



Model Benchmarking

Forrest M. Hoffman, *Task Team co-lead*



The Model Benchmarking Task Team



1. Rebecca Beadling, *USA*
2. Ed Blockley, *UK*
3. Jiwoo Lee, *USA*
4. Valerio Lembo, *Italy*
5. Jared Lewis, *Australia*
6. Jianhua Lu, *China*

7. Luke Madaus, *USA*
8. Elizaveta Malinina, *Canada*
9. Brian Medeiros, *USA*
10. Wilfried Pokam Mba, *Cameroon*
11. Enrico Scoccimarro, *Italy*
12. Ranjini Swaminathan, *UK*

Diversity in expertise (realms and methods), user group representation, gender, location, career stage

Overarching goals:

- Systematic and rapid performance assessment of the expected models participating in CMIP7 (including the model output and documentation)
- Enhancing existing community evaluation tools that facilitate performance assessment of models
- Integration of evaluation tools into CMIP publication workflows and fostering publication of their diagnostic outputs alongside the model output on the ESGF

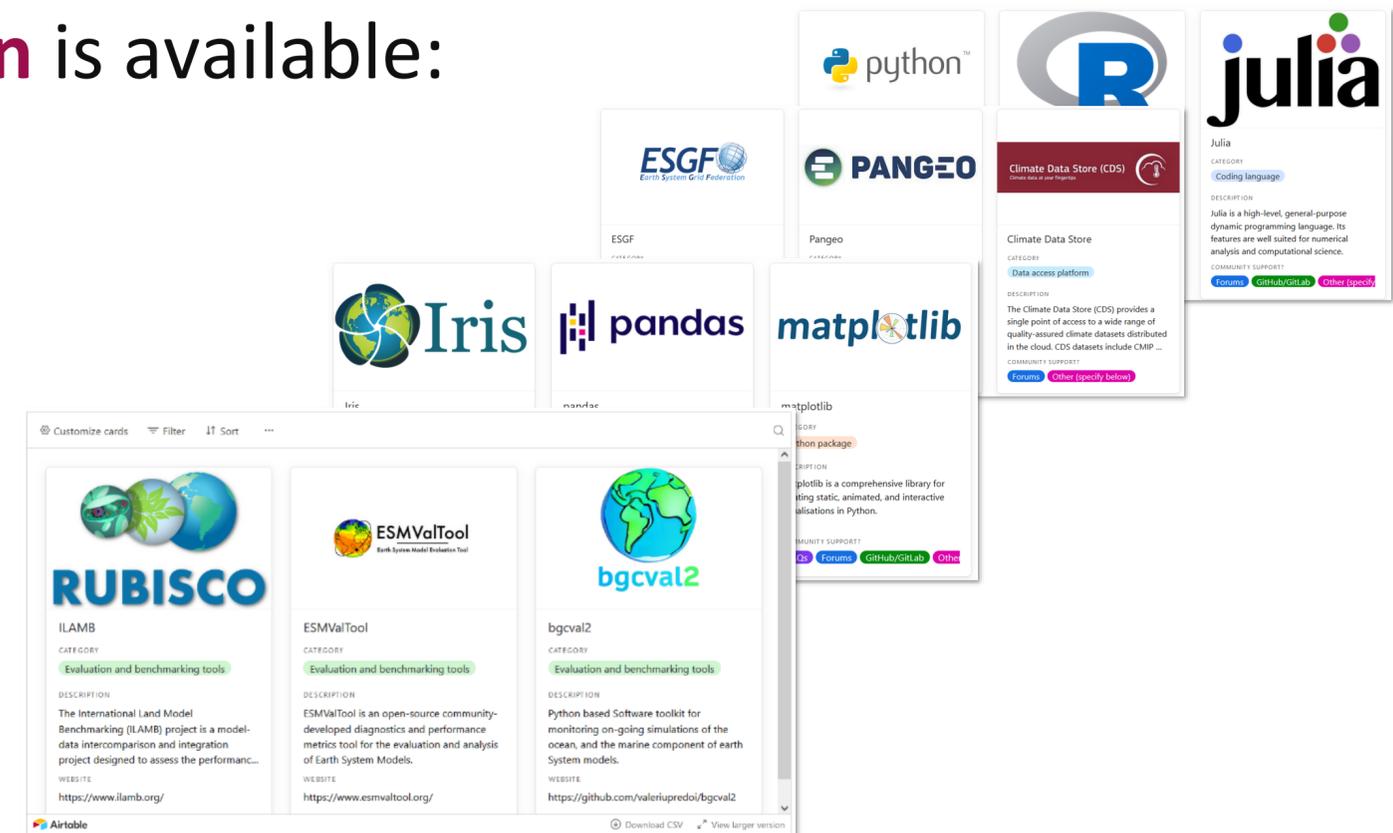
Collaboration with two **Fresh Eyes on CMIP** Subgroups

