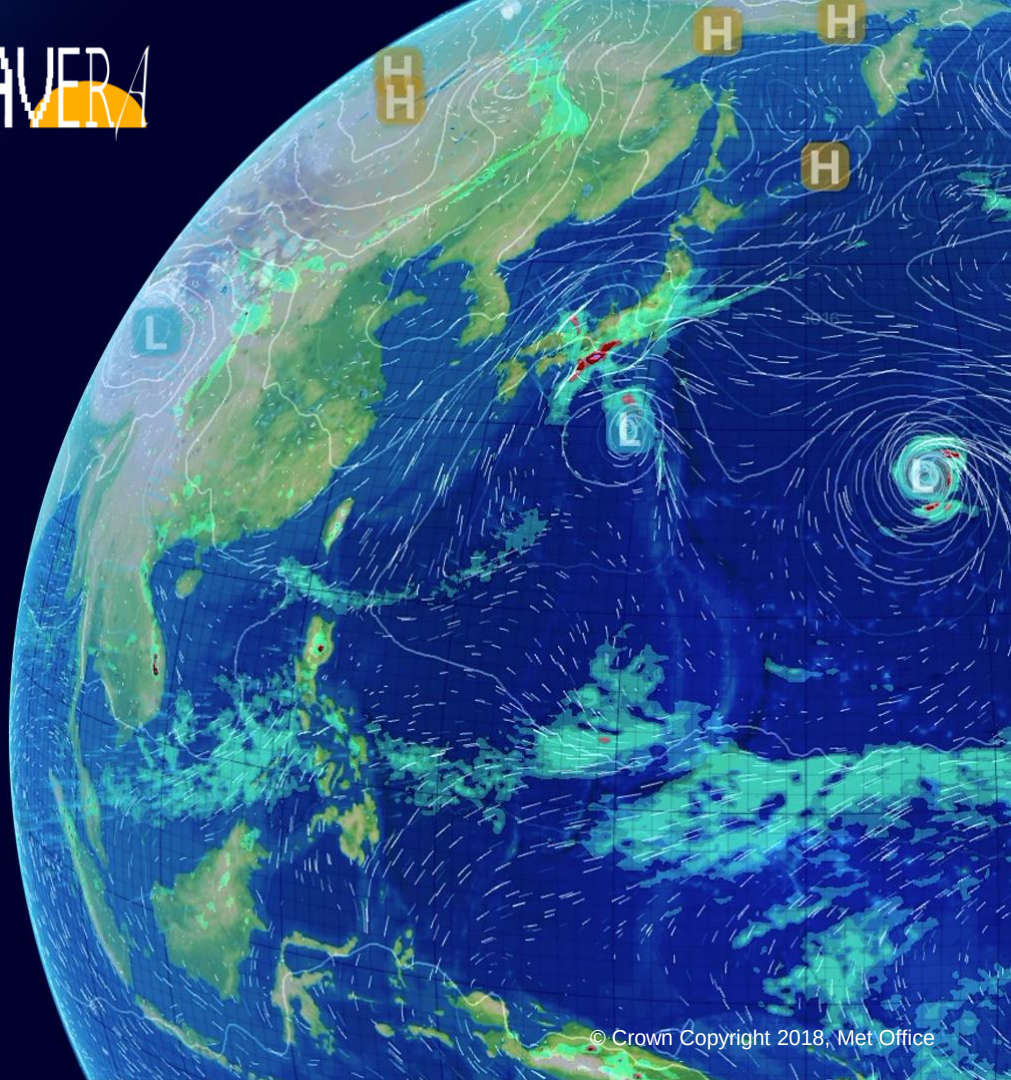
Workflow tools for PRIMAVERA at the Met Office and JASMIN

Jon Seddon

jon.seddon@metoffice.gov.uk

www.primavera-h2020.eu




Met Office
Hadley Centre


PRIMAVERA


AWI


BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación


CERFACS


cmcc


DKRZ
DEUTSCHES
KLIMARECHENZENTRUM


ECMWF


ISAC
Consiglio
Nazionale delle
Ricerche


Royal Netherlands
Meteorological Institute
Ministry of Infrastructure
and the Environment


