EUROPEAN COMMISSION

Community research

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

IS-ENES2 DELIVERABLE (D -N°: 4.2)

Workflow Solutions Initial Workshop Report

File name: {IS-ENES2_D*4_2*.docx }

Authors: ***Kerstin Fieg, Jeremy Walton, Andrew Clark and Reinhard Budich***

Reviewer(s): ***Mick Carter Christian Page***

Reporting period: e.g. ***01/10/2014 – 31/03/2016***

Release date for review: ***03/01/2015***

Final date of issue: ***17/02/2015***

Revision table			
Version	**Date**	**Name**	**Comments**
1	15/01/2015	Mick Carter	First formal release
2	16/02/2015	Christian Pagé	Final version with reviewer's comments integrated

Abstract

{500 characters max}

Workflow is a rapidly developing area within the climate modelling community as a result of the increased complexity due to seasonal and decadal predictions systems with initialisation and increasingly complex experiment design. The deliverable reports on results from the first workshop that shared the experiences of the community using a variety of tools to see if there is any scope for concerned of approach or sharing of best practice.

Project co-funded by the European Commission's Seventh Framework Programme (FP7; 2007-2013) under the grant agreement n°312979		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants including the Commission Services	
RE	Restricted to a group specified by the partners of the **IS-ENES2** project	
CO	Confidential, only for partners of the **IS-ENES2** project	

Table of contents

Executive Summary

Background

This is the description of work for NA3 Task 1 on workflow solutions:

Typical workflow solutions used in climate modeling today are rather inflexible and mostly hard-wired. There are a number of workflow solutions that are being or will be evaluated within the climate modeling community such as SMS (Supervised Monitor Scheduler), developed by the European Centre for Medium-Range Weather Forecasts, and its replacement, ecFlow (when it becomes available), as well as Cylc (a meta-scheduler from the National Institute of Waters & Atmospheric Research, New Zealand). These have been developed for Numerical Weather Prediction suites and are of interest to the climate community because they have the flexibility to meet the more complex requirements of the increasingly important S2D systems (potentially including data assimilation) and more complex ensembles.

Other, more generic workflow solutions, such as those based on the KEPLER toolbox or the BPEL (Business Process Engineering Language) and workflow description languages (WDL) are also being investigated at DKRZ. BSC and UNIMAN can offer experience of workflows used in wider contexts. Some ESM sites in Europe have experience with a number of such tools; others are receiving intense investment (e.g. the MIKLIP project in Germany). Those with relevant experience will be invited to workshops and asked to verify the workflow solutions provided by partners.

There will be two workshops coordinated by DKRZ. A first workshop will identify issues and opportunities to be explored in more depth. The second workshop will also discuss the available post processing solutions in use in the community and how they are integrated into workflows.

The first of these workshops was held at DKRZ from Tuesday 3rd to Thursday 5th of June in order to deliver these aims.

The agenda and attendees of the workshop are provided in Appendix 1.

Statistics: There were 38 attendees from 18 institutions, of which 10 institutions were within the IS-ENES2 consortium. The external contributors were:

- Ufuk Turunczoglu, Istanbul Teknik Universitesi, Turkey
- Hilary Oliver, NIWA, New Zealand (main developer of the Cylc metascheduler)
- Rob Haines, University of Manchester (one of the main developers of the community workflow tool Taverna)
- John Dennis, NCAR / UCAR, USA
- Bernadette Fritzsch, AWI, Germany
- Steve Easterbrook, University of Toronto, Canada
- Amy Langhorst, Chandon Wilson, V. Balaji, Princeton University, USA

- Dean Williams, PCMDI, USA

A link to the workshop and the presentations can be found here:
https://verc.enes.org/ISENES2/events/isenes2-workshop-on-workflow

Summary of Outcome

The workshop met its aims.

A key finding was that there was a lot of interest in Cylc. Many groups are using it, are evaluating it, or are planning to evaluate it. It would be worth investigating whether it makes sense to invest in coordinated support / maintenance of Cylc.

With respect to CMIP6, we will be facing a data deluge. The question of whether we can afford this and if there is room for optimization was discussed.

Collecting experiment / provenance data is a topic which is receiving greater attention. This includes garnering details about the operating system, versions of libraries etc.

We need software which correctly handles hardware failures in exascale architectures.

