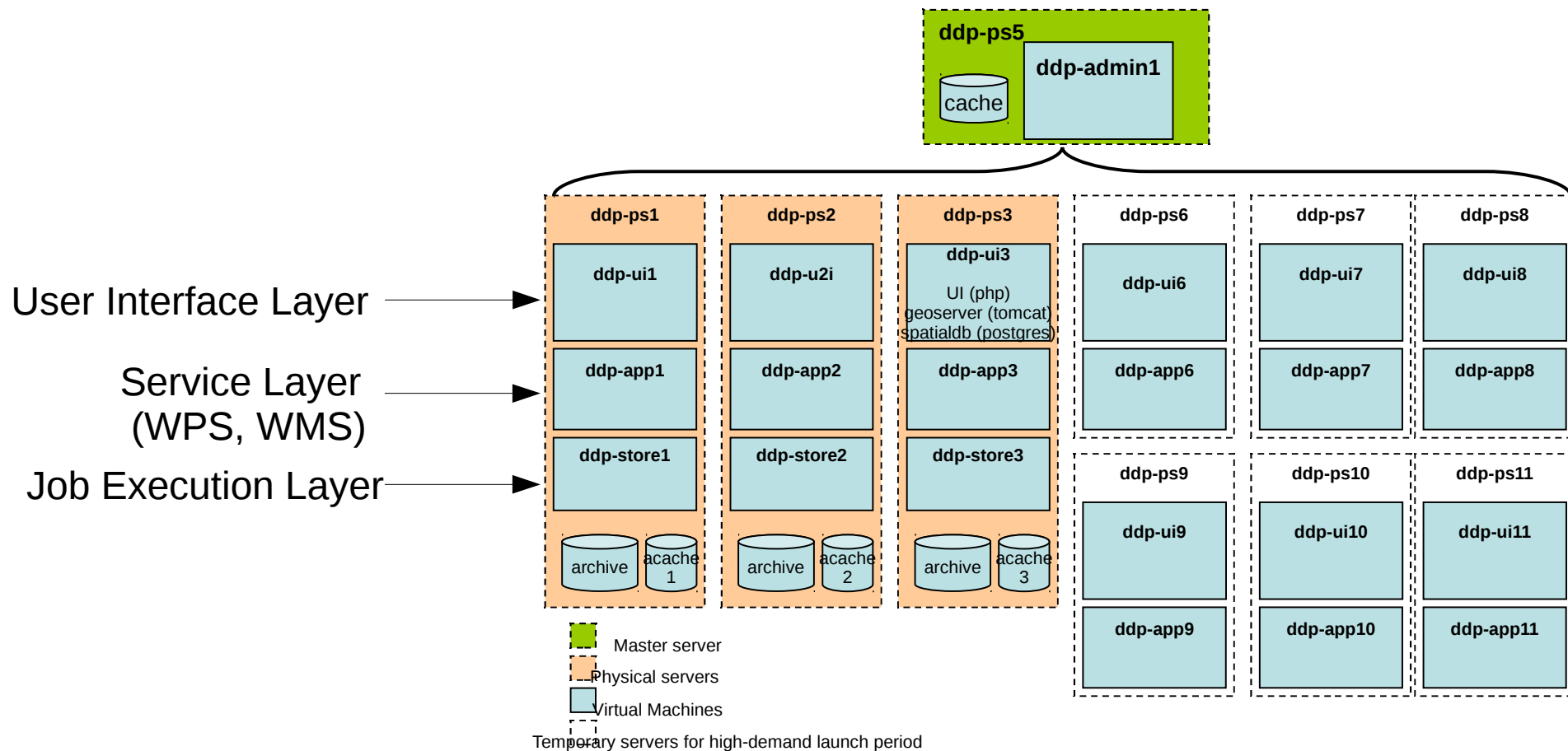
Architecture Process Modules

- Implement a process as a Python module
- Configuration file defines
 - Inputs / outputs
 - Synchronous / Asynchronous
 - Caching
 - Workflow

```
File Edit Options Buffers Tools Python Help
Example COWS-WPS Process
"""
# ... IMPORTS HIDDEN
class ExampleProcess(ProcessBase):
    def _executeProc(self, context):
        # Call standard _setup
        self._setup(context)
        # Now set status to started
        context.setStatus(STATUS.STARTED, 'Job is now running', 0)
        # Add output file
        outputNC = os.path.join(context.outputDir, "output.nc")
        outputPNG = os.path.join(context.outputDir, "output.png")
        # ... DO STUFF
        # ...
        nc_size = os.path.getsize(outputNC)
        png_size = os.path.getsize(outputPNG)
        # Add the stations file to the outputs
        self._addFileToOutputs(outputNC, 'NetCDF File', size = nc_size)
        self._addFileToOutputs(outputPNG, 'PNG File', size = png_size)
        process_support.finishProcess(context, self.fileSet,
                                     self.startTime, keep = True)
    def _validateInputs(self):
        """
        Runs specific checking of arguments and their compatibility.
        """
        pass
--:** example_process.py All L9 (Python)-----
```



UKCP09 Deployment



Generalised Service CEDA-WPS

WPS 1.0.0 Features

- A web service interface, using POST or GET.
- Asynchronous reporting and control of jobs.
- A defined XML interface for responses, including exceptions.
- A common format for passing arguments to the server.
- Job status interrogation.

Extensions in CEDA-WPS

- Add new processes as Python modules.
- Web-client to auto-generate process submission forms and interrogate current and previous jobs.
- Connection to a parallelised processing back-end.
- Output caching
- Inform users via e-mail when a job has completed (or failed).
- Integration with CEDA Security middleware.
- Estimate the job size and duration.
- Zip up output files



CEDA-WPS Web Client

The UI automatically generates submission forms for each process. This includes bounding box, date/time, float, integer and string types.