Met Office


Max-Planck-Institut
für Meteorologie


NERC
SCIENCE OF THE
ENVIRONMENT


predictia


SMHI


Science & Technology
Facilities Council


UNIVERSITY OF LEEDS


UCL
Université
catholique
de Louvain


UNIVERSITY OF
OXFORD


University of
Reading


Stockholm
University

Met Office Hadley Centre

PRIMAVERA

PRocess-based sIMulation: AdVances in high resolution modelling And European climate Risk Assessment

Met Office Hadley Centre

PRIMAVERA

The data challenge

- 2.5 petabytes of model output data from seven centres

- Scientists from 19 institutions across Europe are analysing the data

The data challenge

- 2.5 petabytes of model output data from seven centres

- Scientists from 19 institutions across Europe are analysing the data

The solution:

- Take the analysis to the data

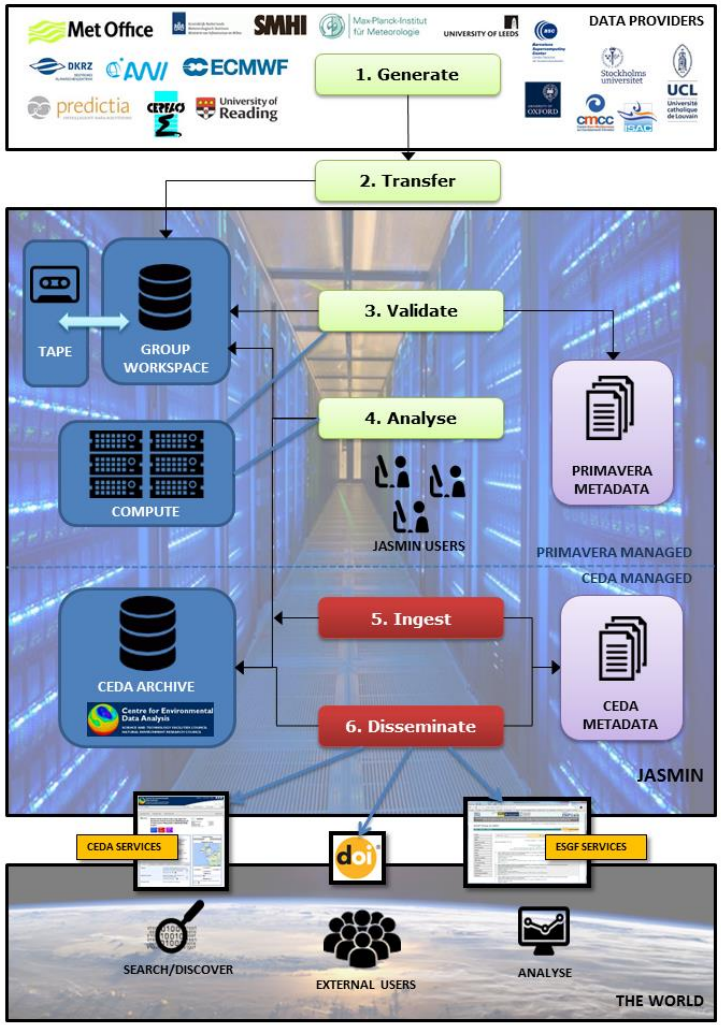
JASMIN

- 38 petabytes of disk storage plus tape
- 4000 compute cores on LOTUS plus interactive data analysis servers
- High-performance internal network and connections to the Internet

- PRIMAVERA has 440 terabytes of disk storage available to it

Met Office Hadley Centre

PRIMAVERA



Met Office use of Workflow Tools in PRIMAVERA

We use Rose and Cylc for:

- Climate model simulations on the HPCs (XCS/NEXCS and ARCHER)
- “CMORization” of PP files to netCDF
- Submission of data to ESGF

We also have the Data Management Tool web interface to the database.

Variables Received

Secure | https://prima-dm.ceda.ac.uk/received_data/?project=&institute=&climate_model=&experiment=highresSST-present&mip_table=...

