

First report on the Scheduler development and support activity

Deliverable D3.9



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 675191

About this document

Work package in charge: WP3 Usability

Actual delivery date for this deliverable: 28 February 2017

Dissemination level: PU (for public use)

Lead authors:

Met Office (MetO): Mick Carter, Dave Matthews, Matthew Shin

Other contributing authors:

Max-Planck-Institut für Meteorologie (MPI-M), Reinhard Budich (Reviewer)

Contacts: esiwace@dkrz.de

Visit us on: www.esiwace.eu

Follow us on Twitter: [@esiwace](https://twitter.com/esiwace)

Disclaimer: This material reflects only the authors view and the Commission is not responsible for any use that may be made of the information it contains.

Index

1. Abstract /publishable summary	4
2. Conclusion & Results.....	4
3. Project objectives.....	4
4. Detailed report on the deliverable	5
5. References (<i>Bibliography</i>).....	8
6. Dissemination and Uptake.....	8
6.1 Dissemination	8
6.2 Uptake by the targeted audience.....	8
7. The delivery is delayed: <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No	9
8. Changes made and/or difficulties encountered, if any.....	9
9. Efforts for this deliverable	9
10. Sustainability.....	10
10.1. Lessons learnt: both positive and negative that can be drawn from the experiences of the work to date.....	10
10.2 Links built with other deliverables, WPs, and synergies created with other projects	10
11. Dissemination activities	10

1. Abstract /publishable summary

This document describes the development work done to the Cylc meta-scheduler as well as the engagement and support activities done within the project up to the date of the report. This is put in the context of the wider effort of Cylc development. The bulk of the effort in this period has been development and the priorities of that development were agreed at a workshop on workflows organised by the IS-ENES2 project.

2. Conclusion & Results

During the period covered by this report, 1 September 2015 start date of ESiWACE to 28 February 2017, there have been 351 pull requests (changes) merged into the Cylc repository. 270 of these pull requests were created by MetO staff and all have been reviewed by MetO staff on behalf of ESiWACE.

Highlights are:

- Client-server interaction enhancements, including moving from the obsolete Pyro3 protocol to industry standard HTTPS protocol;
- New built-in support for parameterised tasks for cleaner workflow;
- Restart robustness and runtime database improvements;
- Performance and feature enhancement to the task/job management subsystem;
- New suite and task event handlers;
- Performance and feature enhancements to "cylc gui" and "cylc gscan" GUIs;
- Performance and efficiency improvements to the suite validation and runtime;
- Significant growth in the number of automated tests;
- Improvement to the documentation.

3. Project objectives

This deliverable contributes directly and indirectly to the achievement of all the macro-objectives and specific goals indicated in section 1.1 of the Description of the Action:

Macro-objectives	Contribution of this deliverable?
Improve the efficiency and productivity of numerical weather and climate simulation on high-performance computing platforms	Yes
Support the end-to-end workflow of global Earth system modelling for weather and climate simulation in high performance computing environments	Yes
The European weather and climate science community will drive the governance structure that defines the services to be provided by ESiWACE	Yes
Foster the interaction between industry and the weather and climate community on the exploitation of high-end computing systems, application codes and services.	No
Increase competitiveness and growth of the European HPC industry	No

Specific goals in the workplan	Contribution of this deliverable?
Provide services to the user community that will impact beyond the lifetime of the project.	Yes

Improve scalability and shorten the time-to-solution for climate and operational weather forecasts at increased resolution and complexity to be run on future extreme-scale HPC systems.	Yes
Foster usability of the available tools, software, computing and data handling infrastructures.	Yes
Pursue exploitability of climate and weather model results.	Yes
Establish governance of common software management to avoid unnecessary and redundant development and to deliver the best available solutions to the user community.	Yes
Provide open access to research results and open source software at international level.	Yes
Exploit synergies with other relevant activities and projects and also with the global weather and climate community	Yes

4. Detailed report on the deliverable

MetO has provided all the ESIWACE resources to this deliverable but details are also provided on the effort provided from non-ESIWACE partners in the development of Cylc and in Cylc related outreach to the wider community showing the benefit of the open and collaborative nature of the Cylc software development. The team acknowledges the critical importance of the input and leadership provided by Hilary Oliver (NIWA¹) in the development of Cylc.