The screenshot shows a web form for configuring a bounding box. It includes a 'PlotTitle' field with the value 'The Name of The Plot' and a placeholder 'Please insert a value of type: string.'. Below it is a 'BoundingBox' section with a map of Africa and a red dashed bounding box. The map shows coordinates: West: -30.0, East: 30.0, North: 30.0, South: -30.0. The current bounding box is defined as '-30.0,-30.0,30.0,30.0'. A 'Submit' button is located at the bottom left of the bounding box section.

The screenshot shows the 'COWS Web Processing Service' interface. The title is 'COWS Web Processing Service' with the subtitle 'Distributed Processing in an OGC / Pylons framework'. Navigation links include Home, Capabilities, View, Submit, Jobs, Documentation, and Login. The main section is titled 'Define your inputs' and contains the following text: 'Please complete the form below to submit a request to the CEDA Web Processing Service. Note that some processes are restricted to registered users only. Click the 'Submit' button to submit your request.' Below this is a paragraph: 'This process includes arguments for which the possible values are dependent on calls to an external process. Please click the 'Update form' button at any time to update the argument options based on the selections you have made.' The form consists of three rows: 'Dataset' with a dropdown menu showing options like '/usr/local/cows_venv/local_dists/cows_wps/cows_wps/tests/data/cdml/eg_hadcm3.xml' and '/usr/local/cows_venv/local_dists/cows_wps/cows_wps/tests/data/cdml/eg_cdml.xml'; 'Variable' with a dropdown menu showing 'theta' and 'u'; and 'Description' with a dropdown menu showing 'long'. At the bottom, there are 'Update form' and 'Submit' buttons with instructions: 'Click to update the options above based on the selections you have made.' and 'Click this button when you are happy with your selections.'



Test Beds: Service Chaining WCS

COWS Web Processing Service
[Home](#) [Capabilities](#) [View](#) [Submit](#) [Jobs](#) [Documentation](#) [Login](#)

Define your inputs

Please complete the form below to submit a request to the CEDA Web Processing Service. Click the 'Submit' button to submit your request.

This process includes arguments for which the possible values are dependent on calls to update the argument options based on the selections you have made.