PRIMAVERA Data Management Tool

Home | Views | Login

Variables Received

The following data has been received:

Project | Institute | Climate Model | highresSST-present | day

Variant Label | tas | Variable Name | Clear | Filter

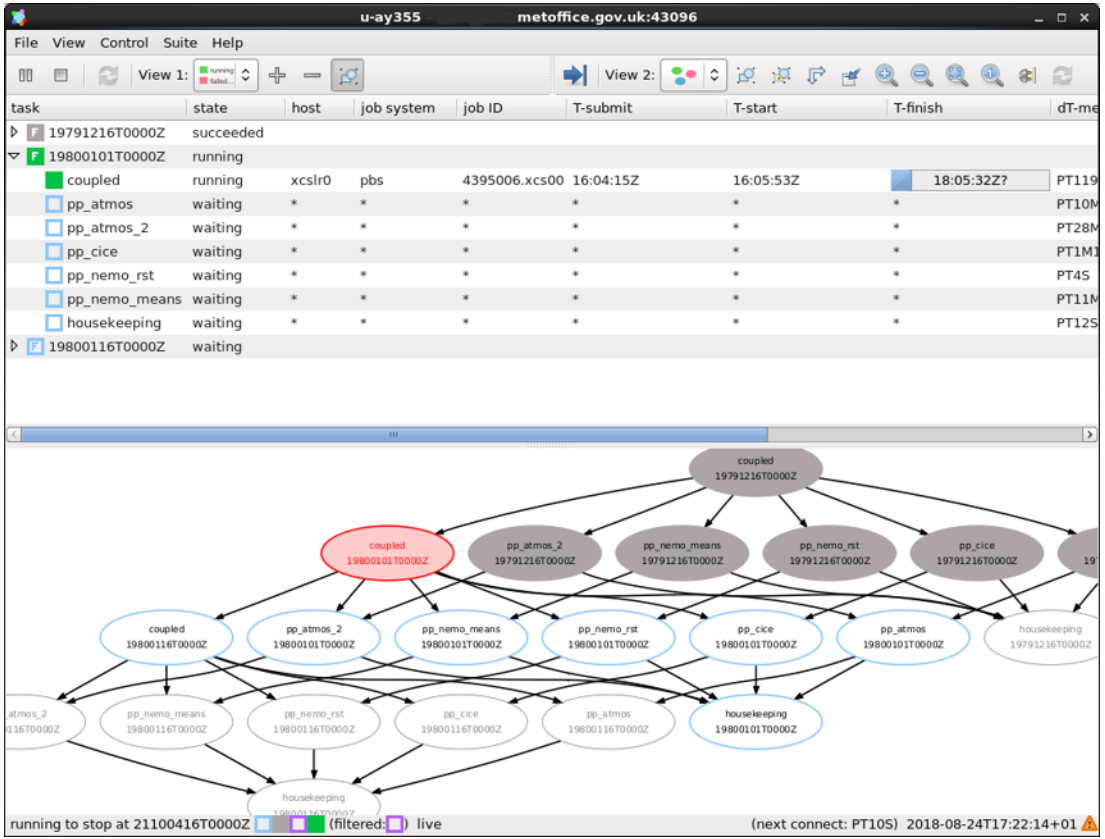
Project	Institute	Climate Model	Experiment	MIP Table	Variant Label	CMOR Name	Start Time	End Time	Online Status	# Data Files	# Data Issues	Tape URLs	File Versions	Data Size	Request Retrieval?
CMIP6	CNRM-CERFACS	CNRM-CM6-1	highresSST-present	day	r21i1p1f2	tas	1950-01-01	2014-12-31	offline	2	0	et:13127	v20180718	1.6 GB	
CMIP6	CNRM-CERFACS	CNRM-CM6-1-HR	highresSST-present	day	r1i1p1f2	tas	1950-01-01	2014-12-31	offline	13	0	et:13688	v20180823	11.5 GB	
CMIP6	EC-Earth-Consortium	EC-Earth3	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2015-12-31	online	792	0	et:9526, et:9487...	v20170911	12.6 GB	
CMIP6	EC-Earth-Consortium	EC-Earth3-HR	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2015-12-31	online	792	0	et:9128, et:9328...	v20170811	49.7 GB	
CMIP6	ECMWF	ECMWF-IFS-HR	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2014-12-31	online	65	0	et:9570	v20170915	10.0 GB	
CMIP6	ECMWF	ECMWF-IFS-LR	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2014-12-31	online	65	0	et:9569	v20170915	2.7 GB	
CMIP6	MOHC	HadGEM3-GC31-HM	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2014-12-30	online	65	0	moose:/adhoc/pr...	v20170831	30.8 GB	
CMIP6	MOHC	HadGEM3-GC31-LM	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2014-12-30	online	65	0	moose:/adhoc/pr...	v20170906	1.2 GB	
CMIP6	MOHC	HadGEM3-GC31-MM	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2014-12-30	online	65	0	moose:/adhoc/pr...	v20170818	5.8 GB	
CMIP6	MPI-M	MPIESM-1-2-HR	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2014-12-31	online	65	0	et:9906	v20171003	2.6 GB	
CMIP6	MPI-M	MPIESM-1-2-XR	highresSST-present	day	r1i1p1f1	tas	1950-01-01	2014-12-31	online	65	0	et:9673	v20171003	9.7 GB	

