



Scheduler support and development plan (Task 3.3.1: Scoping of the Work)

Milestone MS1



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

Lead Beneficiary: Met Office (MET O), Mick Carter and David Matthews

Other contributing authors: Max-Planck-Institut für Meteorologie (MPI-M), Reinhard Budich (Reviewer)

Delivery date Annex 1 (DoA): month 6

Means of verifications: This milestone allows the deliverable D3.9 to meet the needs of the user community by the date of delivery of D3.9.

Achieved: Yes

Comments: A full report is available on the ESIWACE intranet

1. Abstract /publishable summary

This document supports the work of Task 3.3 of Work Package 3 (Usability) for the ESiWACE project. Specifically, the aim of Task 3.3 is to build a supported, user-driven community around the cylc meta-scheduler for complex climate and weather suites on HPC systems. The document provides an initial Cylc development plan based on existing interactions with the cylc user community and information gathered at the ENES workshop on workflow.

2. Conclusion & Results

This document concludes that the best way to get community engagement is to coordinate with the ENES workflow workshop expected in September 2016 and to advertise support services via ESiWACE, ENES and other dissemination channels.

An interim development programmes is provided based on:

1. fixing bugs as exposed by wider set of applications and platforms;
2. moving functionality into cylc from a complementary software package;
3. improving scalability to meet the need to cope with increased workflow complexity and code modernisation.

A support plan has been delivered recognising the phases of uptake and light weight governance structure has been agreed with the work-package lead.

3. Project objectives

This milestones 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 milestone? |
|--|---------------------------------|
| 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. | Possibly |
| Increase competitiveness and growth of the European HPC industry | Yes |

| Specific goals in the workplan | Contribution of this milestone? |
|---|---------------------------------|
| 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. | No |

| | |
|--|----------|
| Foster usability of the available tools, software, computing and data handling infrastructures. | Yes |
| Pursue exploitability of climate and weather model results. | No |
| Establish governance of common software management to avoid unnecessary and redundant development and to deliver the best available solutions to the user community. | Possibly |
| 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 achievement of this milestone

Introduction

This document supports the work of Task 3.3 of Work Package 3 (Usability) for the ESiWACE project. Specifically, the aim of Task 3.3 is to build a supported, user-driven community around the cylc meta-scheduler for complex climate and weather suites on HPC systems

Scoping the work

ESiWACE Task 3.3.1 is to establish development priorities and the most effective form of user support services via an initial user workshop. Currently, we have the following information:

- Input from the first IS-ENES2 workflow workshop;
- An existing forum to prioritise the needs of current cylc users through both the ticketing systems on the cylc development site and input from a Technical Advisory Group¹ (TAG) representing a large proportion of the current user base;
- A clear set of development requirements from the current broad user base of cylc users based on the forum and guidance from the TAG;
- An opportunity to engage with the ESiWACE user community through a second IS-ENES2 workflows workshop later in 2016.

Hence it has been proposed that, to maximise attendance of potential users, we will coordinate the next stage of user engagement alongside the IS-ENES2 workflows workshop in September 2016. This will not affect the development activities since the short term priorities are already well defined.

Cylc adoption is best achieved through a coordinated organisational approach to avoid duplication of effort. Hence, it is important to emphasise the ability of cylc to support a very broad range of activities and platforms and to advertise this and the breadth of existing institutions using cylc. The ability to sustain the needs of this broad range of users and applications should act as evidence to encourage new institutions to invest. A delayed initial user workshop will allow us to gather further information and evidence across the following:

- Current institutions using cylc, including:
 - Weather forecasting centres;
 - Climate research institutions and groups
- For a wide range of uses:

¹ The Technical Advisory Group is a group that provides technical advice to a programme of activities delivering software improvements to a partnership of institutions using the Unified Model and related software, including Cylc.

- The scheduling of complex, real-time operational weather forecasting suites and operational seasonal prediction suites;
- Climate modelling teams developing cylc driven production activities feeding into CMIP6 and supporting cross platform and institution activities;
- Teams using cylc to support automated testing to improve quality assurance in development processes;
- Researchers using cylc to manage bespoke experiment design;
- Developers looking for a framework that will allow forcing data to be generated automatically within reproducible workflows;
- Scientists wanting to provide a higher degree of automation to allow them to manage ever increasing volumes and complexity of model output and observational data.
- A broad range of experience levels, including:
 - Projects using cylc for new purposes through to mature activities looking to learn from their experiences.
 - Organisations new to installing and supporting a cylc service to those running robust, real time operations.
- On a variety of supporting platforms from individual workstations, through to clusters and onto a range of HPC facilities with a varying set of security constraints.