WCSEndPoint

Coverage

TimeStep

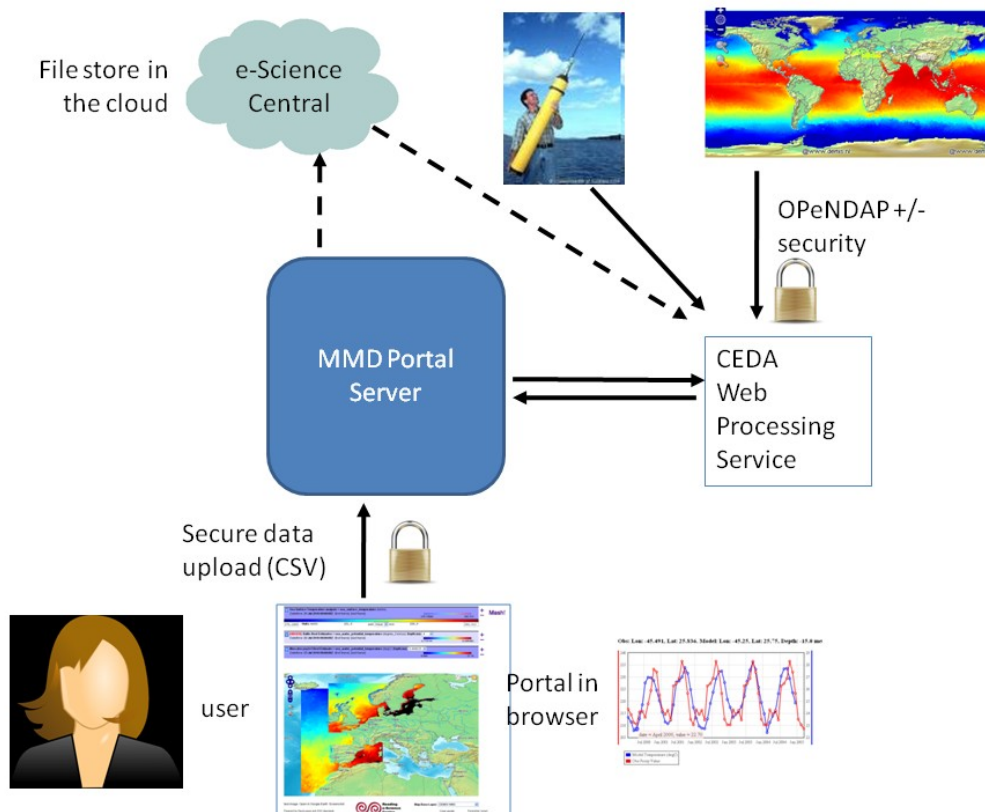
Click to update the options above based on the selections
 Click this button when you are happy with your selections

- COWS WPS UI generates a form for the “WCSWrapper” process.
- User can view the WCSEndPoint options.
- User clicks the “Update form” button to load the available “Coverages”
- The Coverages are extracted from a call to “GetCapabilities” at the WCSEndPoint.
- Further options are loaded.
- Once all selections have been made the user can click “Submit” to make a request to the WPS.
- The WPS calls the WCS “GetCoverage” method at the WCSEndPoint