Login to create Retrieval Requests

Met Office Hadley Centre

PRIMAVERA

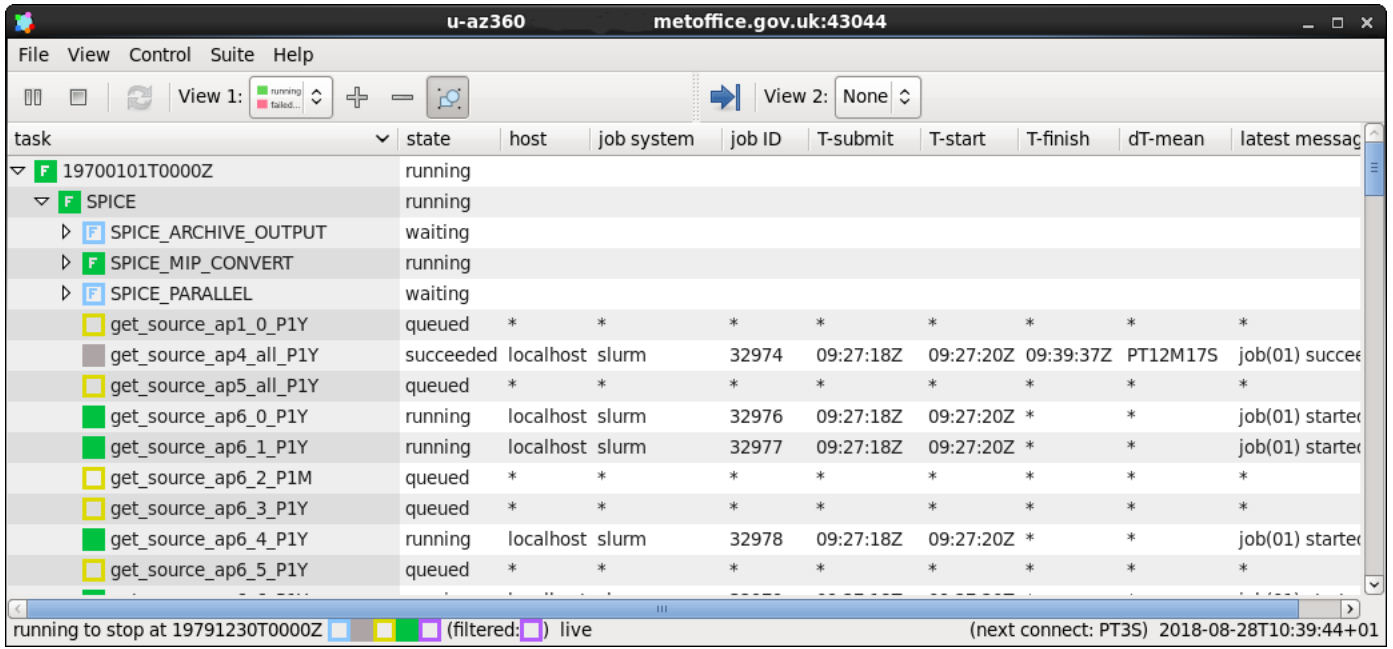
Climate Simulations



Met Office Hadley Centre

PRIMAVERA

CMORization



CMORization – suite generation

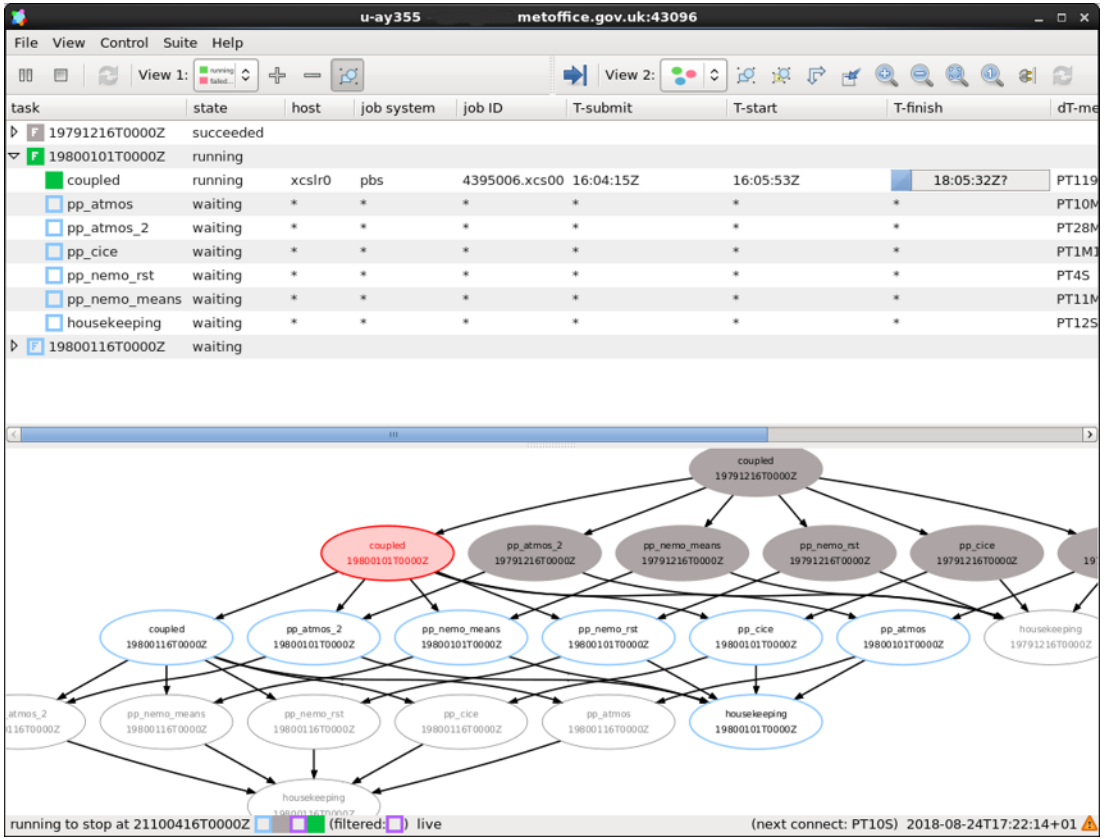
```
[jinja2:suite.rc]
ARCHIVE_OUTPUT=True
DRS_VERSION="v20180705"
FETCH_INPUT=True
FINAL_YEAR=1969
FIRST_YEAR=1953
GENERATE_VALIDATION=True
LOG_ROOT_DIRECTORY="$DATADIR/primavera_conversion_logs"
MODEL="u-ay585"
PARALLEL_TASKS=200
PROCESSING_INTERVALS=['P1Y', 'P6M', 'P3M', 'P1M']
SOURCE_ID="HadGEM3-GC31-HM"
VARIABLES={'ap4_all_P1Y': {'CMIP6_Amon': ['tas',
        =                          'ts',
        =                          'tasmin',
        =                          'tasmax',
        =                          'psl',
        =                          'ps',
        =                          'uas',
```

