

# Rapid Evaluation Framework

# Housekeeping

- Abide by the WCRP code of conduct
- This meeting is being recorded and will be uploaded to YouTube – if you do not wish to appear in the video do not turn your video on.
- Minimise data usage – during presentation, only the presenter should have their video on.
- Raise hand to ask question or place in slido –we will go to slido first

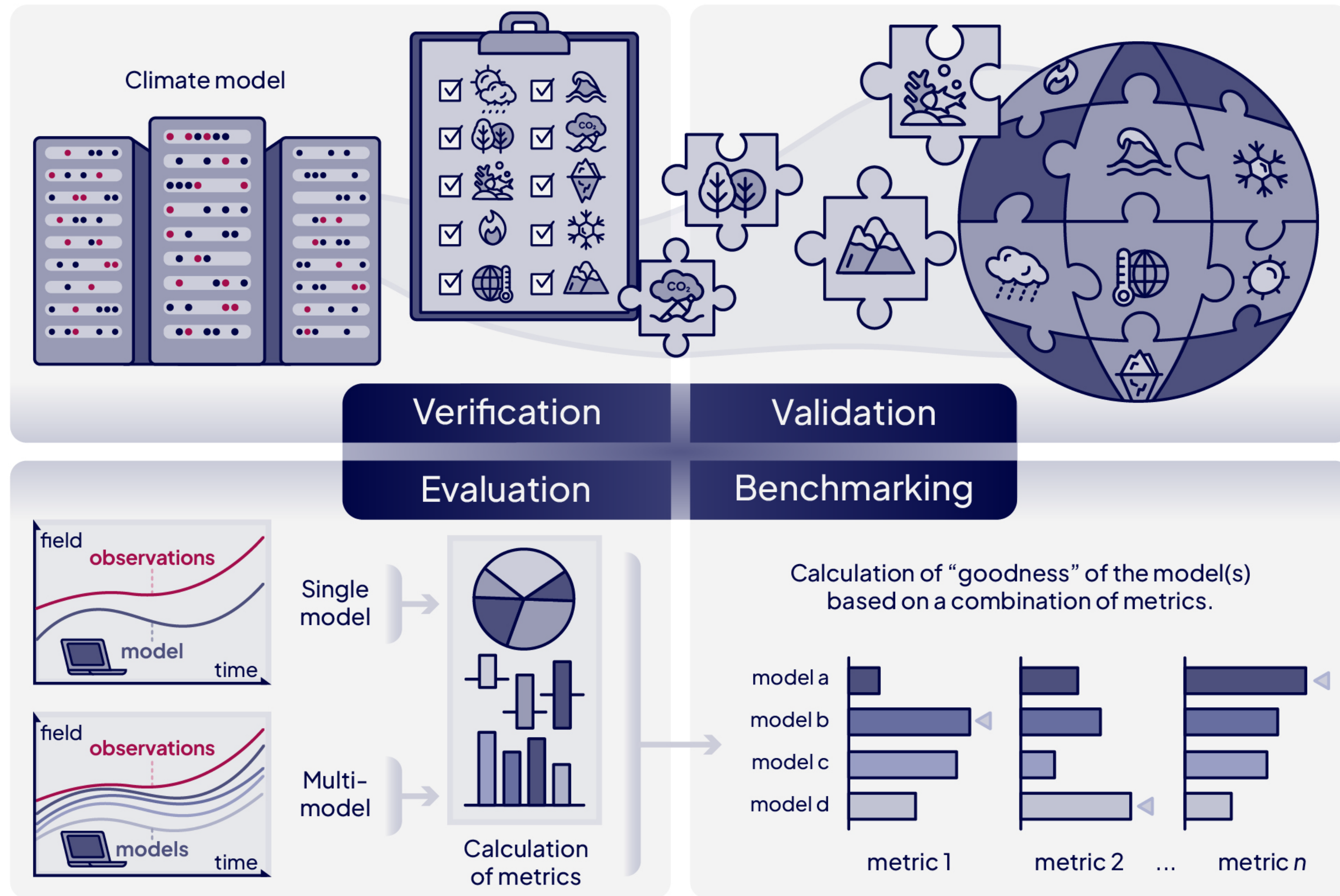
# **Session 1**

# **Vision and**

# **Implementation Plan**

# **Overview**

# Importance of model evaluation and benchmarking



As climate models become increasingly complex, there is a need for systematic and comprehensive evaluation and benchmarking through comparison with best-available observational data to assess model fidelity across a wide range of variables.