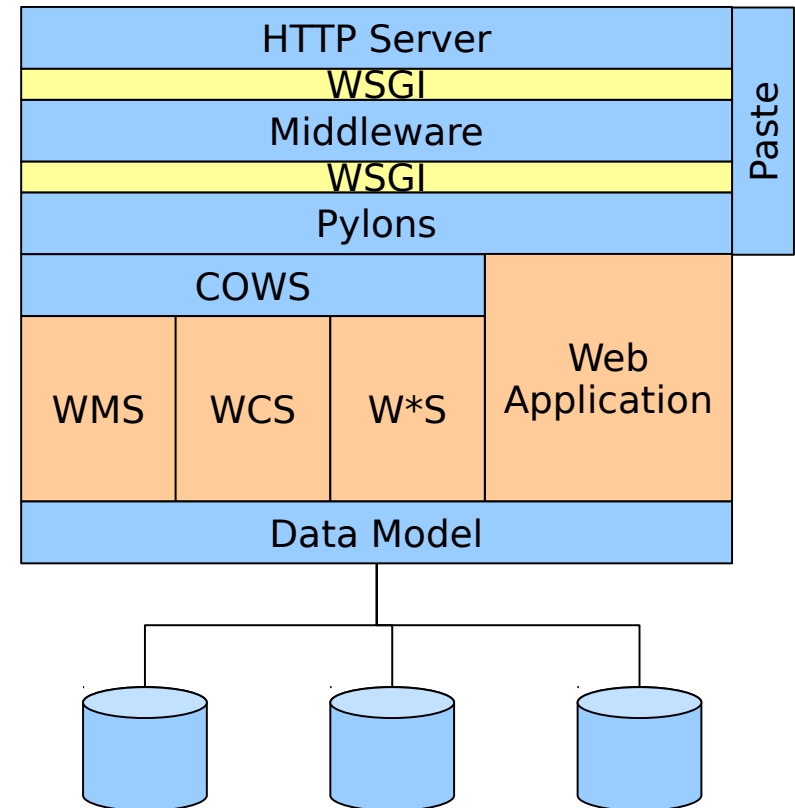
Test Beds: MashMyData

- Proof-of-concept web portal
- Scientists will be able to simultaneously visualize data from many sources, including their own uploaded data
- Scientists will be able to perform simple quantitative comparison calculations
- Data access will require Authentication and Authorisation



Test Beds: NDGSecurity

- Single sign-on with OpenID and PKI credentials
- WSGI filters for Authentication and Authorisation
- Centralised authorisation policy described by XACML (Oasis Standard)
- XACML policy generated from COWS-WPS configuration





Test Beds: MashMyData

OGC and OPeNDAP Services in a Secured Workflow

1) A user accesses the MashMyData Portal

2) They select some datasets for an intercomparison operation requiring a process run by the WPS.

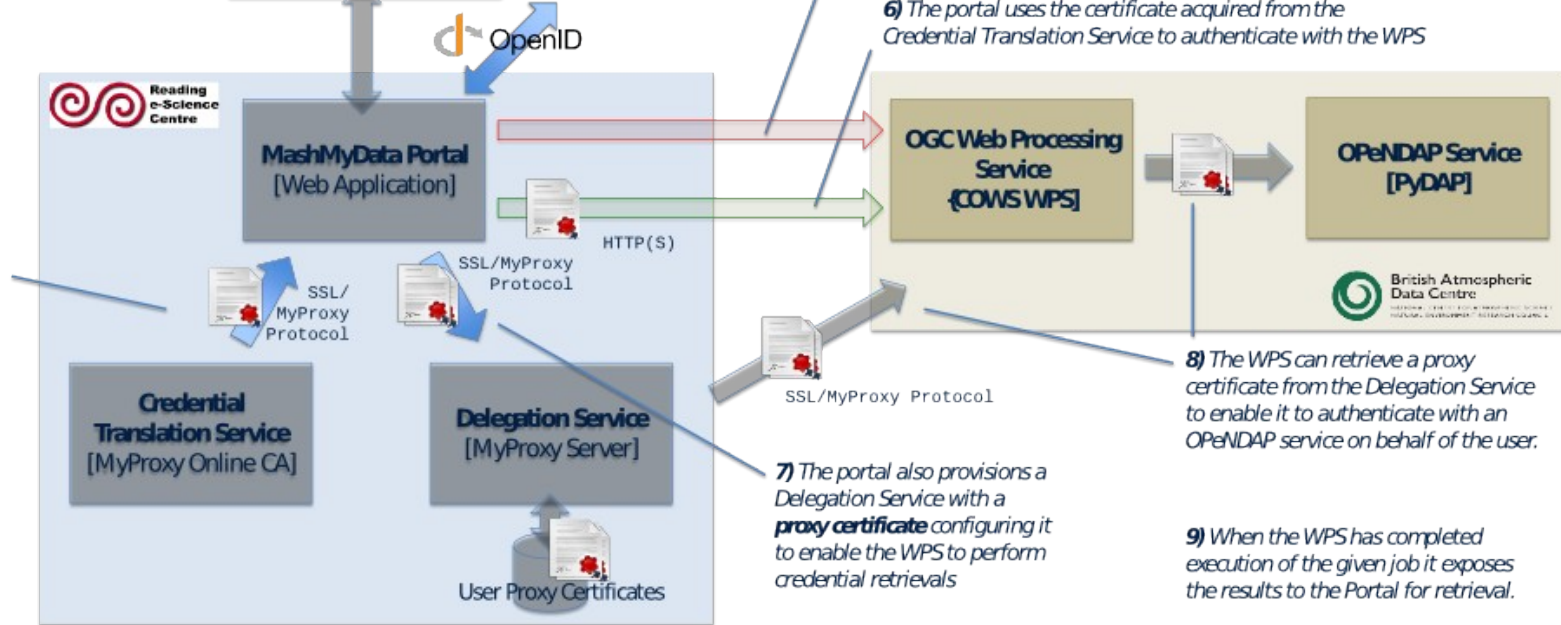


3) The portal requests the BADC's WPS execute this calculation on its behalf, but it responds indicating that access credentials are required to access secured datasets