1. The Presentations

1.1 Project Introduction, R. Budich, MPI-M

Reinhard Budich gave a brief introduction on the background of ENES and the goals of IS-ENES2, reminding the attendees of the aims of the workshop and the task within the IS-ENES2 networking activity, NA3, task 1.

1.2 Workflow Tools

1.2.1 Ufuk Turunczoglu, ITU: Kepler & Climate Modelling

Kepler was designed as a tool to deliver an abstraction layer which hides complexity from the user, orchestrates the workflow and collects metadata and provenance data. Kepler is an open source, platform independent (Java) workflow environment, tested in a prototype application (together with NOAA, with the ESMF team) in a grid environment and on a conventional computing cluster

Ufuk highlighted the need to be able to track the provenance of an experiment (both code and output); this turned out to be a recurring theme throughout the workshop. Under Kepler this was being addressed by means of a python script which collected information about the OS and compiler versions when running tasks. Intriguingly, version control was not being used. Additionally, the implemented system did not provide a generic solution and was described as being sensitive to changes in the model used.

Ufuk pointed out that web services / client–server architectures should be checked in terms of their usefulness to capture environment states for experiments. Another important point Ufuk made was that integrating climate model components into an off-the-shelf workflow toolkit like Kepler (or Taverna – see below) isn't easy because the components invariably use non-standard interfaces - that is, ASCII namelists, as opposed to something which is more likely to have built-in support like XML.

Lessons learned: to create a ready-to-use, domain-specific, end-to-end workflow is ambitious. Kepler provided a useful framework, but it was found that the integration of new model components is difficult, there is no mechanism to track versioning, there are no functionalities for checkpoints/restarts, the collection of provenance data is insufficient and the overall environment is not easy to handle, not efficient and needs continuous interaction.

1.2.2 Rob Haines, University of Manchester: Taverna

Taverna is a comprehensive scientific workflow management system with auxiliary tools for interactive or batch use. It provides services to analyze and manage data as well as collect provenance data. Taverna has been made available on github.

Its development focus is mainly the biodiversity community (*circa* 10,000 downloads, 1,000 users), in which the requirements are to find, share and organize data manipulation workflows. The workflows are set up using a GUI. Taverna enables flexible use of plug-ins and provides an easy-to-use interface for the incorporation of tools. HPC batch work relies on

polling a web service to check the queue status. Rob noted that nobody - to his knowledge - had used Taverna in ESM applications.

1.2.3 Hilary Oliver, NIWA: Cylc

Cylc is a meta-scheduler which is used to manage tasks and control related data processing. It has been described as easy to learn and to use, and is easy to adapt to individual requirements. A complete scientific workflow can be split up into thousands of individually manageable tasks; this means that starting, stopping and restarting (as well as failure correction and recovery) is possible during a model experiment.
A Cylc suite can optimally interleave tasks from multiple workflow cycles for fast catch up, fast trials, and offset parallel tests; they seamlessly transition between optimal catch up and real time operation with flow around delayed or failed tasks.
Hilary highlighted forthcoming changes such as ISO8601 and 360 day cycling for climate suites. An extensive Q&A followed this talk, during which it became apparent that a number of centers were using or looking to make use of Cylc for managing their experiments. In particular, it was noted that the Met Office were using Cylc for their operational suites; this long-term technical investment and the associated degree of support and development was deemed significant for the future of the system. EU and US involvement and future funding possibilities were discussed.

1.2.4 Andy Clark, UKMO: Rose

Rose provides a toolset to ease managing, configuring and running ESMs and other scientific applications in a transparent and as simple as possible way. Rose is built on top of the Cylc meta-scheduler to run suites. In Rose, suites are built up using so-called *apps*: these contain everything needed to run a given executable, and are simple to edit, compare and review. There is an optional GUI to ease configuration. It is also possible to version control suites.
Following the talk, there was discussion about whether Rose would support other version control systems (such as git) in future.

1.3 Experiences: data generation 1

1.3.1 Deike Kleberg, MPI-M: ORM based workflow management

For the MIKLIP project, there is a plan to replace the old shell-based workflow infrastructure by a modular object-oriented workflow management system (WfMS) based on Cylc which will be orchestrating the individual tasks. Objects are used to record commands, I/O and options used to run an experiment. In a final step, they are converted to database entries to enable provenance tracking. The main advantage of Rose is clearly the auto-recovery feature.

