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

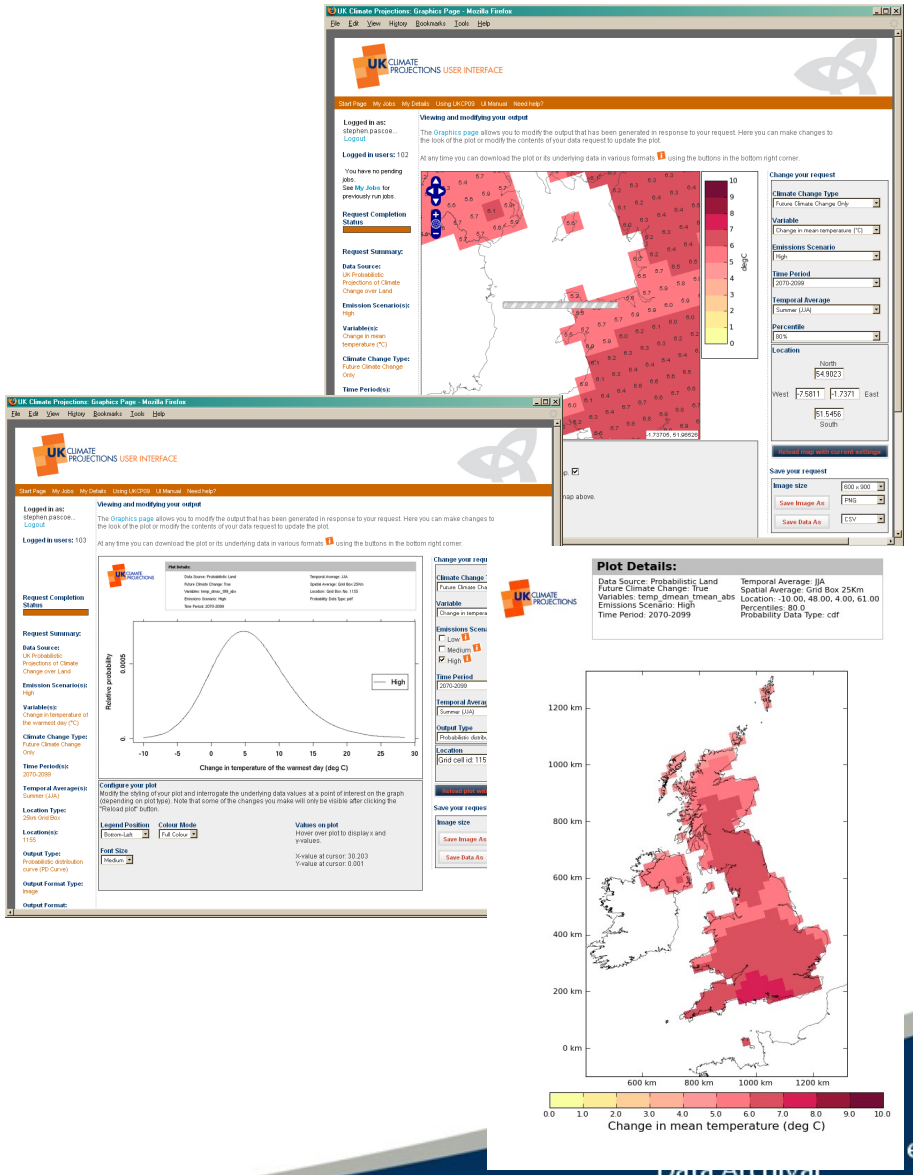
- History – first implementation and deployment
- Architecture
- Generalising first implementation
- Test Beds
- Operational Services

History
UKCIP09 User Interface

Requirements

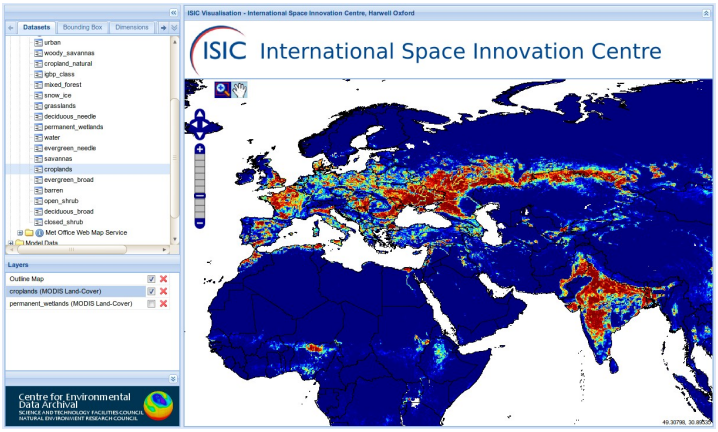
- Map and non-map plots
- Interactive and configurable
- Long-running tasks
  - Large data extractions
  - Secondary simulations (Weather Generator)
- Resilliant to variable load