rose-suite.conf

Met Office Hadley Centre

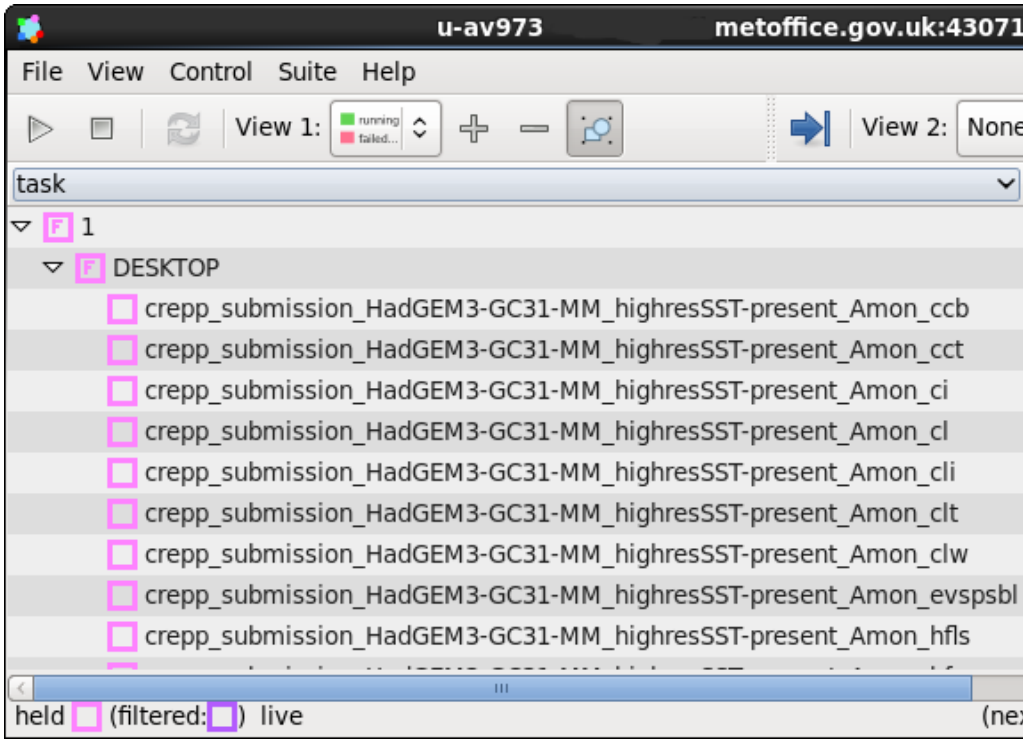
PRIMAVERA

Climate Simulations



ESGF Submission

- Over 10,000 submissions from the data at JASMIN to CEDA's ESGF node
- No cycling (but many tasks)
- Uses Cylc's familiar UI and Rose Bush for log file searching and reading



ESGF Submission

```
#!jinja2

[cylc]
    [[parameters]]
        streams = {{ TASK_FILE_NAME | load_json_tasks() }}

[scheduling]
    [[dependencies]]
        graph = crepp_submission<streams>
    [[queues]]
        [[[default]]]
            limit = {{ NUM_CONCURRENT_SUBMISSIONS }}
```

```
def load_json_tasks(filename):
    """
    Load the task names from the specified JSON file into a
    comma separated string.

    :param str filename: The full path of the file to load.
    :returns: The comma separated list of task names.
    :rtype: str
    """
    with open(filename) as fh:
        return ', '.join(json.load(fh))
```

Met Office Hadley Centre

PRIMAVERA

Lessons Learnt

- Data Management Tool
- Climate simulations on the HPCs
- CMORization
- Submission of data to ESGF

Questions?

For more information please contact

www.primavera-h2020.eu

jon.seddon@metoffice.gov.uk