## With thanks to...

- Acknowledgement of funders



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# History of the Rapid Evaluation Framework

- Objective of the Model Benchmarking Task Team
- Conceptual design was developed at a workshop held at DLR in May 2024
- Approved by CMIP Panel in July 2024

# Consultation to date

**May 2024** Model Evaluation Survey – analysed by FEoC model evaluation sub group.

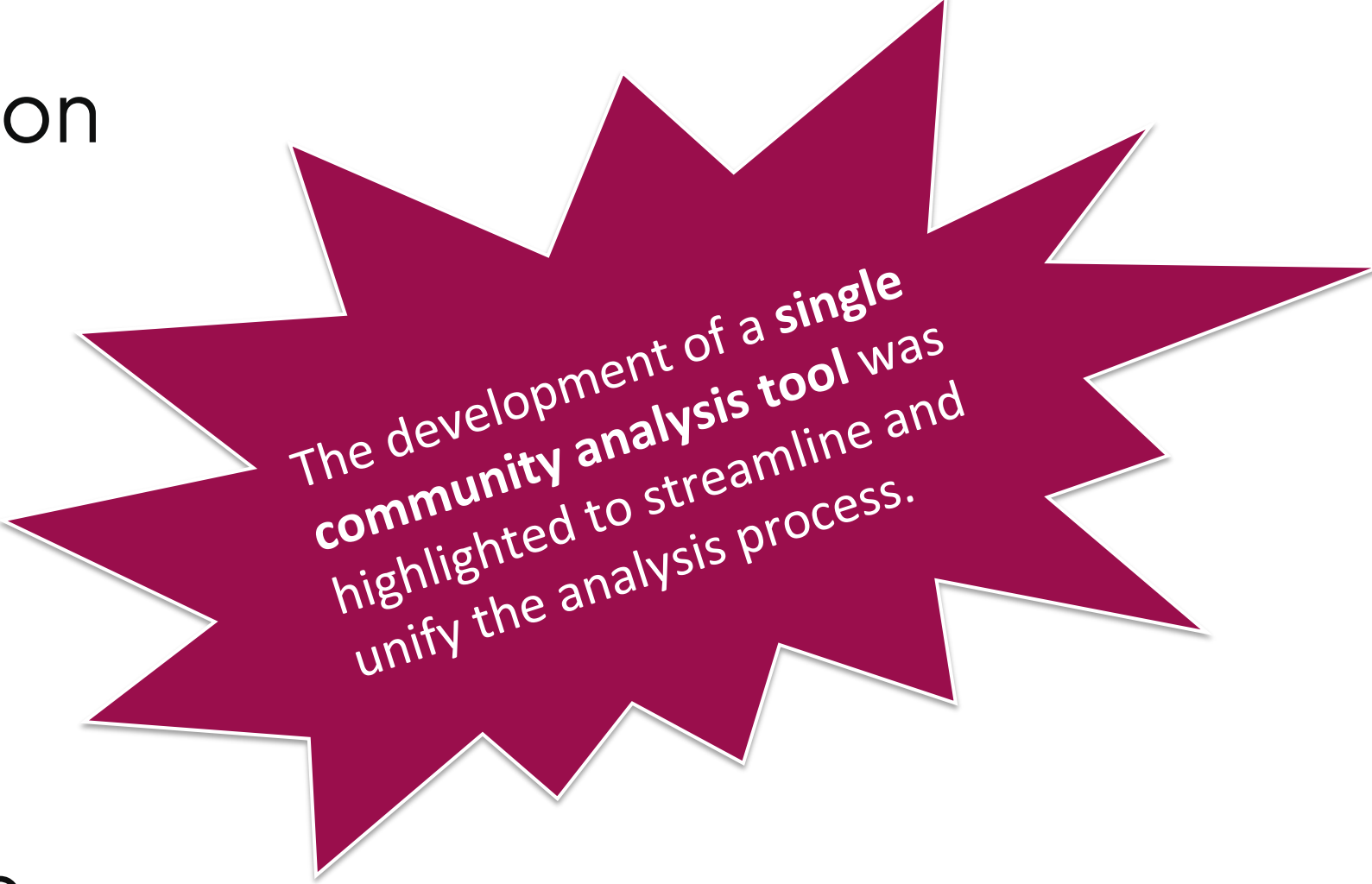
Three areas of focus:

1. how did respondents engage with CMIP?
2. What was their satisfaction with CMIP models & observations?
3. What are their aspirations for model attributes & benchmarking

This provided context for the May workshop where the propose for the REF was initiated.

**September 2024** Diagnostics Survey – analysed by IPO and FEoC members

**October 2024** Observations for diagnostics session at ESA Climate Modellers Users Group Co-location meeting



The development of a **single community analysis tool** was highlighted to streamline and unify the analysis process.