Service Orientated Architecture



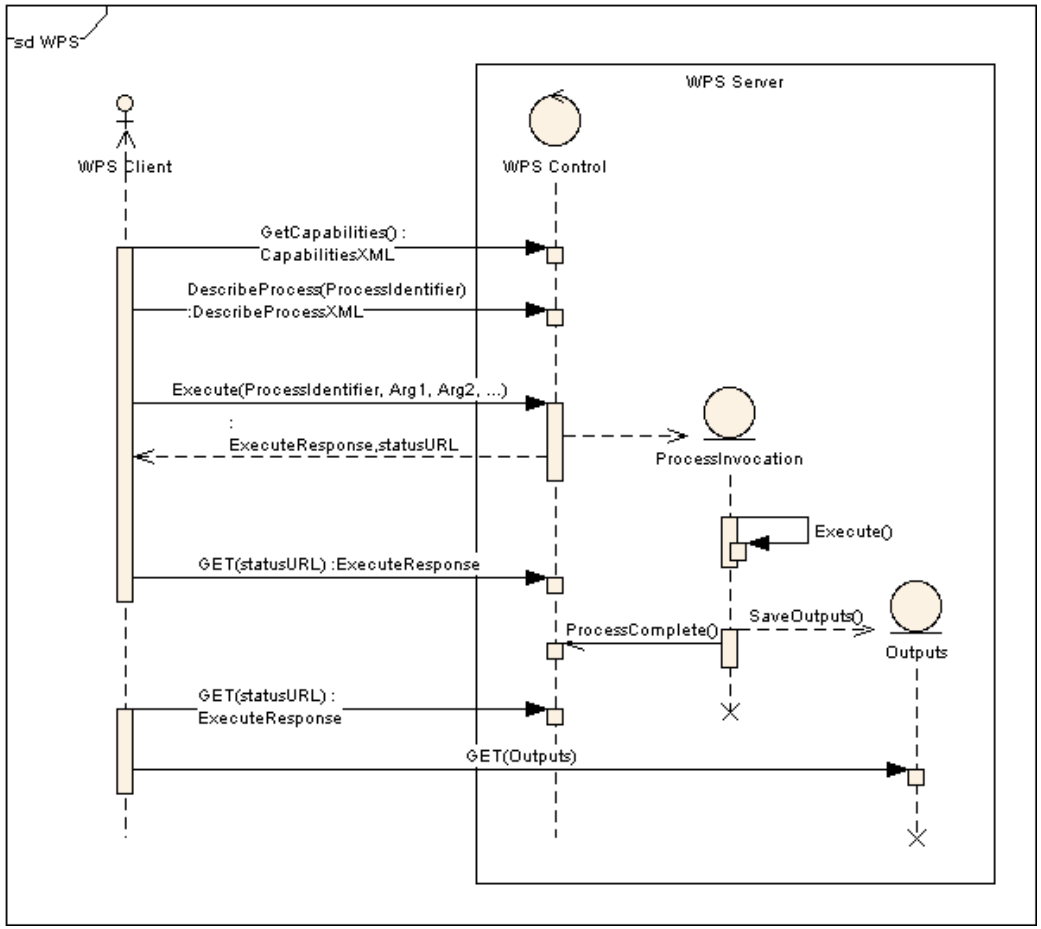
History
OGC Web Services at CEDA

- 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

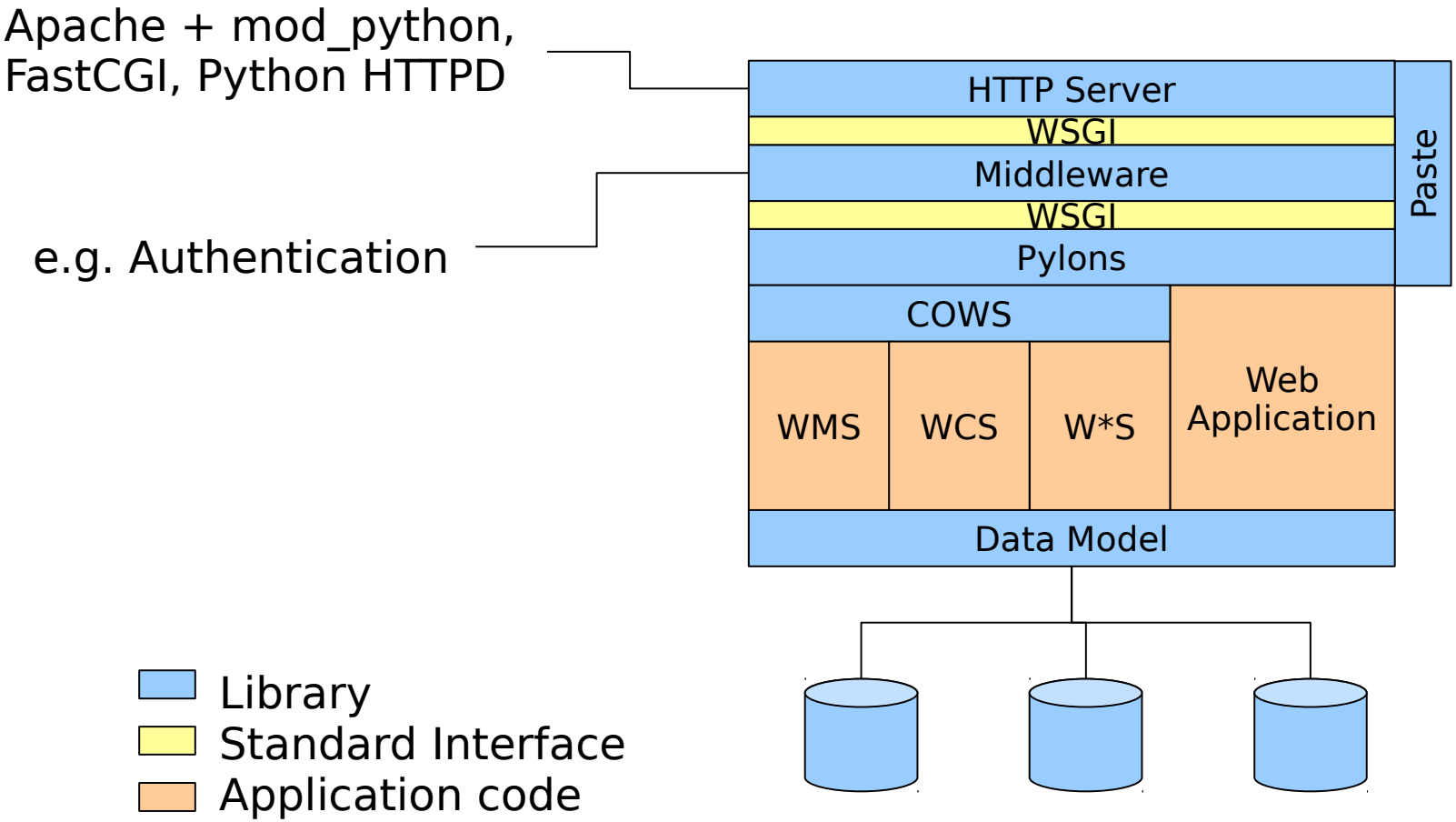
- GET DescribeProcess resource to discover process arguments
- POST to create a process execution resource (unique URL)
- GET to poll status of process execution
- Navigate to outputs when available