Introduction

This report covers the activities undertaken so far by the Met Office under task 3.3 of WP3 Usability. Note that this includes tasks 3.3.1 - 3.3.3. Task 3.3.4 is not led by the Met Office and is out of scope for this report.

Task 3.3.1: Scoping the work

An initial ESIWACE Cylc development and support plan was published in February 2016 (milestone 1): <https://www.esiwace.eu/results/milestones/ms-1-report/view>

This proposed the development priorities, the support approach and an initial governance process. The original intention was that this would be informed by an initial workshop. However, in consultation with both potential attendees and with the work package lead it was decided to delay this workshop and, instead, hold a joint workshop with the related IS-ENES2 activity. This approach had the advantage of having a much wider audience.

This workshop took place in Lisbon in September 2016:

<https://is.enes.org/events/final-is-enes2-workshop-on-workflow-solutions-1>

The plan was presented at the workshop:

<https://is.enes.org/events/final-is-enes2-workshop-on-workflow-solutions/slides/26Matthewsesiwace.pdf>

The workshop attendees agreed to the plan without any changes.

¹ <https://www.niwa.co.nz/>

Task 3.3.2 Cylc development

The main contributors (with 10+ pull requests) to cylc in this period were:

- Dr Hilary Oliver, NIWA (Cylc's original author)
- Matt Shin, MetO
- Oliver Sanders, MetO (joined team Jan 2016)
- Andy Clark, MetO
- Ben Fitzpatrick, MetO (left team Oct 2016)

Looking at the period 1 September 2015 (start date of ESIWACE) - 28 February 2017 (the date this report was written) there have been 351 pull requests merged into Cylc master and/or other supported branches. Each of these is reviewed and signed off by at least one member of the Met Office. Furthermore, 270 of these pull requests were created by Met Office staff. Each pull request is a self-contained change to the code base and they vary from simple one-liners through to large changes which may have taken months to code and test.

Details of the key changes in each Cylc release are detailed in the change log: <https://github.com/cylc/cylc/blob/master/CHANGES.md>

Key highlights during this period are:

- Client-server interaction enhancements, including moving from the obsolete Pyro3 protocol to industry standard HTTPS protocol.
- New built-in support for parameterised tasks for cleaner workflow.
- Restart robustness and runtime database improvements.
- Performance and feature enhancement to the task/job management subsystem.
- New suite and task event handlers.
- Performance and feature enhancements to "cylc gui" and "cylc gscan" GUIs.
- Performance and efficiency improvements to the suite validation and runtime. (See later.)
- Significant growth in the number of automated tests.
- Improvement to the documentation.

In June 2016, Dr Hilary Oliver spent 2 weeks at the Met Office. This was the first face-to-face meeting of the full development team for 2 years and it was used as an opportunity to review all of the open issues for Cylc. This gave the team a chance to prioritise each issue, agree the way forward on a number of unresolved problems and also close down and consolidate a number of issues. At the end of the 2 weeks we had reduced the number of open issues from ~160 to ~130. We also identified any small issues which could be easily fixed and, by focussing some effort on these, we have since reduced the number of open issues down to ~100. The resulting plan and priorities was presented at the Lisbon workshop (see later) and agreed by the participants – see later.

A lot of effort over the last few years has gone into making Cylc more efficient. We recently did some tests to investigate how much impact this work has had. As an example, the graph below shows dramatic improvements in memory usage in recent Cylc releases for a large ensemble suite which is used at the Met Office.