1.3.2 Craig MacLachlan, UKMO: GloSea5: operational forecasting system using Rose and Cylc

For seasonal forecast prediction with a coupled atmosphere-land-ocean-sea-ice model, a new system called GloSea5 has been introduced which is based on Rose and Cylc. Its main benefits are automatic retry / recovery functionality, reduced complexity and task parallelism.

Craig described the way GloSea4 had evolved to GloSea5, including the migration to Rose and Cylc. He discussed the benefits obtained from the migration and the ease with which this had been effected. Craig said that the main thing still missing was a tool for more powerful comparison of app configs than that offered by diff; it was noted that this is something, which other users have requested, and is on the ToDo list for the Rose team.

1.3.3 Jeremy Walton, UKMO: Climate data dissemination using the CREM workflow system

Jeremy predicted that it will become more complicated to handle Model Intercomparison Projects (MIPSs) in future, owing to greater amounts of data, more complex experiments and the appearance of multiple sub-MIPs. These additional complications, together with the difficulties with the workflow for CMIP5 experienced by all modeling groups have led to more work being done on systems such as CREM (Climate Research Experiment Management system) in the Met Office. This system can facilitate the automatic gathering of experiment metadata from other sources such as Rose and MOOSE (the Met Office's archiving system), and hence frees users from – for example – having to enter the same experiment metadata in several places. Here, Rose is used to maintain detailed configuration information on each simulation performed for the MIP, whilst MOOSE contains details about all data stored. Finally, model output is disseminated to ESGF (Earth System Grid Federation) after data conversion using CMOR tools, and quality checking.
The discussion following this talk focused on all participants' reported experiences and difficulties with CMIP5 - for example, problems with versioning data - with Dean Williams (see below) indicating some of the ways in which these would be improved for CMIP6.

1.3.4 John Dennis, UCAR: Redesigning the CESM post-processing workflow

John Dennis described his work on redesigning the CESM post-processing workflow. His focus was on the efficient processing of the data, and points at which lossy data compression could be performed without causing impacts which were scientifically significant. This sparked discussion as to the best ways to compress, and what to parallelize. It was noted that any use of lossy compression would need agreement of required accuracy for each parameter.

1.4 Experiences: Data Generation 2

1.4.1 Stéphane Sénési, MétéoFrance: CNRM-CM, ECLIS & EM

Stephane said that CNRM-CM, a workflow system developed by CERFACS, is able to handle projects like CMIP experiments. It consists of a bundle of shell scripts, customizable for individual requirements, uses git as version control, is relatively platform independent and has been tested for major systems like BullX, Cray, IBM. Once again, the focus was on capturing provenance information and being able to repeat experiments. Stéphane mentioned the increased requirements of CMIP6, and showed a visualization of the CMIP5 experiment database which had been done by Patrick Brockmann: - see http://bl.ocks.org/PBrockmann/raw/09571d7326e2ca057ede

1.4.2 Domingo Manubens, IC3 - Autosubmit: A Tool for Managing Climate Prediction Experiments

Domingo described Autosubmit as an object-oriented python tool to create, manage and monitor experiments. He discussed IC3's use of the tool to perform multi-member multi-ensemble comparison of models on different platforms (e.g. for IS-ENES2 WP9/JRA1).
At this stage, EC-Earth is the only model supported under Autosubmit, although there are future plans to develop it to cope with alternative models and ensembles. Another ISENES2 work package (WP9/JRA1) will be comparing Autosubmit with Cylc to analyze their potential suitability to run High-Resolution climate ensembles. There are plans to release Autosubmit as open source.
A major issue will be the adaption of its templates to handle IC3's MareNostrum. It was noted that it appears to require special code to be added for each site where it is to be used. A multi-member task wrapper is provided (for use on systems which require submission of very large jobs) but it was less clear how failures are handled. There is a plan to integrate with SAGA, (http://en.wikipedia.org/wiki/Simple_API_for_Grid_Applications) .
Task communication is realized by handing over log-files (polling the queuing system). Up to now, most parts of the workflow are serial, but ensemble simulation runs can be run in parallel. It was pointed out that there seemed to be no freedom to define the workflow, and there was a possibility of using Cylc for this to get more flexibility.