4) The portal prompts the user to sign in with OpenID

6) The portal uses the certificate acquired from the Credential Translation Service to authenticate with the WPS

5) Once the user has signed in, the portal can translate their authenticated status by obtaining a short-lived certificate representing them from the Credential Translation Service - This credential can be used by the Portal to authenticate the user on their behalf with other services.



8) The WPS can retrieve a proxy certificate from the Delegation Service to enable it to authenticate with an OPeNDAP service on behalf of the user.

9) When the WPS has completed execution of the given job it exposes the results to the Portal for retrieval.



Test Beds: OPeNDAP

- WPS NetCDF Outputs available over OpeNDAP
- Implemented via an embedded PyDap server
- Interrogate NetCDF Metadata and Subset outputs
- Secured with NDGSecurity

Job Capabilities	download
Job Duration	1 second
Output Size	0.01 MB

OUTPUT FILES

The following file outputs are available from your job.

output.nc	[Download] [OPeNDAP: DAS DDS DODS]	0.01 MB
-----------	---	---------



Operational Service MIDAS

- New Service at BADC (beta)
- UK Met Office MIDAS database
- Land surface observations data from the Met Office station network
- daily and hourly weather measurements
- Extract by UK county of bounding box
- CSV output

The screenshot shows a web browser window with the URL `ceda-wps2.badc.rl.ac.uk/submit/form?proc_id=ExtractUKStationData`. The page contains a form with the following fields:

- StartDateTime**: . Instruction: Please insert a date/time field in the format `YYYY-MM-DDTHH:mm:ss` such as `2009-01-01T00:00:00`.
- EndDateTime**: . Instruction: Please insert a date/time field in the format `YYYY-MM-DDTHH:mm:ss` such as `2009-01-01T00:00:00`.
- OutputTimeChunk**: . Instruction: Please select an item from the list.
- BBox**: . Below the input is a map of the UK with a red bounding box. The map shows coordinates: West: [-6.8], South: [55.5], East: [-5.1], North: [57.5]. Instruction: This input is optional. Please select a valid bounding box within the following geographical extent: -12.0, 49.0, 3.0, 61.0.
- Counties**: . A dropdown menu lists counties: ABERDEENSHIRE, ALDERNEY, ANGUS, ANTRIM, ARGYLL (IN HIGHLAND REGION), ARGYLL (IN STRATHCLYDE REGION), ARGYLLSHIRE, ARMAGH, ASCENSION IS. Instruction: This input is optional. Multiple inputs are allowed. Please select one or more items from the list.



Future

- CDO Operators available as WPS Processes (ExArch Project)
- Run WPS Processes on the CMIP5 Archive
- Porting process execution backend to other schedulers (TORQUE)
- Improved WPS-1.0.0 compliance



Lessons Learned

- OGC Standards provide a useful blueprint for implementing operational SOA
- Slavish adherence to the standards can lead to limited payback for NWP/atmospheric science community
 - Process is slow
 - Premature standardisation
 - Focus is GIS not atmospheric science
- Our approach is to comply with commonly deployed versions whilst staying within striking distance of the latest standards