Architecture

CEDA OGC Web Services Framework: COWS

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



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

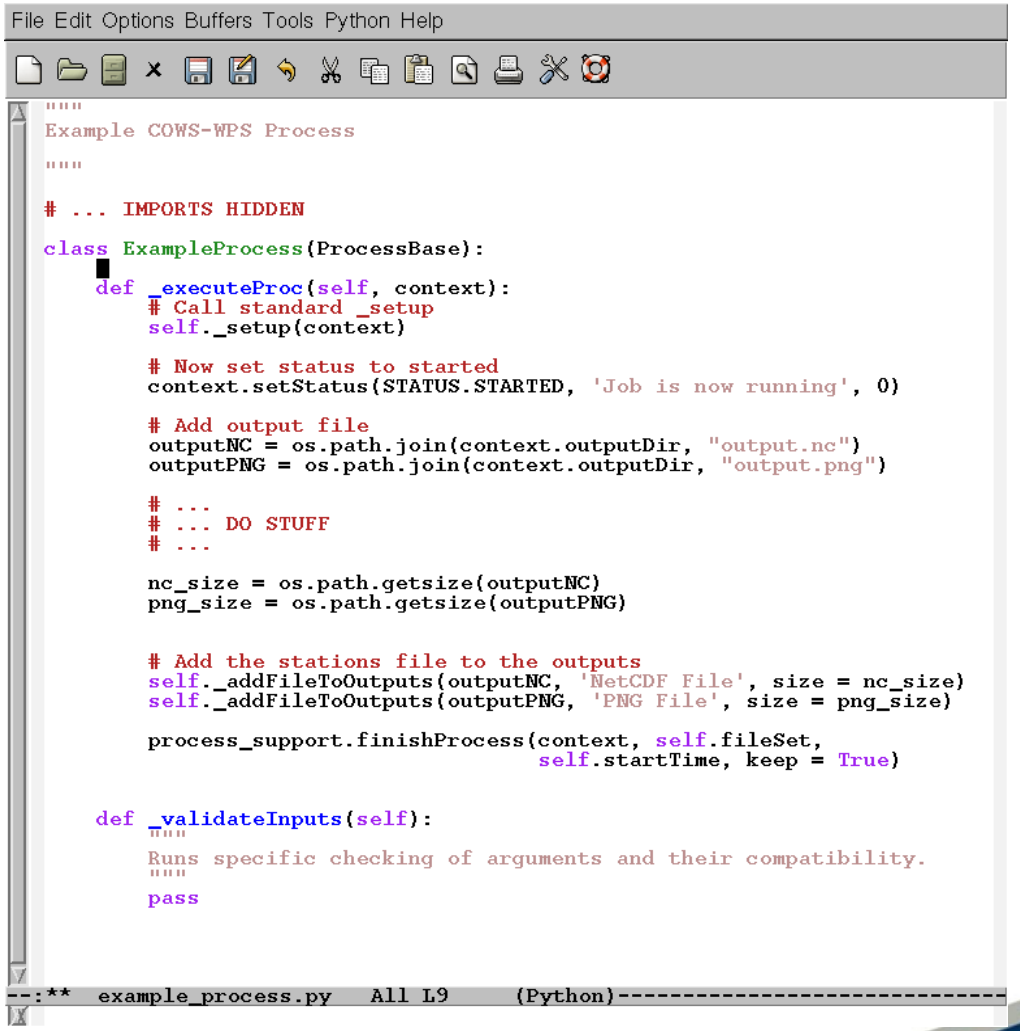
cows.csmlbackend

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

7

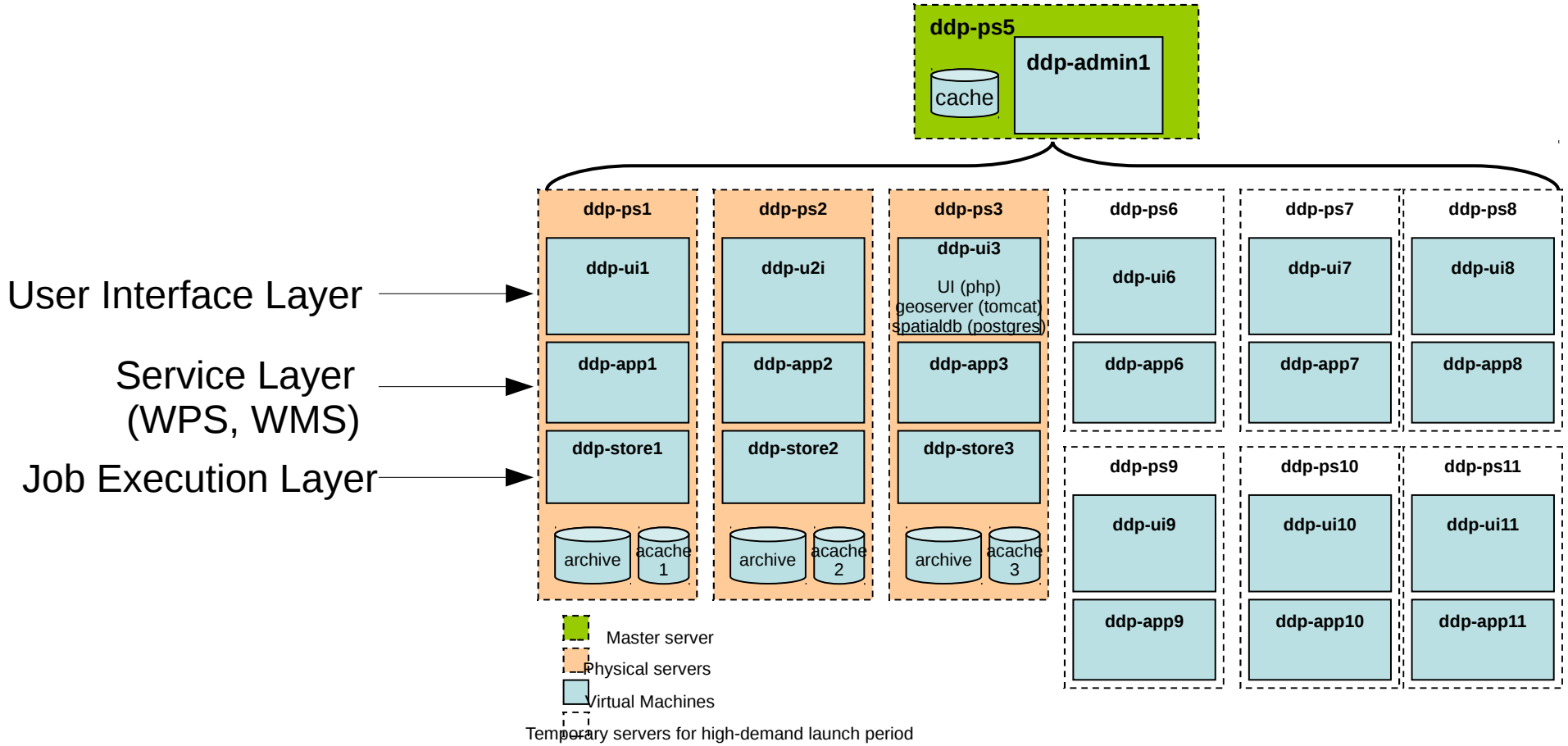
Architecture Process Modules

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



