

C. Ehbrecht¹, J. Meyer², S. Kindermann¹, T. Kipp¹

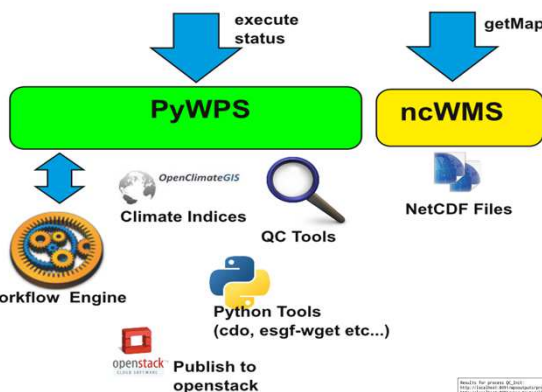
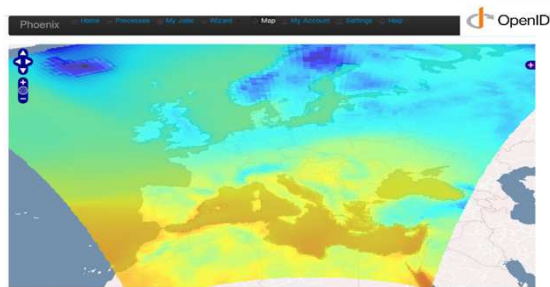
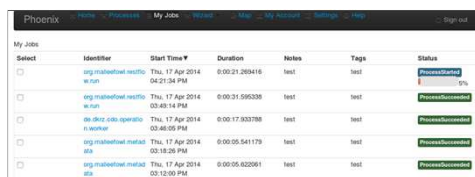
1) Deutsches Klimarechenzentrum DKRZ, Hamburg 2) Karlsruher Institut für Technologie KIT, Karlsruhe

Motivation


- Support of end users in data life cycle management activities involving multiple distributed data centers
- Data management activities are composed of a set of basic operations
- Services are accessible to end users via GUIs as well as client tools
- Services are discoverable and expose standardized interface descriptions

The Approach

- Basic data life cycle support operations are exposed as services by data centers
- Services can be composed to build up complex data management activities (workflows)
- A web framework supports users in invoking, composing and executing services / workflows
- Services are exposed conforming to the OGC WPS standard

Select	Identifier	Start Time	Duration	Notes	Tags	Status
<input type="checkbox"/>	org.individual.metad	Thu, 17 Apr 2014 00:21:20Z	0:00:21.209416	test	test	Finished
<input type="checkbox"/>	org.individual.metad	Thu, 17 Apr 2014 00:31:59Z	0:00:31.595338	test	test	Finished
<input type="checkbox"/>	org.individual.metad	Thu, 17 Apr 2014 00:49:14 PM	0:00:17.303786	test	test	Finished
<input type="checkbox"/>	org.individual.metad	Thu, 17 Apr 2014 03:18:26 PM	0:00:05.541179	test	test	Finished
<input type="checkbox"/>	org.individual.metad	Thu, 17 Apr 2014 03:12:00 PM	0:00:05.022061	test	test	Finished



```

Results for process @_init:
Step: 1: Download netcdf files from iRods with rsync
Step: 2: QC Init: validates file/directory names
Step: 3: QC Check: runs quality check on contents of netcdf files
Step: 4: QC Evaluate: evaluates the quality check, generates metadata and quality files and generates PIDs
Step: 5: QC Publish: publishes generated metadata and quality files to openstack cloud
Results for process @_publish:
Step: 1: Publish netcdf files to openstack cloud
Step: 2: Publish metadata files to openstack cloud
Results for process @_evaluate:
Step: 1: Evaluate netcdf files
Step: 2: Generate metadata and quality files
Step: 3: Generate PIDs
Results for process @_check:
Step: 1: Check netcdf files
Step: 2: Check metadata files
Step: 3: Check quality files
Results for process @_init:
Step: 1: Download netcdf files from iRods with rsync
Step: 2: QC Init: validates file/directory names
Step: 3: QC Check: runs quality check on contents of netcdf files
Step: 4: QC Evaluate: evaluates the quality check, generates metadata and quality files and generates PIDs
Step: 5: QC Publish: publishes generated metadata and quality files to openstack cloud
Results for process @_publish:
Step: 1: Publish netcdf files to openstack cloud
Step: 2: Publish metadata files to openstack cloud
Results for process @_evaluate:
Step: 1: Evaluate netcdf files
Step: 2: Generate metadata and quality files
Step: 3: Generate PIDs
Results for process @_check:
Step: 1: Check netcdf files
Step: 2: Check metadata files
Step: 3: Check quality files
  