# Overview

- **Vision:** A community owned evaluation framework, built upon, and compatible with, existing community evaluation packages that incorporates an application programming interface (API) for executing metrics generation from those community evaluation packages, across the globe.
- **Goal:** The Rapid Evaluation Framework (REF) is a complete end to end system providing a systematic and rapid performance assessment of the expected models participating in the CMIP AR7 Fast Track, supporting the next IPCC Assessment Report 7 (AR7) cycle
- **Outcome:** The REF provides ability to fully integrate evaluation tools into the CMIP publication workflow, and their diagnostic outputs published alongside the model output on the Earth System Grid Federation (ESGF) through an easily accessible website.
- **Wider community use beyond CMIP:** The REF is designed to be a starting point for the community to develop and build upon, with applications across the WCRP Modelling Multiverse.
- **Diagnostics for CMIP AR7 Fasttrack model evaluation:** The CMIP Model Benchmarking Task Team through consultation with the CMIP modelling community, have collected a set of informative diagnostics for five different climate model themes. These will be announced this afternoon.

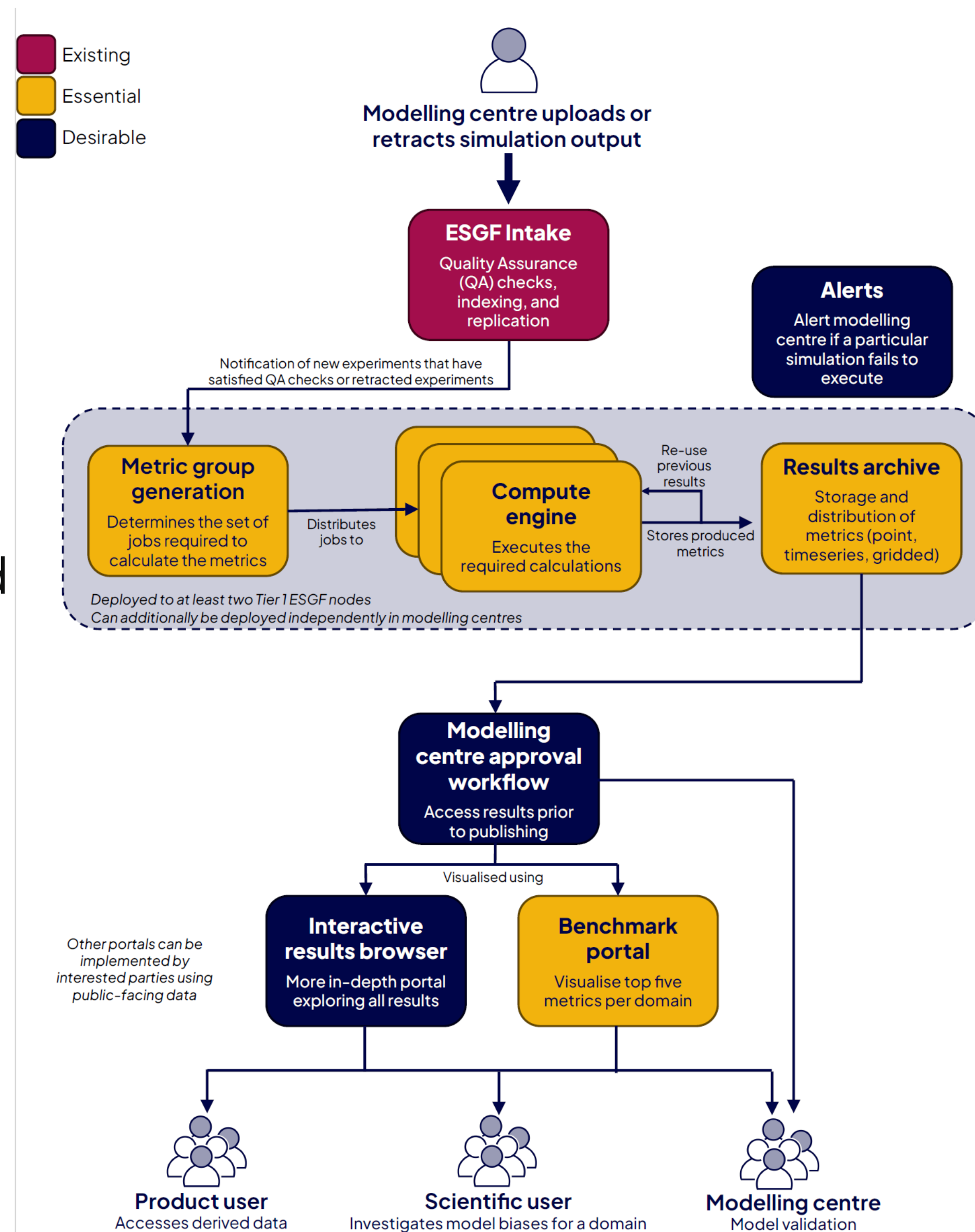


# Overview continued

The REF will be:

- **Open source**, capable of being used by, and further expanded by the model benchmarking and evaluation community
- **Open access**, available both online and through a portable containerised system for use by modelling centres in-house

It will only be run for experiments that have satisfied QA checks via the ESGF QA process.