8

UKCP09 Deployment



Generalised Service CEDA-WPS

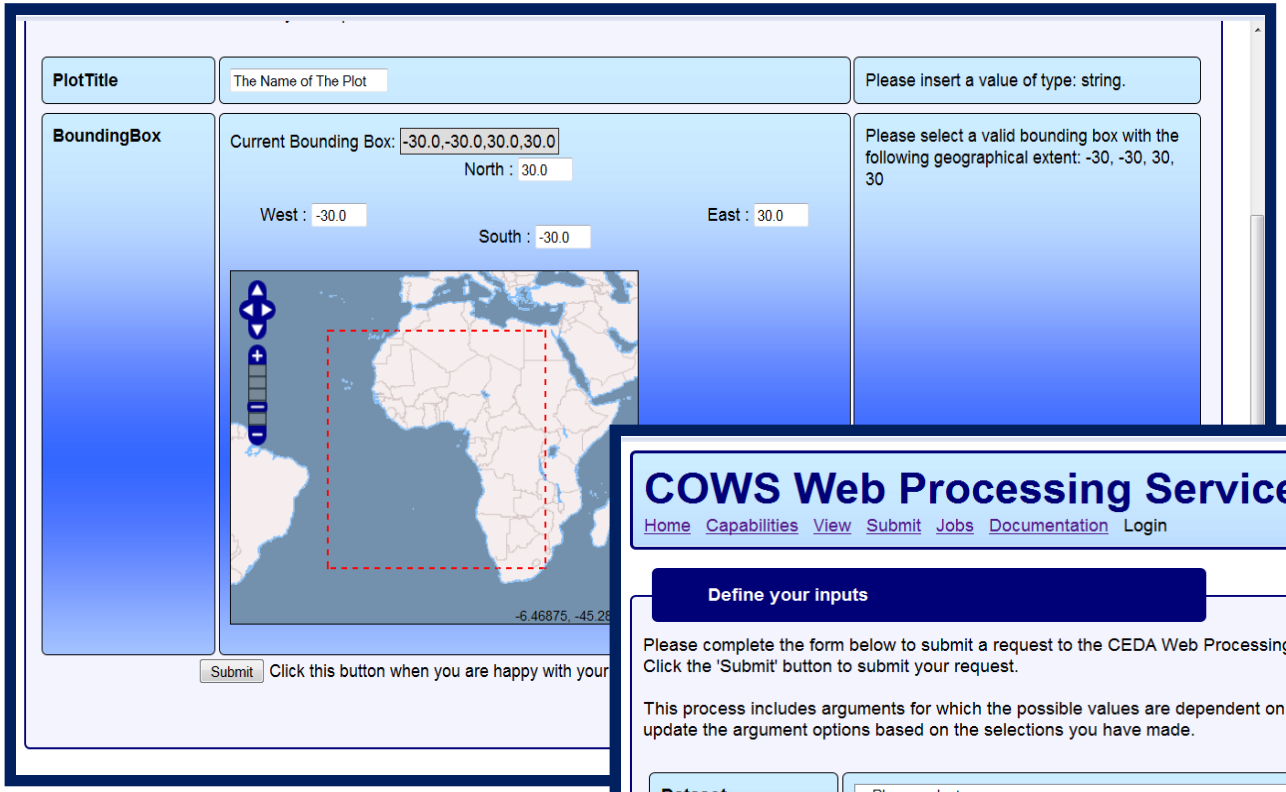
WPS 1.0.0 Features

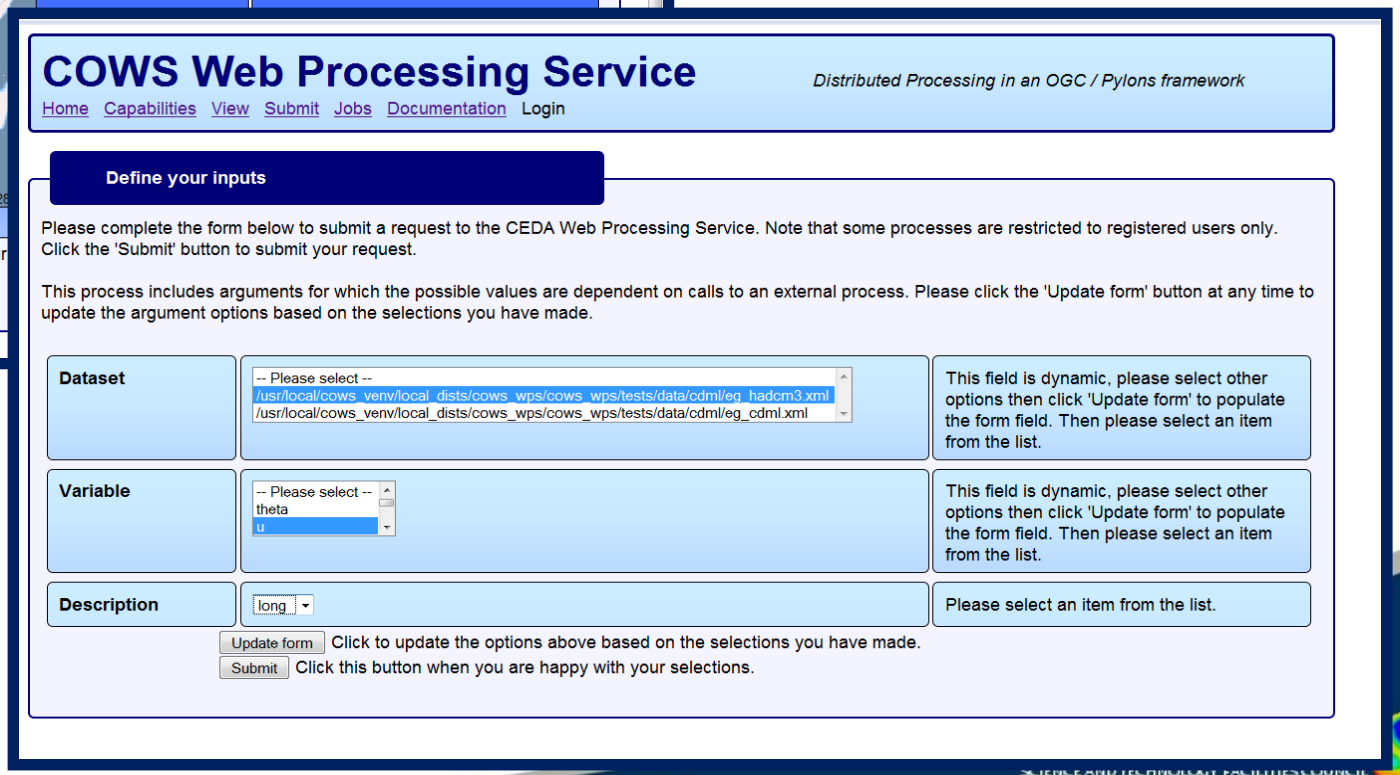
- A web service interface, using POST or GET.
- Asynchronous reporting and control of jobs.
- A defined XML interface for responses, including exceptions.
- A common format for passing arguments to the server.
- Job status interrogation.