1.4.3 Amy Langhorst, GFDL: Lessons learned over a decade of workflow at GFDL

Amy said that GFDL's FMS (Flexible Modeling System) provides infrastructure and interfaces to enable reproducibility, robustness, efficiency and error handling. In addition, FRE (FMS Runtime Environment) is a set of workflow management tools that enables all steps of a climate model run from configuration to post-processing.
FRE offers templates for standard experiments which are then customized by the scientist (by making a series of entries in an XML file) to produce their configuration of their experiment. A list of the utilities available for managing and running jobs was also presented.
If a team member makes changes to site-level experiment configurations they are automatically tested using Jenkins against reference outputs to ensure code changes don't alter results.
With increasing scale (more data, higher resolution, more ensembles), data storage and data movement has to be reduced, and this increased complexity means that task handling by FRE is not possible anymore. For example, CM2.6 is generating 2.6 Tb per simulated year. Amy said that they need to enhance robustness and fault tolerance; their future plans involve a re-write of their system to incorporate a scheduler which they can trust (Cylc was indicated as their choice) and to handle the transition to post-processing that is more parallelized.

1.4.4 Chandon Wilson, GFDL: Infrastructure underpinning the GFDL workflow

Chandon presented an approach which makes HPC access available across multiple sites with minimal pain to the user. This involved a combination of single sign-on into a unified user space and use of expiring certificates.

He also discussed a utility that allows users to easily copy tasks across sites. This uses the Globus toolkit, which involves a powerful but complex setup to make it easy for the user. Chandin explained they'd written a generic copy tool to isolate the user from having to worry about the optimum way of transferring data between particular file systems. They were running unit tests hourly to ensure data transfers between all the various systems were working correctly. A generic copy tool allows logging of all data transfers to help with optimization and problem diagnosis.

1.5 Experiences: data procession and distribution

1.5.1 Stephan Kindermann, DKRZ: Climate processing web services and workflows

Stephan pointed out that, because data volume grows faster than network capacity and storage, compute resources have to be shifted towards data resources to reduce temporary storage and data movement. He said that the WPS interface standard provides rules for I/O, and eases interoperability and processing. He presented ClimDaPs as a WPS example with RestFlow as the workflow engine which orchestrates the WPS services in a data flow system.

1.5.2 Christian Page, CERFACS: Workflows in EUDAT: ENES and cross-communities

The main target for Christian's work was avoiding the download of TBs of unnecessary data by bringing computing resources to data storage in order to enable either data selection or data reduction without data copying / moving.
He described EUDAT as a cross-community scientific and technical collaboration to handle big data, which included looking at – for example – data processing / management / analysis problems using a common interface among different scientific communities.

1.5.3 Sandro Fiore, CMCC: Data Analytics workflow for climate

Sandro said that key data analytic requirements have been identified and addressed as big data challenges for the Ophidia project. Ophidia's architecture consists of a server, a front-end layer, a compute layer, operators, I/O nodes, an I/O server, primitives, a storage layer and the system catalogue. The system provides about 100 array-based primitives to, for example, perform data reduction (by aggregation), statistical analysis and compression. It uses parallel operators for analytics, and its storage model is suitable for multidimensional data. Sandro said that performance evaluation of Ophidia is still ongoing.

1.5.4 Bernadette Fritzsch, AWI: Workflow treatment in C3

Bernadette said that the Collaborative Climate Community (C3) Data and Processing Grid provides a user interface that allows searching, processing and downloading data from distributed data resources among the C3Grid federation. The infrastructure of the C3Grid delivers a scheduling system, data management service and data information system. The benefit for the user is the reduction of data traffic and the provision of replica management, whilst its main problems are a complicated technology, the need to educate the users and an

implementation of workflows which is mainly hard-coded. The alternative path is to use WPS.

1.5.5 Martina Stockhause, DKRZ: Long-term archiving workflow in CMIP5

Martina said that the purpose of long-term archival (LTA) and the IPCC DDC is to provide stable data for long-term interdisciplinary re-use. She presented the workflow needed to ingest data into LTA and the DOI process for the CMIP5 project. Some of her suggestions on workflow improvement for CMIP6 were related to project management structures, CMOR2, ESGF, informal and formal citation as well as external data services.