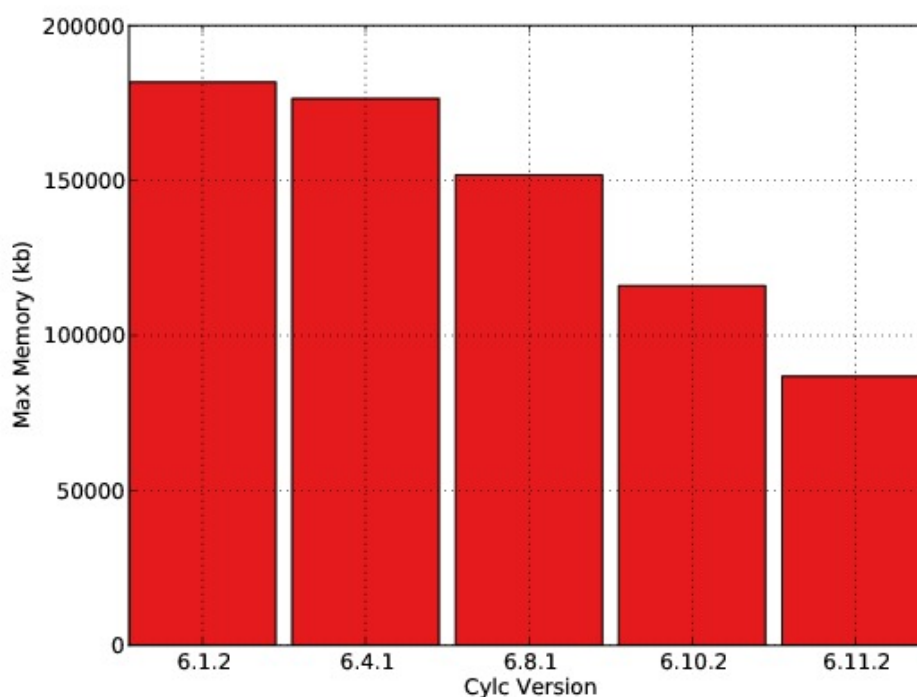


Figure 1: This, and other performance improvements enhance the scalability of Cylc which will, in turn, better support the NWP and ESM communities using Cylc

In addition to the work on Cylc, there are also important improvements to Rose (<https://github.com/metomi/rose>) over this period. Some of Rose’s functionality will migrate into Cylc (which is their more logical home) in later stages of the project. One of the key parts is Rose Bush - a powerful web based interface which allows you to view all the outputs generated by a Cylc suite. There have been some significant enhancements to Rose Bush during this period resulting in a much improved view of the task and job states.

It should be noted that the development effort described here covers the entire Met Office contribution to Cylc development which is considerably more than is funded by ESIWACE. The funding provided by ESIWACE enables the Met Office to maintain a much greater pace of development than would otherwise be possible. It also means we can prioritise the developments which are of most benefit to the Cylc community as a whole rather than focussing solely on Met Office priorities.

Task 3.3.3 Support services

We took the opportunity of the workshop in Lisbon to provide a Cylc Tutorial. The tutorial was run jointly by the Met Office and Dr Hilary Oliver. We had 17 people register interest in advance of the workshop agenda being finalised. Due to a clash with the final agenda, the number of people who actually took part in the tutorial was 10, from 5 different institutions. The tutorial was based on virtual machines that had been prepared in advance which the participants installed on their own laptops. A big advantage of this approach is that participants were able to take away a working installation which they could use for continued experimentation. The tutorial was well received and it leaves us in a good

position to offer similar tutorials in the future.

Although several different institutions gave presentations at the Lisbon workshop which described their use of Cylc, the demand for Cylc support is generally rather low. This can be seen as encouraging since it implies that sites that adopt Cylc do not, typically, need a large amount of support. The low level of support needed has also meant that issues or questions which have been raised via Github, the Cylc Google group or via email have all been responded to promptly. Further, the effort has been invested into product improvements that are driven by requests in the active user community and will hence make Cylc more attractive.