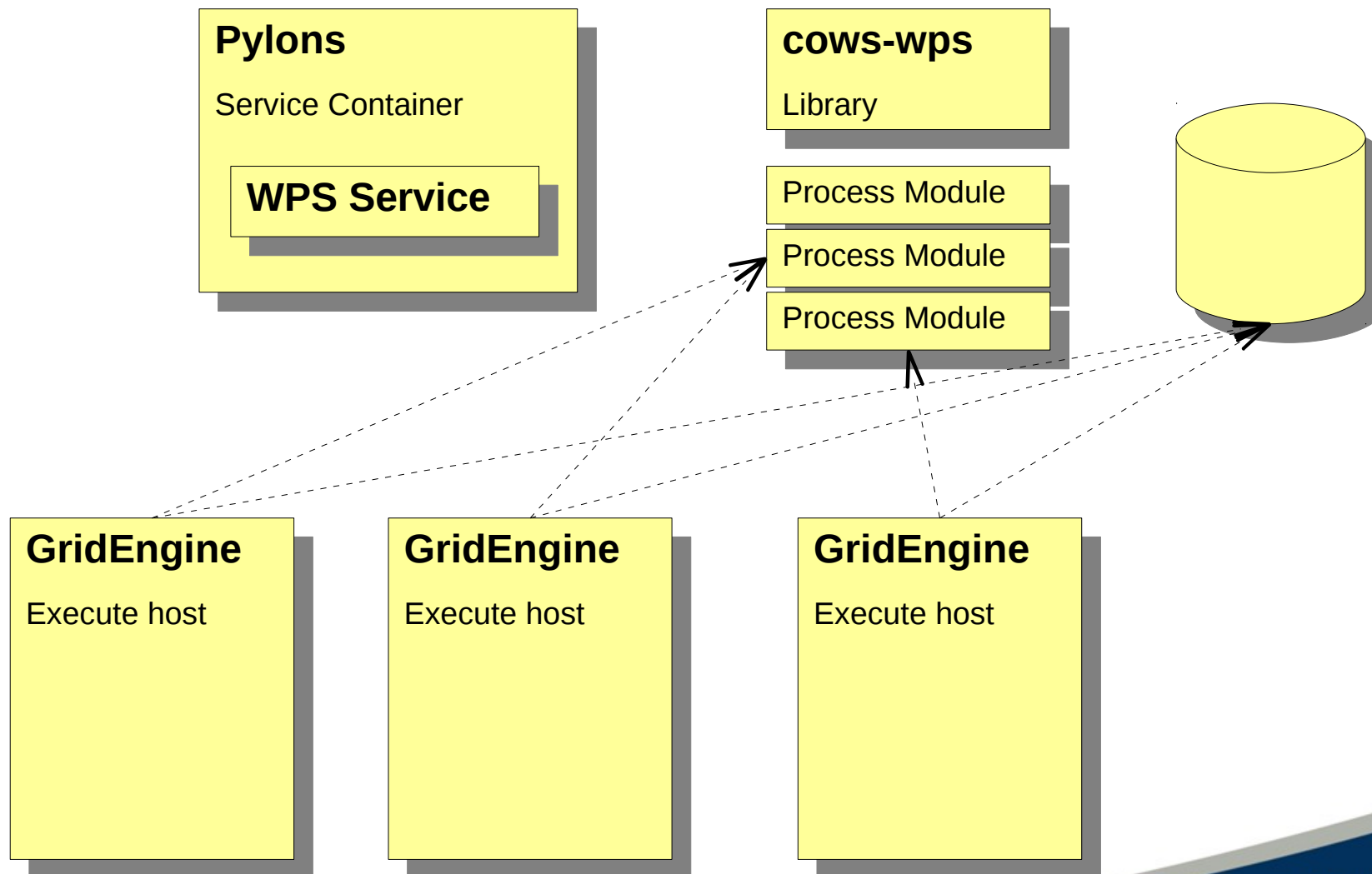
Thanks

Stephen Pascoe, Ag Stephens, Phil Kershaw

*<http://ceda-wps.badc.rl.ac.uk>
http://cows.badc.rl.ac.uk/cows_wps.html*



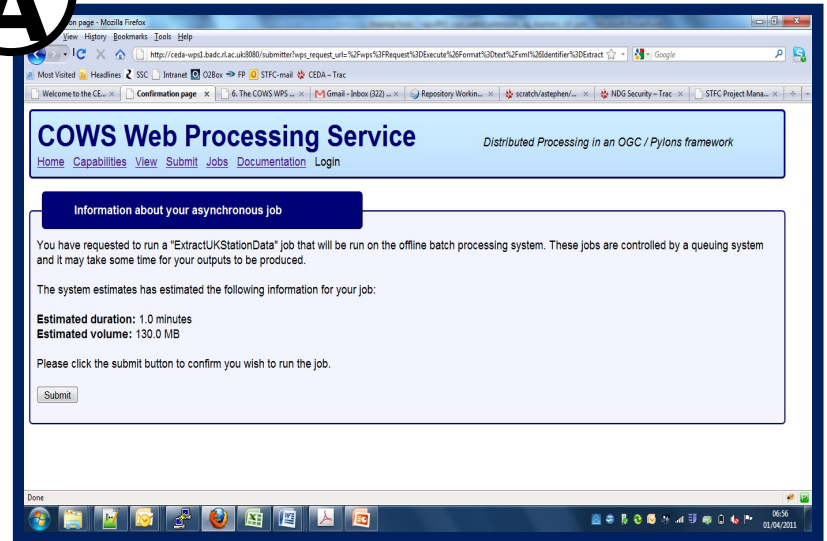
Architecture: COWS-WPS



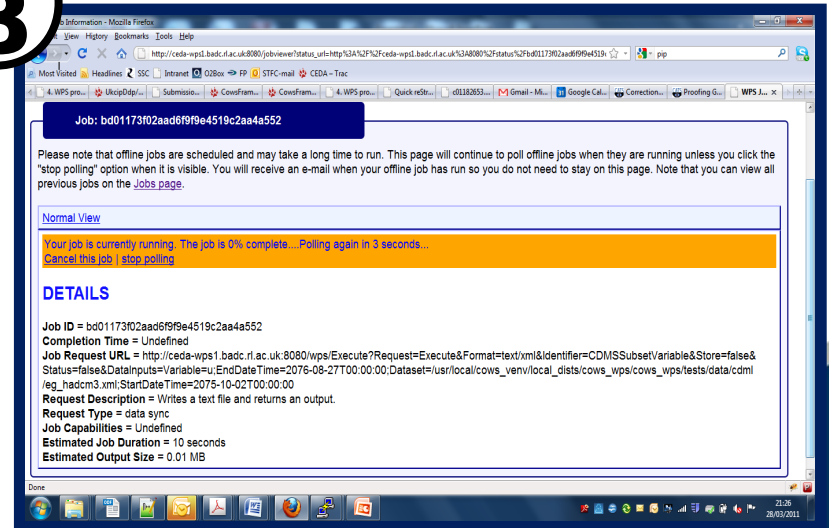
CEDA-WPS Workflow

- (A) "cost-only" call.
- (B) poll to update the job status.
- (C) retrieve job details and outputs
- (D) Previous jobs can be displayed
- (E) outputs can be downloaded.

A



B



CEDA-WPS Workflow

- (A) "cost-only" call.
- (B) poll to update the job status.
- (C) retrieve job details and outputs
- (D) Previous jobs can be displayed
- (E) outputs can be downloaded.



Request ID	Job type	User	Submission time	Job status	Action
dc8f1ebda35308d1af517159cea3c957	1	Ag.Stephens	2011-03-22 13:12:36	COMPLETED	View job info/outputs
39a3d570ed985722f5e5087710139e20	1	Ag.Stephens	2011-03-22 13:12:10	COMPLETED	View job info/outputs