- Model Evaluation
- Data Analysis



Getting started with benchmarking and evaluation tools

- CMIP International Project Office has started to collect and host information about **open-source data analysis tools** on their website
- For each tool, the following **information** is available:
 - tool category;
 - tool description;
 - types of available community support;
 - tool website link;
 - link to tool documentation;
 - links to tool tutorials which may be useful; and
 - links to access any available community support.



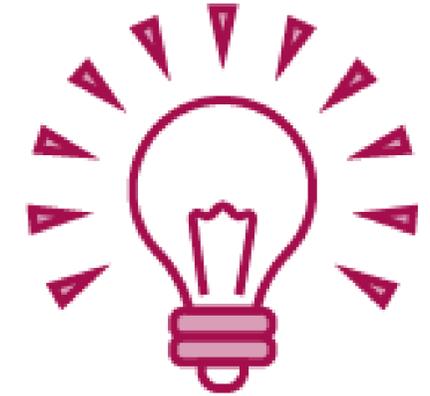


Retrospective paper

- **Definitions** of “evaluation”, “validation”, and “benchmarking”
- Retrospective look at **evolution of evaluation & benchmarking metrics**
- What tools were **available** for CMIP6 (methods, philosophies, tools)?
- What approaches were **used** for CMIP6?
- Which of them **worked well** for CMIP6 and what **did not work** for CMIP6?
- **Extensive information** about different benchmarking and evaluation tools

Status: Currently being finalized

Planned submission: August 2024



What is the way forward?

Based on the findings of the extensive information collected about different tools, and the retrospective paper – What do we think should be the **benchmarking/evaluation focus for CMIP7?**

- What **framework** would ideally be available for instantaneous benchmarking and evaluation at the time of data submission? Is such a framework even possible?
- How to **avoid the bottlenecks** encountered in CMIP6 benchmarking/ evaluation?
- Comprehensive community evaluation in **near-real time** possible?