Demand for support services could increase dramatically if any new sites decide to start evaluating Cylc and ask for help with installation, training, etc. The ESIWACE funding means that this support is available if required. However, if demand for the support services remains low this allows us to continue focussing on the development activities which are of clear benefit to the whole Cylc community.

5. References (*Bibliography*)

None

6. Dissemination and Uptake

6.1 Dissemination

- Cylc was presented to the community at the joint IS-ENES2 workshop on meta-data and workflow in Lisbon in September 2016: <https://www.esiwace.eu/events/joint-final-workshop-on-is-enes2-workflow-solutions-in-earth-system-modelling-and-on-meta-data-generation-during-experiments>
- The plan was presented at the workshop: <https://www.esiwace.eu/results/conference-contributions/scale-and-breadth-of-cylc-usage-at-met-office-27-29-september-2016-costa-da-caparica/view>

Peer reviewed articles

None

Publications in preparation OR submitted

None

6.2 Uptake by the targeted audience

As indicated in the Description of the Action, the audience for this deliverable is:

X	The general public (PU)
X	The project partners, including the Commission services (PP)

This is how we are going to ensure the uptake of the deliverables by the targeted audience

This deliverable will be uploaded in the website.

7. The delivery is delayed: Yes No

8. Changes made and/or difficulties encountered, if any

In the Description of the Action Part A, description of the Work package 3, the following was envisaged:

“Task 3.3.1: Scoping of the Work [Lead: MetO; Participants: MPG] Networking Activity, Development of service: A user-centric, initial workshop will establish the development priorities and the most effective form of user support services. It will, with the aid of WP1, establish a governance process for continued support and development activities within and after the project (consistent with the ESIWACE business plan [D3.1]). MetO will organize and evaluate the workshop; the other partners will help prepare, participate, and will assist in the evaluation. We plan to involve in this task two supporting institutions; NIWA and GFDL (see the letters of support provided in Appendix 1)”

However, with the support of the work package lead, the initial workshop was delayed so that a **wider audience** could be consulted at the Lisbon workshop on workflow and meta-data in September 2016.

9. Efforts for this deliverable

Person-months spent on this deliverable:

Beneficiary	Person-months	Period covered	Names of scientists involved, including third parties (if appropriate) and their gender (f/m)
DKRZ			
ECMWF			
CNRS-IPSL			
MPG			
CERFACS			
BSC			
STFC			
MET O	15.4	01/09/2015 to 28/02/2017	Dave Matthews (m) Matthew Shin (m) Ben Fitzpatrick (m) Oliver Sanders (m) Andy Clark (m)
UREAD			
SMHI			
ICHEC			
CMCC			
DWD			
SEAGATE			
BULL			
ALLINEA			

Total	15.4		
--------------	-------------	--	--

10. Sustainability

10.1. Lessons learnt: both positive and negative that can be drawn from the experiences of the work to date

The uptake of an “industrial strength” workflow manager such as Cylc needs both working level engagement and institutional level support. For any single activity, it is often quicker to improve existing (often script-based) workflow solutions. However, the benefits outweigh the overheads when implemented at the institutional level because of the economies of scale. The opportunities to realise these benefits are rare and are most likely to occur when there is a need for a significant step change in workflow complexity or a significant existing problem with workflow that needs to be overcome.

The performance and memory footprint of a workflow manager is as important as functionality to allow the solution to scale to meet institutional needs.

10.2 Links built with other deliverables, WPs, and synergies created with other projects

Not applicable

11. Dissemination activities

Type of dissemination and communication activities	Number	Total funding amount	Type of audience reached In the context of all dissemination & communication activities	Estimated number of persons reached
Training: Cylc tutorial	1	1,470€	Scientific Community (higher education, Research)	10
Participation to a workshop: Workshop on Workflow Solutions in Earth System Modelling and Meta-Data Generation during Experiments	1	-	Scientific Community (higher education, Research) Industry	52
Total funding amount		1,470€		

Intellectual property rights resulting from this deliverable

None. Cylc is open source.