1.5.6 Grenville Lister, University of Reading: NCAS-CMS typical supported workflows

Grenville introduced the National Centre for Atmospheric Science (NCAS) as a NERC-funded distributed organization with a focus on climate change, modeling, prediction, observing and storing data; Computational Modeling Support (CMS) helped users with their modeling and data gathering. He gave an overview of the typical workflows that are supported by NCAS-CMS, which covered a discussion of the various platforms involved (ARCHER, JASMIN, PUMA etc.) and the running of models. He mentioned the requirement (or desire) for outputting files in NetCDF format directly from the Unified Model (UM).

1.5.7 Steve Easterbrook, University of Toronto: Doing science by building models

Following the conference dinner, Steve discussed his findings from his studies of the various "quirks" of model development as seen in various climate centers. Some highlights were a look at the model development lifecycle, software bugs, and a comparison of the codebases for the various ESM models used in CMIP5, concentrating on measures such as lines of code, the amount of shared code and the relative sizes of the different components (atmosphere, ocean, coupling, etc) in the different models.

1.6 Plans and Perspectives

1.6.1 Sebastian Denvil, IPSL: Self-healing workflows at IPSL

Sebastian described Convergence, a five-year project to develop a platform capable of running large ensembles of simulations with a suite of models, handling complex and voluminous datasets and facilitating model evaluation and validation, together with the use of high-resolution models.
Sebastian said that an effective I/O strategy is to use XML IO Server (XIOS) to generate data in a standardized format. He also pointed out that the requirements of CMIP6 mean that the volume of data will grow very quickly (by a factor of 50), and described IPSL's experiences of trying to do parallel writes of data which incorporated compression.
Sebastian mentioned the use of RabbitMQ for centralized logging – see http://en.wikipedia.org/wiki/RabbitMQ.

He also mentioned that the Simulation Control Environment is using one single main ksh script. The goal of the Environment is to hide the architecture from the user, making it easy and transparent for them to use very different infrastructures.
This talk concluded with some technical discussion about failure rates for large jobs on exascale systems, and the best ways to handle failures when they occur.

1.6.2 Ralf Müller, MPI-M: CDOs and Workflows

CDO is an example of a community-wide used toolset for data processing, which is a collection of command-line operators that manipulate and analyze climate data. Ralf described his work with CDO, and his attempt to remove the need to create many shell scripts by creating a python/ruby library which wrapped that toolset, acting as a smart caller of the CDO programs. The library chains CDO commands in order to minimize the use of (large) temporary files; this is particularly important when the final result is a small collection of values (for example, annual mean / global mean).

1.6.3 Dave Matthews, UKMO: Migrating to Rose and Cylc

Dave described how the Met Office migrated the infrastructure of its NWP system to Cylc. They planned to make the unified model (UM) fully configurable with Rose and Cylc from June 2014 onwards. The new system will be able to cover more complex and larger ensemble suites than would have been possible with the previous NWP environment. Dave noted that further advantages will be improved change control, the exposure of the full functionality of the UM to the scientist, and a big improvement in the way release management is handled. However, he pointed out that a lot of parameter values which were previously hidden from the user are now exposed through the Cylc / Rose interface

1.6.4 Luis Kornblueh, MPI-M: Rationales and optimization potentials of workflow management

Luis said that workflow management should both reduce the workload for the scientist and also secure responsible experimentation. He noted that an important piece of the workflow documentation is the creation of provenance data which documents the history of the data to enable re-use and reproducibility, and which gives credit to the creator of the data set. Analogies for workflow were made with school science experiments where the user presents their hypothesis, lists their apparatus, gathers results, analyses them and finally draws conclusions.
He said that future experiment organization and provenance data collection at MPI-M will be done using Cylc .

1.6.5 Dean Williams, PCMDI: Workflow requests for CMIP6

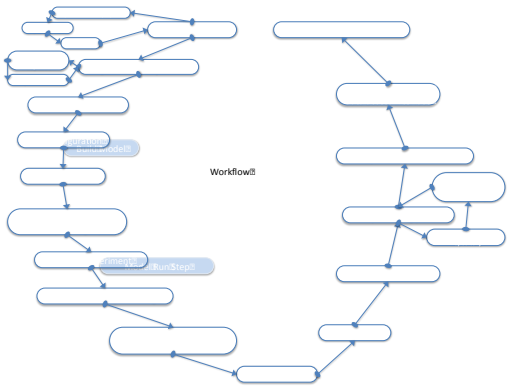
Dean discussed the workflow requirements of CMIP6. He pointed out that, assuming the amount of data and number of experiments for the next set of IPCC comparisons projects will grow from 3.5 PB in CMIP5 to around 3 EB in CMIP6, the most important issue is where to store the data and what to store.