Exensions 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

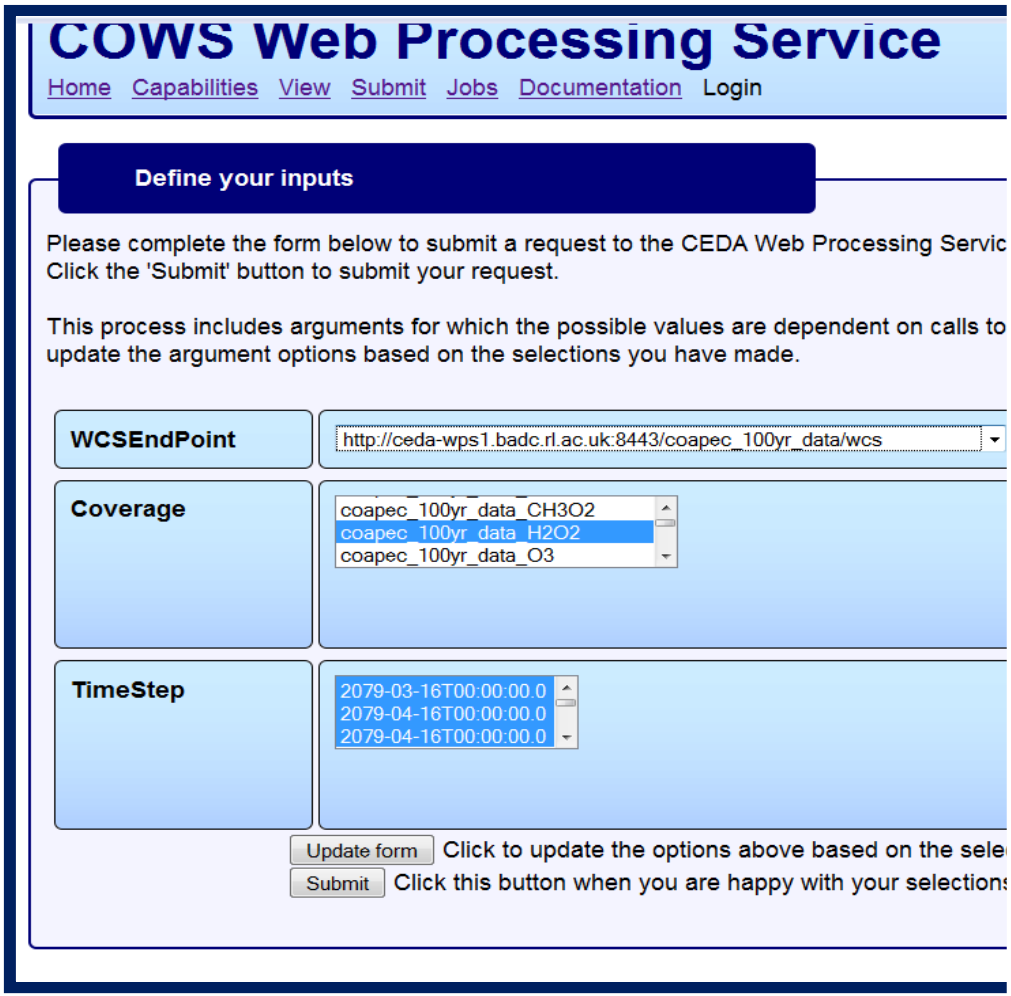
CEDA-WPS Web Client





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

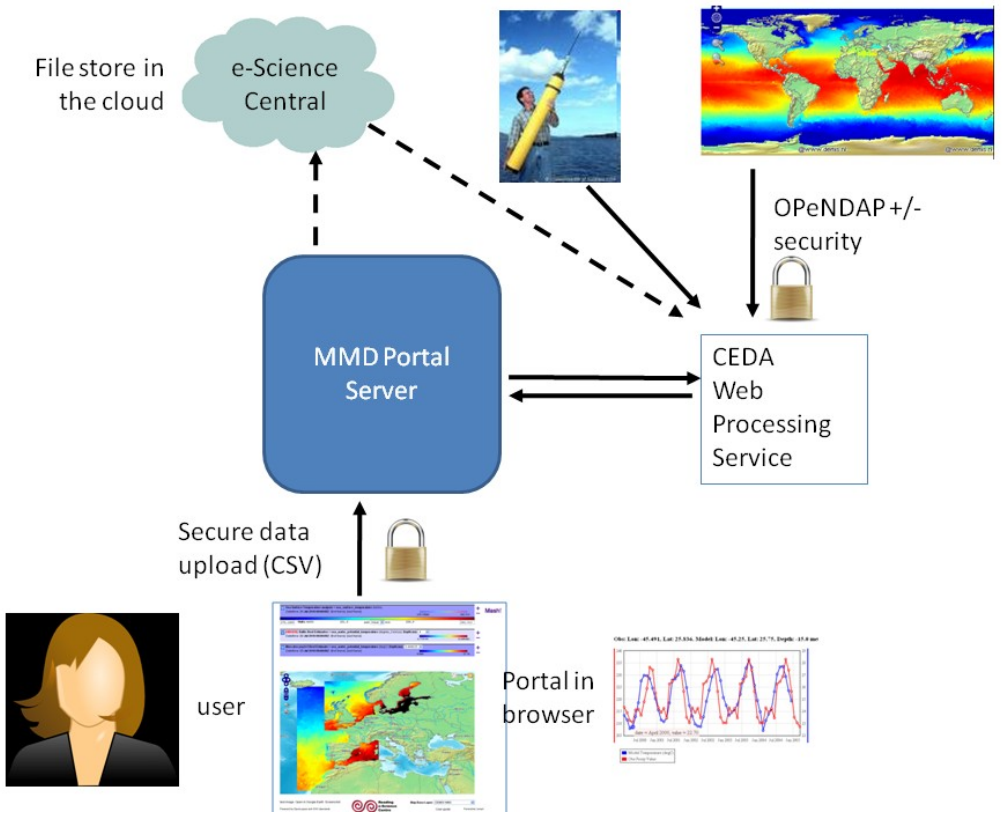
Test Beds: Service Chaining WCS



- 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

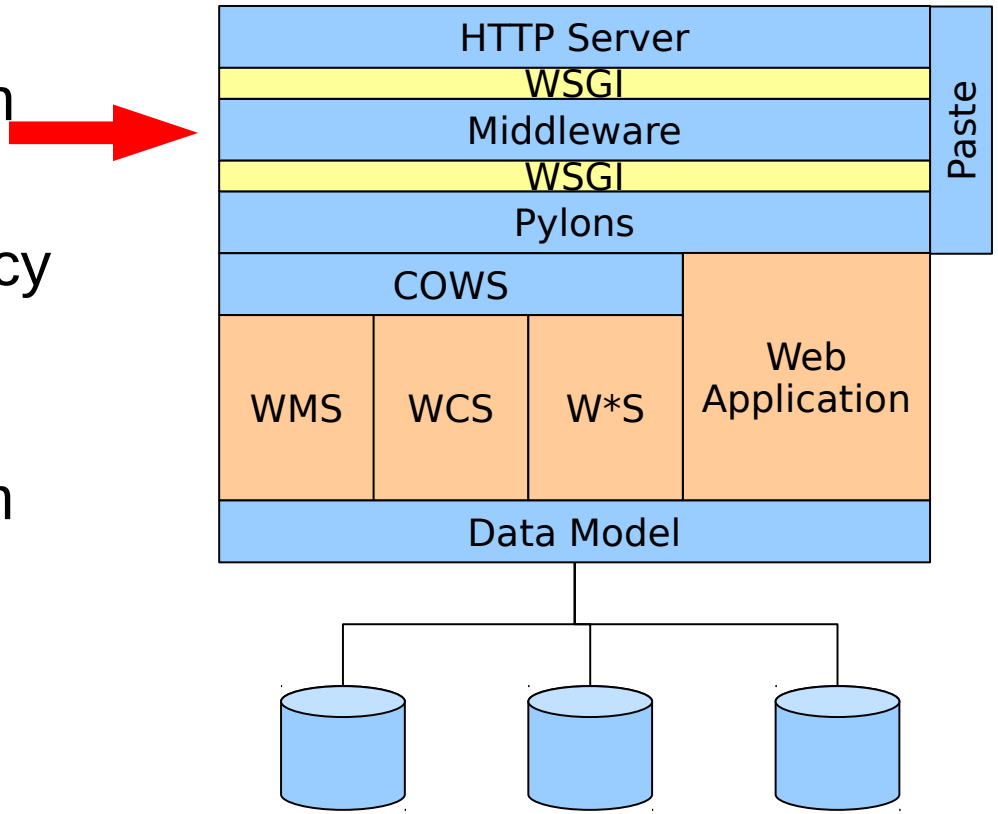
Test Beds: MashMyData

- 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