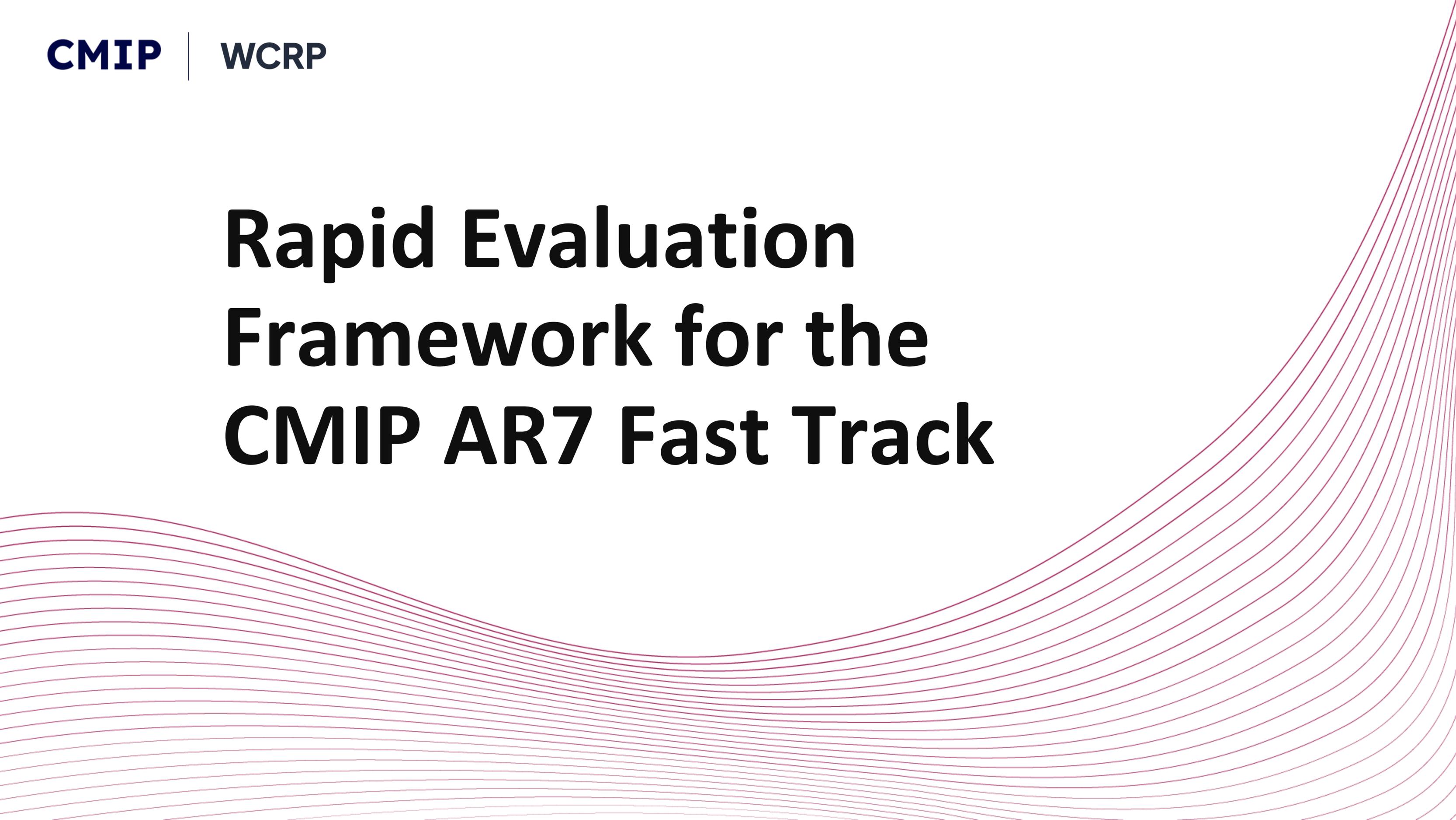
Status: Under development

Planned submission: Autumn 2024

CMIP

WCRP

Rapid Evaluation Framework for the CMIP AR7 Fast Track



Mapping the components

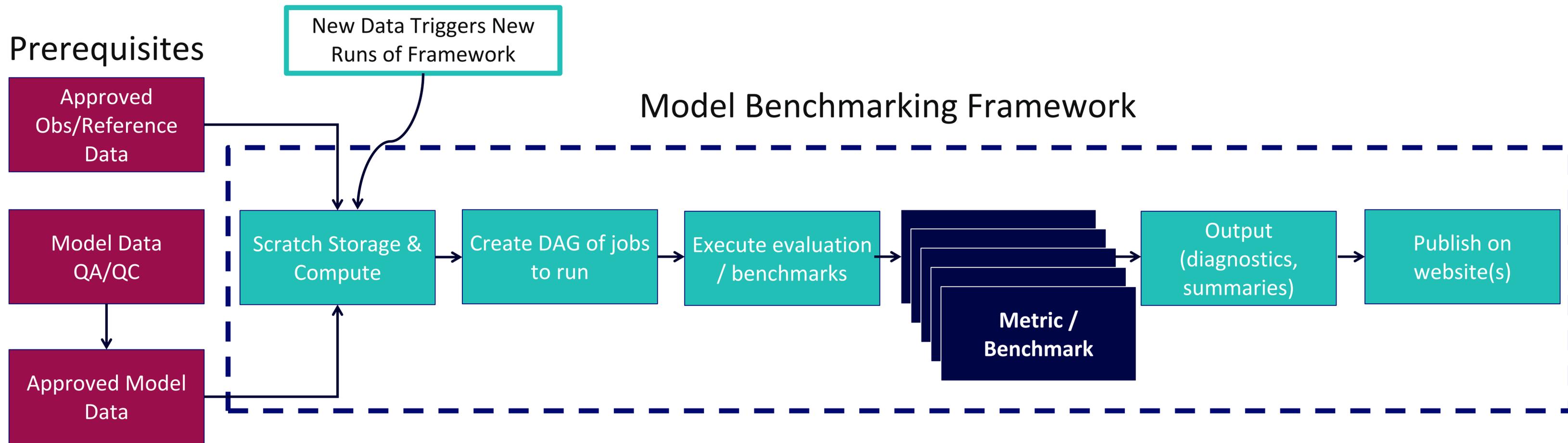
For each component required in the framework the task team are identifying the:

- ESSENTIAL
- MINIMAL
- DESIRABLE

We are then identifying the combination of component levels required for a minimal and optimal framework to meet the needs of the AR7 Fast Track.