Dean said that looking at this issue will have scientific implications for both model runs and workflow types: there will be the need to be able to capture and record runs and their settings during model development, and to be able to quickly evaluate and compare coupled model behaviors. In addition, he pointed out that there will be a general need to collect and process provenance data automatically in order to enable reproducibility and enhance productivity, to do diagnostics on the fly during the coupled model run and within one software system, and to handle heterogeneous hardware and software. Finally, he said that there will be an increasing demand for automatic processing, and that there was a strong need to reduce the volume of data transfer from the archive to local resources (i.e. the possibility of doing more processing on the archive machine should be investigated).
Dean also announced improvements to the Climate Model Output Rewriter (CMOR) software package. He also noted that a versioning procedure for data will be provided (or the procedure which was already in place would be re-emphasised): users of a dataset will be automatically notified if it has been updated, provided it has been updated in the correct way..

1.7 Discussion

Reinhard Budich led the final session in which he wrapped up the workshop and discussed possible action items. The workshop conclusions included:

- The complete end-to-end workflow can be described as a ring (see figure) starting with workflows for data production (model configuration, data preprocessing, model run, monitoring) and ending with workflows for data management (post-processing, storing, archiving).



- Model and data workflow are showing the tendency to merge, because the tasks – previously clearly separated – start to interact more intensively (e.g. preprocessing needs post- processed data, whilst archiving needs to harvest provenance data processed during model run).
- There is a clear requirement for a metascheduler which is able to orchestrate a number of individual tools (i.e. workflow tools, post-processing tools, configuration management tools, meta data capture tools, coupling tools). A coordinated effort suggests either the development of a new tool, or the coordination of the ongoing development and maintenance of an existing package.
- There is a lot of interest in Cylc – i.e., many groups are either using it, or are in the process of evaluating or planning to evaluate the package.
- There is much speculation about how much data will be generated and used during CMIP6, but it is clear that it cannot be handled with old-fashioned manual shell methods. Instead, an increasing amount of automatic processing will be required.

- Workflows of the next generation have to be prepared for exascale computers, and be able to cope with hardware failures with regard to recovery, restart and reproducibility. Experiment repeatability and result reproducibility (which isn't the same thing) will become a key issue.
- There is an emerging trend to bring computing resources to the data storage in order to minimize data movement and downloading.
- There will be a growing need to store community data (e.g. for CMIPs) into federated storage devices like ESGF with agreed interfaces and to guarantee the sustainability of this infrastructure.

There is a general problem regarding security: HPC systems are not set up to allow workflow solutions across defined trust zones.

2. Next Steps and Action Plans

The ISENES2 DoW foresees that "*There will be two workshops coordinated by DKRZ. A first workshop will identify issues and opportunities to be explored in more depth. The second workshop will also discuss the available post processing solutions in use in the community and how they are integrated into workflows"*.

Given the conclusions from the first workshop described above, the European workflow community should start to support Cylc development and maintenance as a possible future standard for scheduling as example of a new community tool. The EC call "E-Infra-5-2015 Centres of Excellence for computing applications" could represent an opportunity for this activity. ENES is currently preparing an application for this call, and will aim to incorporate workflow activities.

In the EUDAT context, there are activities on-going which address the challenge to "bring computing to the data". A working group on workflows has been initiated to handle such issues and would welcome more ENES involvement. Those interested should contact reinhard.budich@mpimet.mpg.de.

The next CMIP will pose serious challenges to the community, not only in scientific, but also in organizational and technical terms. ENES is currently coordinating requirements and governance suggestions for submission to the organizing bodies for CMIP6, but more engagement with, and contributions from, the community are always welcome. At the same time, the community institutions need to understand the fact that more technical expertise and resources within the institutions, closely linked to the CMIP6 organizing bodies and their activities, will be necessary to cope with this challenge. We suggest that there is a danger of having too many chiefs and too few Indians here.

Concerning the second workshop, gathering the right input on post-processing will be the major focus.