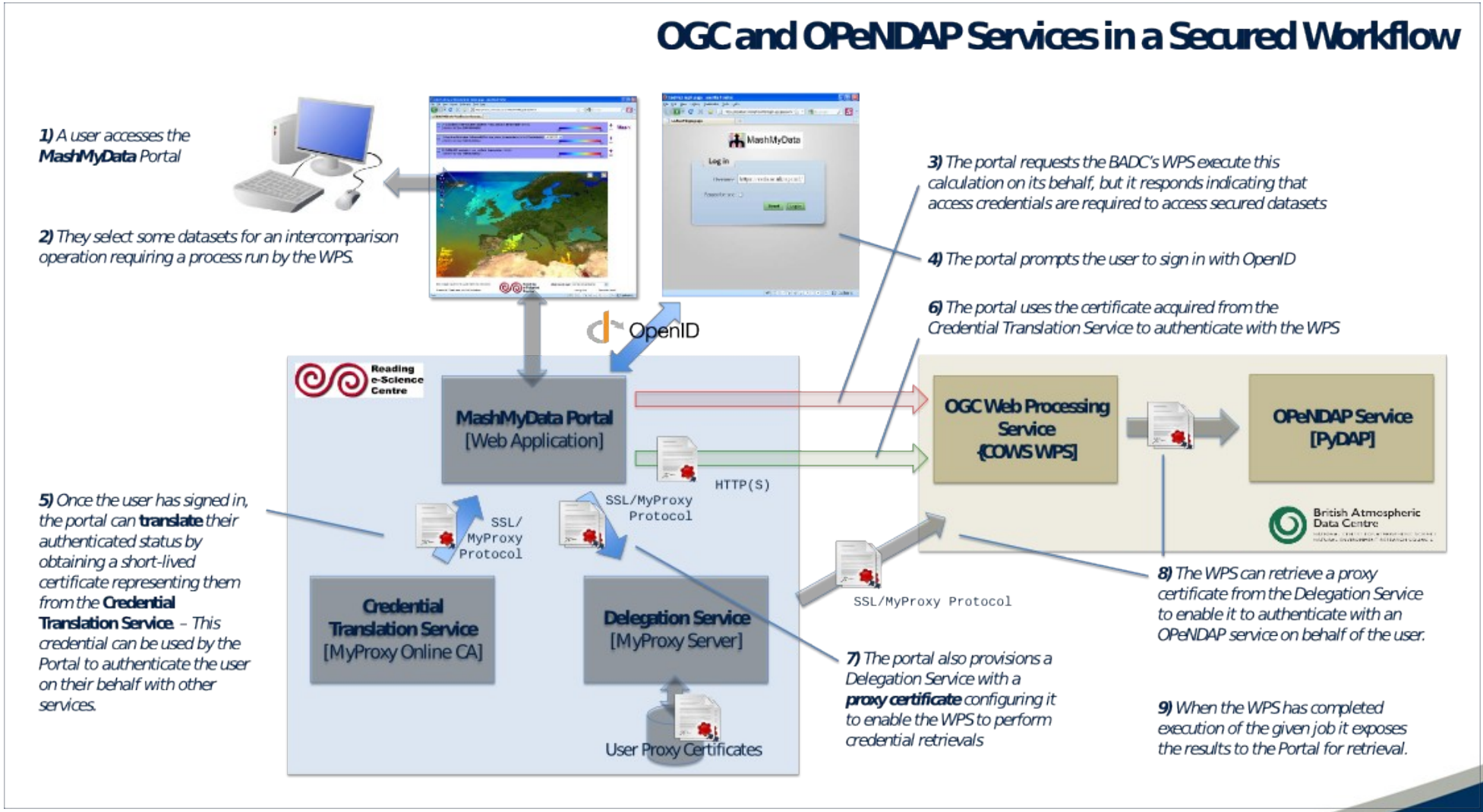


Test Beds: NDGSecurity

- 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

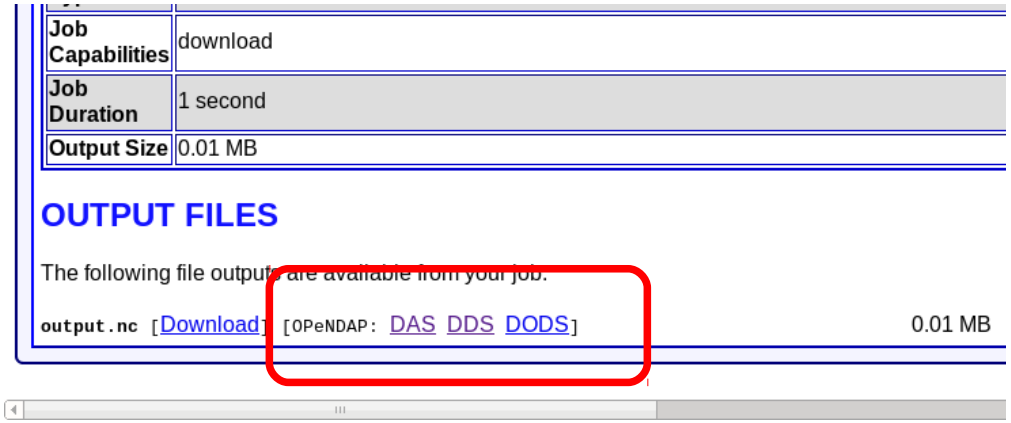


Test Beds: MashMyData



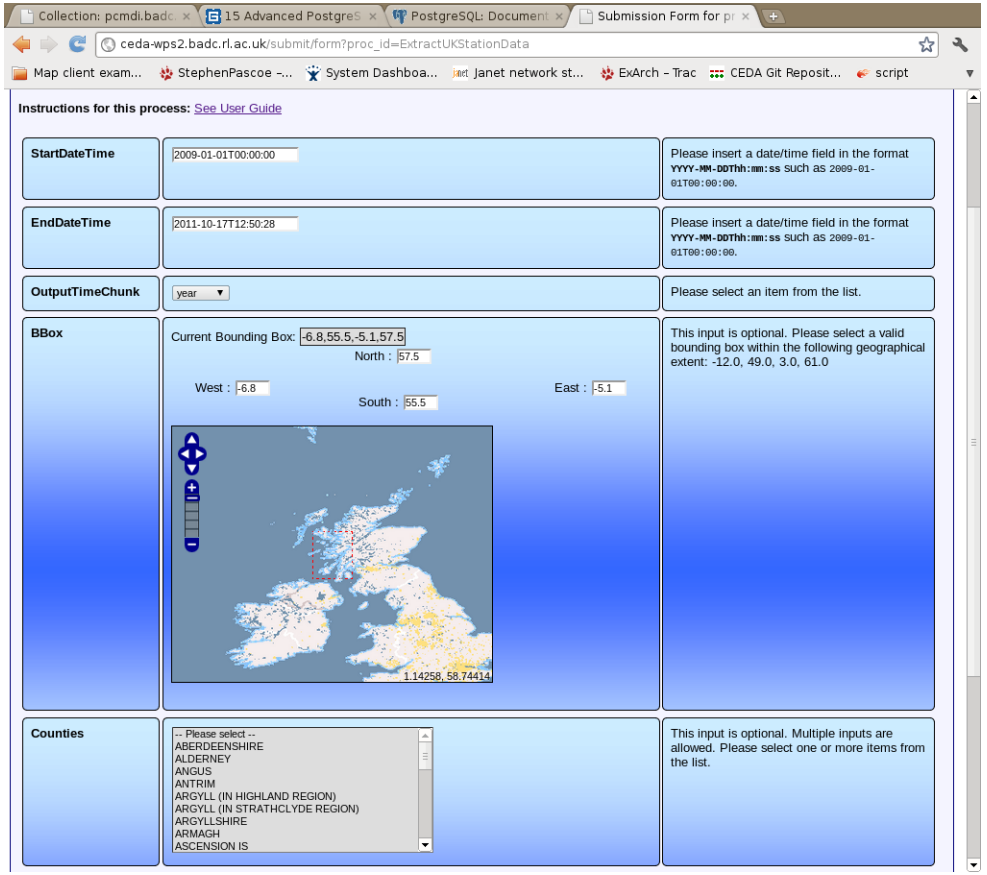
Test Beds: OPeNDAP

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



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



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