```

A Concrete Use Case

Data generation and publication for regional climate model data

Data generation:

- Researcher copies initialization data from KIT to DKRZ
- Researcher at KIT runs climate model at DKRZ
- Researcher collects output at DKRZ storage resources

Data postprocessing:

- Data scientist postprocesses data (at DKRZ)
- Storage of data in standard format/layout

Quality Control of Data:

- Data quality check software is run at Data Center
- Quality check results are communicated for review
- Persistent identifiers (PIDs) are assigned to successfully checked data

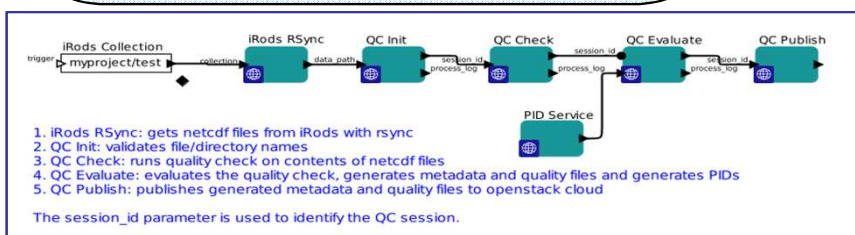
Data Publication:

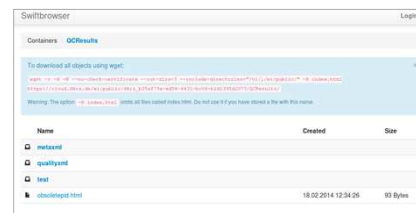
- Data is published to a worldwide data federation portal at DKRZ
- Data is visible and accessible via any of the ESGF portals worldwide

Data Archival:

- Important parts of published data are archived

- Close collaboration necessary between Researcher and Data Scientist (coming from different institutions)
- Large variety of services and tools are involved in this collaboration





Name	Created	Size
metad		
quality		
test		
observed.html	18.02.2014 12:34:26	53 Bytes

The data services framework ClimDaPs

Components of ClimDaPs:

- Pyramid WPS WebClient with OpenLayers
- PyWPS Web Processing Service
- ncWMS Web Mapping Service for NetCDF files
- Restflow Workflow Engine

Authentication with OpenID
Workflow Engine called as WPS Process/ Workflow calls other WPS Processes

Prototype available at <http://mouflon.dkrz.de>



Services available via ClimDaPs by now

Low level data/metadata operations:

- ISO 19139 and ESGF solr metadata generation
- iRods based data transfer and synchronization
- Handle system based persistent identifier assignment and retrieval for single data products and data aggregations
- Publication of results on an OpenStack storage cloud

Higher level operations:

- Cordex data quality checking
- CDO based climate data processing
- Calculation of climate indices
- Calculation of species distribution

Complex workflows:

- Cordex data quality control and quality result publication

Open Issues

- Security infrastructure
- Discovery and semantic service description
- End user workshops

Project / Funding context

- Current development and application to concrete use cases is funded within the Helmholtz LSDMA project
- Work on a scalable WPS based data quality processing infrastructure is supported by the ExArch G8 project.
- Development and consistent deployment of WPS services at European data centers is supported by the IS-ENES2 FP7 project
- Early ClimDaPs service and framework developments were supported by the German C3-Grid-INAD project