3. Appendix

3.1 List of invitees

Giovanni Aloisio (CMCC)	V. Balaji (GFDL)
Joachim Biercamp (DKRZ)	Patrick Brockmann (IPSL)
Reinhard Budich (MPI-M)	Andy Clark (UKMO)
John Dennis (UCAR)	Sebastien Denvil (IPSL)
Steve Easterbrook (Uni of Toronto)	Kerstin Fieg (DKRZ)
Sandro Fiore (CMCC)	Marie-Alice Foujols (IPSL)
Bernadette Fritzsch (AWI)	Ksenia Gorges (DKRZ)
Rob Haines (Uni of Manchester)	Nils Hempelmann (Climate Service Center)
Stephan Kindermann (DKRZ)	Deike Kleberg (MPI-M)
Luis Kornblueh (MPI-M)	Amy Langenhorst (GFDL)
Grenville Lister (Uni of Reading)	Michael Lautenschlager (DKRZ)
Thomas Ludwig (DKRZ)	Domingo Manubens (IC3)
Dave Matthews (UKMO)	Craig MacLachlan (UKMO)
Ralf Müller (MPI-M)	Hilary Oliver (NIWA)
Christian Page (CERFACS)	Kerstin Ronneberger (DKRZ)
Stephane Senesi (MeteoFrance)	Pavan Siligam (DKRZ)
Martina Stockhause (DKRZ)	Ufuk Turunczoglu (ITU)
Jeremy Walton (UKMO)	Kalle Wieners (MPI-M)
Dean Williams (PCMDI)	Chandin Wilson (GFDL)

3.2 PROGRAMME

June 3

12:00	*Lunch & Registration*	
13:00	Th. Ludwig (DKRZ)	Welcome
13:15	R. Budich (MPI-M)	Project Introduction

Session1, Chair K. Fieg: Workflow Tools and Concepts

13:30	Ufuk Turunczoglu (ITU)	Kepler & Climate Modeling
14:00	Rob Haines (Uni Manchester)	Taverna
14:30	Hilary Oliver (NIWA)	Cylc
15:00	Andy Clark (UKMO)	Rose: a framework for meteorological suites
15:30	*Coffee*	

Session 2, Chair R. Budich: Data Generation 1

16:15	D. Kleberg (MPI-M)	ORM based workflow management
16:40	Craig MacLachlan (UKMO)	GloSea5: operational forecasting system using Rose and Cylc
17:05	Jeremy Walton (UKMO)	Climate data dissemination using CREM workflow system
17:30	John Dennis (UCAR)	Redesigning the CESM post-processing workflow
17:55	Discussion	

June,4

Session 3, Chair Chr. Page: Data Generation 2

09:15	Stephane Senesi (MeteoFrance)	CNRM-CM, ECLIS and EM
09:40	Domingo Manubens (IC3)	Autosubmit: a tool for managing climate prediction experiments
10:00	*Coffee*	
10:45	Amy Langhorst (GFDL)	Lessons learned over a decade of workflow at GFDL
11:25	Chandin Wilson (GFDL)	Infrastructure underpinnings of the gfdl workflow
11:50	Discussion	
12:20	*Lunch*	

Session 4, Chair: M. Lautenschlager: Data Processing and Distribution

13:50	Bernadette Fritzsch (AWI)	Workflow Treatment in C3
14:15	Chr. Page (CERFACS)	Workflows in EUDAT: ENES and cross-communities
14:40	Sandro Fiore (CMCC)	Data Analytics workflows for climate
15:00	*Coffee*	
15:45	St. Kindermann (DKRZ)	Climate Processing web services and workflows
16:10	M. Stockhause (DKRZ)	Long-term archiving workflow CMPI5
16:35	Grenville Lister (Uni Reading)	NCAS-CMS typical supported workflows
17:00	Discussion	

Evening Lecture:

20:00	Steve Easterbrook (Uni Toronto)	Doing Science by building Models

June,5

Session 3, Chair V. Balaji: Plans and Perspectives

09:00	S. Denvil (IPSL)	Self healing workflows at IPSL
09:30	Ralf Müller (MPI-M)	CDOs and Workflows
10:00	Dave Matthews (UKMO)	Migrating to Rose and Cylc
10:30	Luis Kornblüh (MPI-M)	Rationals and optimization potentials of workflow management
11:00	*Coffee*	

Future Requirements

11:45	Dean Williams (PCMDI)	CMIP6 workflow requirements
12:30	Discussion on next steps	
13:15	Reinhard Budich (MPI-M)	Workshop wrap up, action items