58d702eb14076da7e44f83d1351f78ff	2	Ag.Stephens	2011-03-22 09:43:19	COMPLETED	View job info/outputs
13b386bcbe70336d6b16fa0f14f4277f	2	Ag.Stephens	2011-03-22 09:41:44	COMPLETED	View job info/outputs
cd1cb42905eca2878ebbe7a13e548038	2	Ag.Stephens	2011-03-22 09:41:12	COMPLETED	View job info/outputs
ddb88153a56916eccafbec3e5813b36	2	Ag.Stephens	2011-03-22 09:40:31	COMPLETED	View job info/outputs
12732bedad0df01b53cd09186cb9de9	2	Ag.Stephens	2011-03-22 09:39:17	COMPLETED	View job info/outputs
7902d8fb7b808efcbd40dea42e3cf61	2	Ag.Stephens	2011-03-22 09:37:33	COMPLETED	View job info/outputs
8ff421533b9048b02b9956d8749bb97a	2	Ag.Stephens	2011-03-22 09:33:13	FAILED	View job info/outputs
ab5cf53e02b1d97eab2cfb699a8f6b0e	2	Ag.Stephens	2011-03-22 00:32:50	FAILED	View job info/outputs
0a06100122430ea83b37f497cfd82085f	2	Ag.Stephens	2011-03-22 00:15:14	FAILED	View job info/outputs