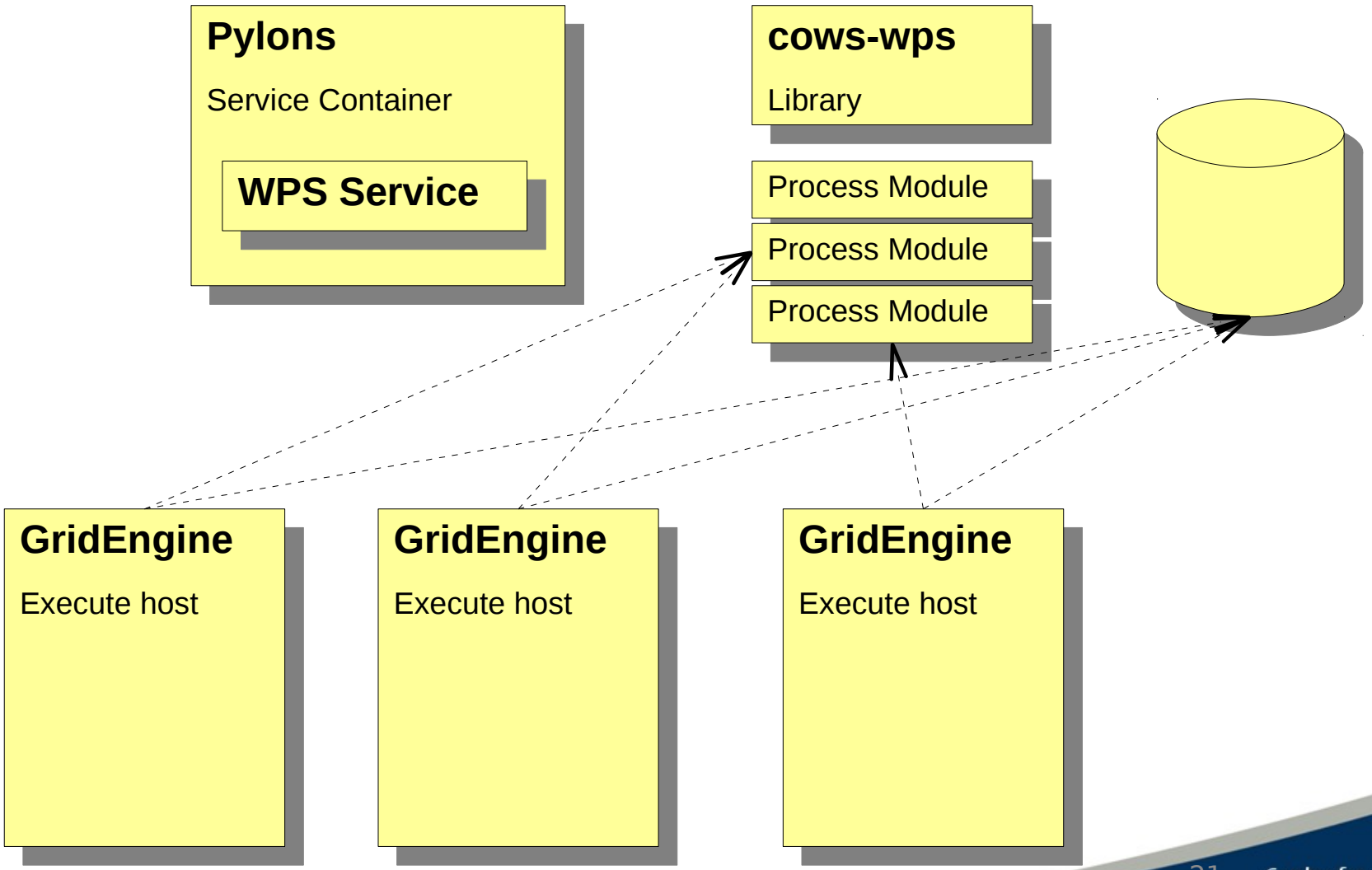
- OGC Standards provide a useful blueprint for implementing operational SOA
- Slavish adherance to the standards can lead to limited payback for NWP/atmospheric science community
  - Process is slow
  - Premature standardisation
  - Focus is GIS not atmospheric science
- Our approach is to comply with 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