We are also considering how this framework could have community benefit beyond the AR7 Fast track

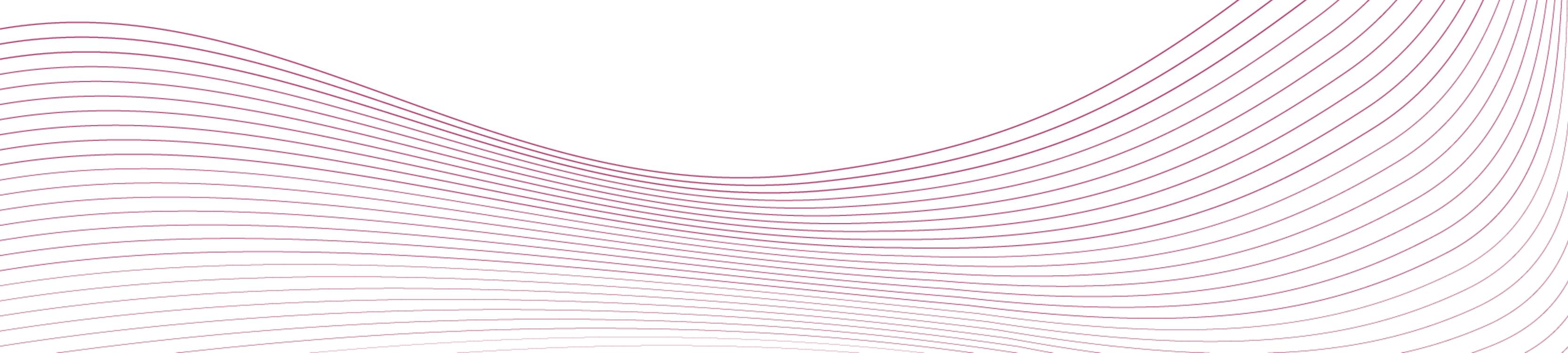
Rapid Evaluation Framework Overview



CMIP

| **WCRP**

Dreaming big

The bottom half of the slide features a decorative graphic consisting of numerous thin, wavy lines in a reddish-pink color. These lines flow from the left side towards the right, where they curve upwards and become more densely packed, creating a sense of movement and depth.

Science

- New diagnostics to reflect on the developments of the models (high-res, AI-based simulations/models, new model components, large ensembles...)
- Thinking about model weighting/excluding outliers to reduce uncertainties on projections
- Considering possibility to adjust quickly to provide metrics for new and emerging topics
- Sustainability of the effort

Technical I

- Framework for automatic evaluation
- Interactive website for the results
- Access to pre-processed data
- Observations (e.g. for benchmarking) alongside the model data
- Improved documentation on all important aspects
- Consistent output for derived variables

Technical II

- Data quality check and assurance before publishing (e.g. on ESGF side, on modeling center side, somewhere in between...)
- Easy data availability and accessibility for everybody (e.g. no need to download the data for analyses...)
- Pre-organized data licensing for easy use (open data preferred)

Strategic

- Funding
- Reduce burden on modelling centres (e.g. with the framework, the different benchmarking tools, etc.)
- Rapid evaluation of different experiments possible (e.g. beyond the historical simulations...)
- Make it easy to use the results of the framework to help decide which model (simulations) to use for your purposes.

Get involved!

Catch up on our CMIP Model Benchmarking TT **face-to-face meeting in May** at the German Aerospace Center (DLR)

<https://wcrp-cmip.org/model-benchmarking-task-team-meet-to-advance-rapid-evaluation-tool/>

Join **Fresh Eyes on CMIP** groups -we are working with members to:

- Develop scope for better quality assurance / quality control (QA/QC) for CMIP model output
- Develop a white paper on data needs for model benchmarking, including uncertainties