Cylc development

ESIWACE Task 3.3.2 covers development of cylc. This includes both development of new features required by the user community and on-going improvements to the code aimed at ensuring continued supportability and allowing a wider community engagement in the code development.

To build confidence in the community, it is important that institutions are aware of the robustness of the development processes and we note that cylc has well established and effective working practises to provide such confidence. The core development team currently consists of Dr Hilary Oliver, NIWA (cylc's original author) plus several members of the Met Office Modelling Infrastructure team. Other active developers are welcome to contribute and the openness of the development was a key issue when cylc was chosen at the Met Office and was recognised as an important feature at the first IS-ENES2 workflow workshop. All development is done via GitHub² using the GitHub Flow workflow³.

All code changes:

1. Are documented and discussed on GitHub;
2. must comply with Python PEP8 coding standards;
3. must have associated automated tests to ensure future stability;
4. must be reviewed and signed off by at least two core developers before they are merged to the master branch.

Cylc is in operational use at multiple sites so it is essential to minimise bugs and ensure that new features do not create future maintenance issues or problems when upgrading. The working practises have evolved with this in mind and the team is experienced in delivering robust and portable releases. Requests for new features (and bug reports) get recorded as issues on GitHub, typically after initial discussion with the development team.

² <https://github.com/cylc/cylc>

³ <https://guides.github.com/introduction/flow/>

There are currently over 160 open issues for cylc. Milestones are used to help define the development priorities. There is also a wiki page which defines some longer term development goals (<https://github.com/cylc/cylc/wiki/Cylc-Refactor-Proposal-%282014%29>).

The key elements that are of most benefit to the ESIWACE community are:

1. **Bug fixing!** Cylc is being used for an ever increasing breadth of activities and new platforms. Bugs will inevitably emerge and responding quickly to these is a priority for the development team.
2. **The migration of some Rose functionality into cylc⁴.** From the first IS-ENES2 workflow workshop, it is clear that many of the ESIWACE community are likely to be primarily interested in cylc and will not want to also use Rose, a separate set of tools developed at the Met Office to complement cylc, at least initially. Hence, support for suite installation and a web interface for viewing suite output will be moved into cylc from Rose. Any cylc users not already using Rose will find these new features of significant benefit. This also has architectural benefits as incorporating these features into cylc should make them easier to maintain and make it easier to develop future enhancements.
3. **Cylc scalability.** Based on feedback within the current user community, we need to allow cylc to scale to meet the needs of larger computing platforms and increasingly complex workflows to support activities such as ensemble weather prediction and operational seasonal prediction systems. We will optimise the suite daemon to reduce memory footprint, speed up the scheduling algorithm and provide an efficient data storage model to document the full lifetime of a running suite. This work will ensure that cylc is fast and robust and does not become a bottleneck for computing resources.
4. **Modernisation.** Ease of installation and conformance with organisational technical constraints is increasingly important as cylc is deployed to more institutions. The team will update the communications layer to use a RESTful API via HTTP to replace the current mechanism based on RPC. Another driver for this is ensure that cylc is not limited by Pyro3 - a legacy Python remote object library which is no longer maintained.

Longer term potential developments could include:

1. A rework of a number of the cylc tools taking advantage of the move to the new RESTful API communications layer. These can be improved using readily available web technology.
2. A re-implementation of the GUIs using newer technology, e.g.: Web technology instead of GTK.
3. Migration to Python 3. This (and the GUI re-implementation) is likely to become essential to future proof cylc, ensure portability, etc.

Ongoing engagement with the current user base across the broad ranges of applications, institutions and platforms makes us confident that, in advance of the ESIWACE workshop on cylc, this roadmap will deliver clear benefits to both existing and potential future users of cylc.

The user workshop planned for September 2016 will be used as an opportunity to further refine these plans and to balance effort between development and support.

Support services

ESiWACE Task 3.3.3 covers the training and support services which will be provided to the community. The support requirements for cylc will need to take into account the various phases of adoption. We

⁴ <https://github.com/metomi/rose>

cannot force cylc adoption, we can only encourage it and so our approach needs to take into account the needs of those groups making decisions to invest the effort necessary to use cylc.

Already, we have a well-established forum for users wanting to find out about and engage with cylc (<https://groups.google.com/forum/?fromgroups#!forum/cylc>). This allows decision makers to openly see what is going on in the community and how the cylc support teams respond to the needs of end users. We will also welcome specific requests for support made directly by email to metomi@metoffice.gov.uk.