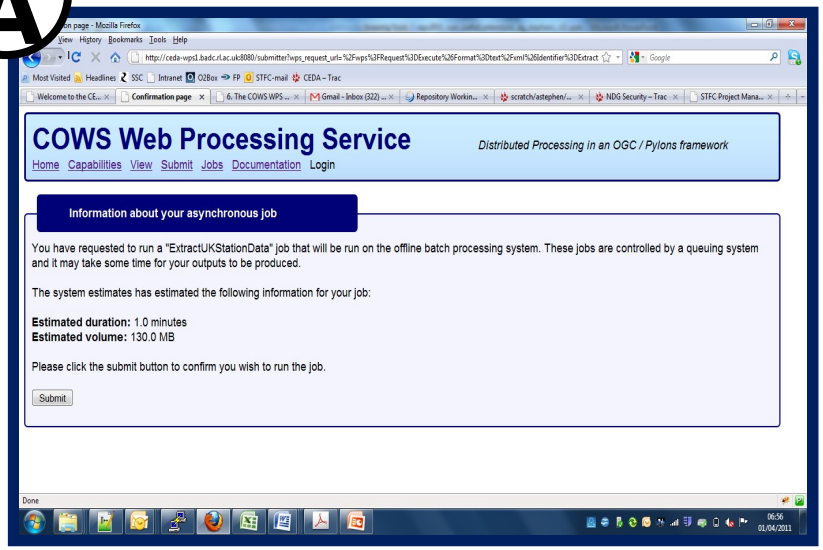
Architecture: COWS-WPS



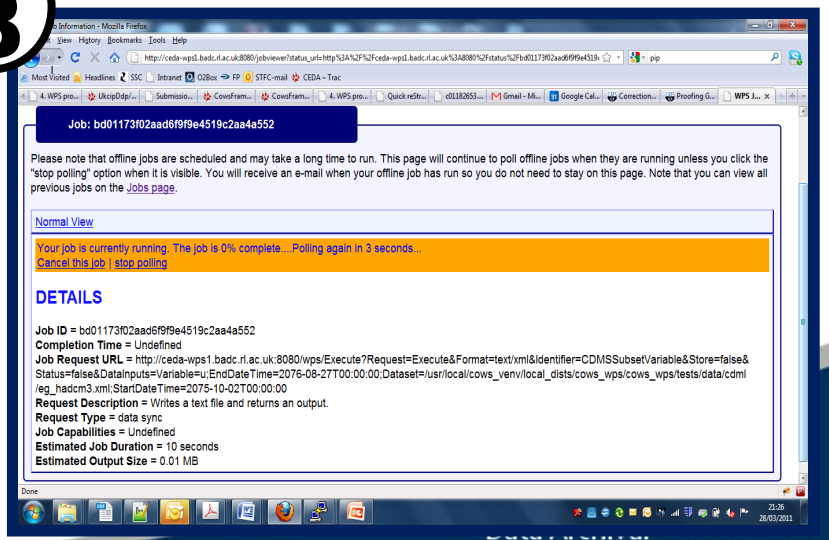
CEDA-WPS Workflow

(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



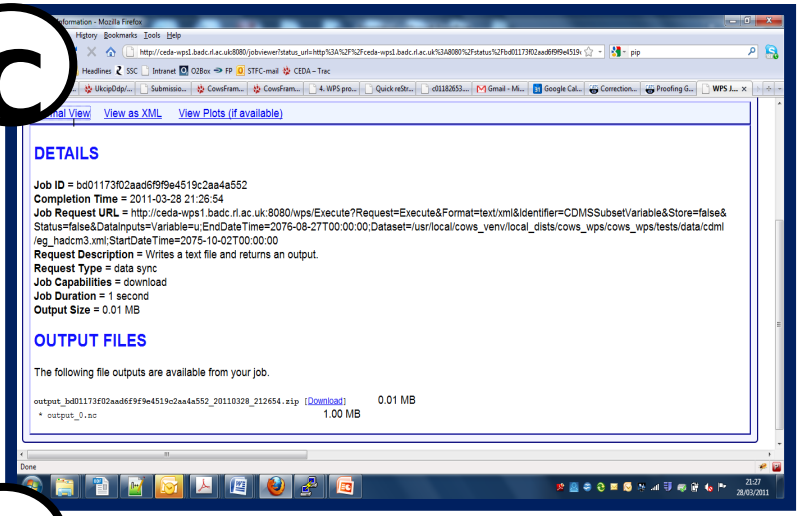
B



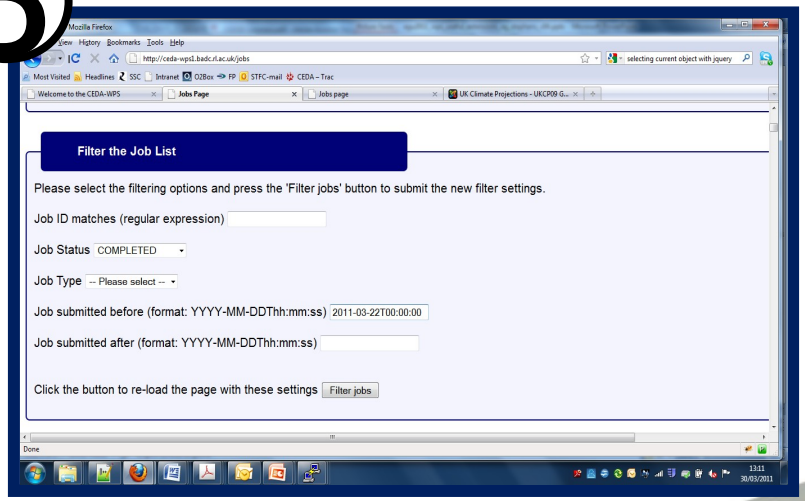
CEDA-WPS Workflow

(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.

C



D



E