Other than the proposed advertising activity, which will start in March 2016, the following proposal will be taken to the community for agreement at the workshop in September 2016.

Phase 1. Evaluation - Advertise and Support Decision Making

We will advertise the availability of cylc support through ESIWACE via the web portal and by email distribution lists including the ENES/IS-ENES community. Other appropriate opportunities will be sought. Once we have user engagement, the support service will concentrate on 1:1 bespoke consultancy which will be tailored to meet the needs of the interested party. We envisage that this could include:

1. Site visits or video conferences;
2. Hosting visits to the Met Office for demonstrations and initial evaluations;
3. In-person or remote-access support for installation.

This engagement will be as broad as resources allow but, if necessary, we will prioritise European weather and climate institutions.

In parallel with this, we will also ensure that the cylc documentation contains all the necessary information to support decision making. This documentation will advise institutions on how to plan to establish a cylc service and what resources they will need. This will be kept at a high level as the primary thing is to provide the direct and bespoke engagement that is best for any one institution.

Phase 2. Implementation - Training and Advice

As decisions are made and institutions move from evaluation into implementation, the needs will change. At this stage, we will need to train local support staff and support the training of end users as well as answering questions as people learn how to exploit the new facility. Existing training materials will be further improved based on feedback. The principles of the support approach will be agreed at the workshop. Priorities will be managed via the governance structure in section 5.

Initial Governance Structure

Priorities and the balance between support and development services, within the ESIWACE funded effort of 1.5 FTE will be controlled by a very light weight user governance process in line with the effort available in the project. It should be noted that the development team already have effective ways of engaging with the community and are considered responsive – see section 3 for development and section 4 for support.

Institutions will be encouraged, but not initially mandated, to have a single point for contact for cylc who would typically be trained to give first-level support and would oversee any priorities of requests from the point of view of their institution. Typically, this would be the person managing the cylc service locally.

Where there are conflicts (in terms of the priority given to particular support or development requests) the work package lead within ESIWACE (Reinhard Budich) will be the point of contact to resolve the dispute. If necessary, Hilary Oliver may be contacted (whilst he remains in his role in cylc development) for independent technical advice.

A more detailed basis for prioritisation will be agreed at the first meeting, in September 2016, but until that time, the following principles will be used for the ESIWACE funded effort:

- We prioritise support over development to encourage uptake and evaluation.
- We prioritise European weather and climate sites over other users.

If this prioritisation fails to meet the user need, a user group will be established under the chairmanship of the work package lead within ESiWACE (Reinhard Budich).

5. References (*Bibliography*)

No references, see the links and footnotes in the document:

- <https://github.com/cylc/cylc>
- <https://guides.github.com/introduction/flow/>
- <https://github.com/metomi/rose>

6. Dissemination and uptake

As indicated in the Description of the Action, the present report will be disseminated to the following target audiences:

| | |
|---|--|
| | The general public (PU) |
| X | The project partners, including the Commission services (PP) |
| X | A group specified by the consortium, including the Commission services (RE) |
| | This reports is confidential, only for members of the consortium, including the Commission services (CO) |

This is how we are going to ensure the uptake of this report by the targeted audiences:

Dissemination via the IS-ENES and ESiWACE communication channels (portals and mailing lists).

7. The milestone is delayed: Yes No

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

An initial workshop was envisaged but not held in the end because:

1. We have enough information from users and interested parties as described in the document;
2. It was our judgement, and that of the work-package lead, that we would get better engagement by coordinating with the IS-ENES2 workflow workshop.

9. Efforts for this milestone

Person-months spent on this milestone:

| Beneficiary | Person-months | Period covered | Names of scientists involved, including third parties (if appropriate) and their gender (f/m) |
|-------------|---------------|----------------|---|
| DKRZ | 0 | | |
| ECMWF | 0 | | |
| CNRS-IPSL | 0 | | |
| MPG | 0 | | |
| CERFACS | 0 | | |
| BSC | 0 | | |
| STFC | 0 | | |
| MET O | 0.25 | February 2016 | Mick Carter (male), Dave Matthews (male) |

| | | | |
|--------------|-------------|--|--|
| UREAD | 0 | | |
| SMHI | 0 | | |
| ICHEC | 0 | | |
| CMCC | 0 | | |
| DWD | 0 | | |
| SEAGATE | 0 | | |
| BULL | 0 | | |
| ALLINEA | 0 | | |
| Total | 0.25 | | |

10. Sustainability

Links will be built via coordination with the IS-ENES2 workflow workshop.