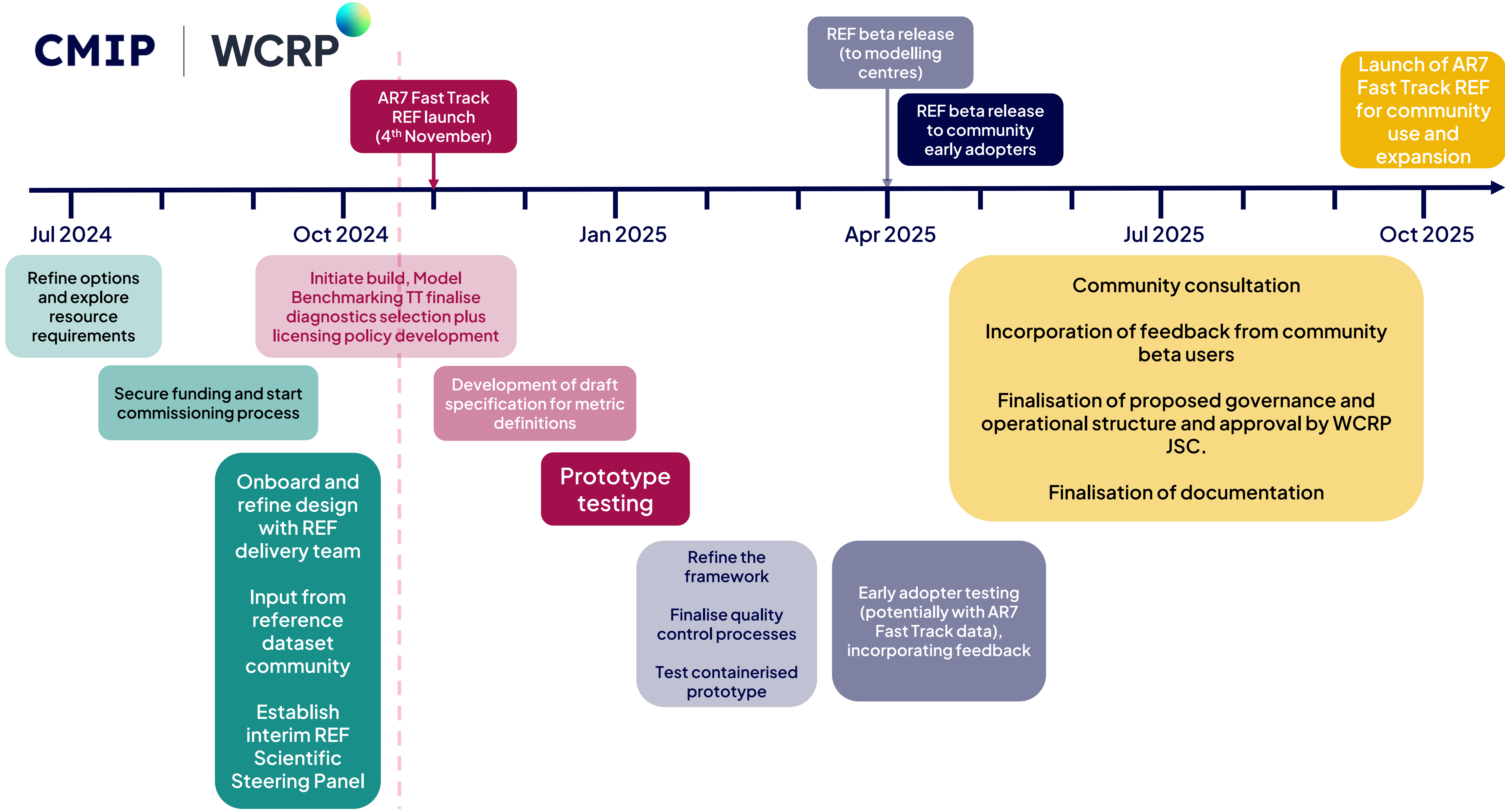
## Current activities

- Funding secured for essential components for AR7
- Welcomed by the community – very positive response in the survey (September 2024)
- REF Benchmarking opportunity has been submitted to the Data Request and peer reviewed by all thematic author teams
- Diagnostics list for AR7 FT being announced this afternoon
- Community are encouraged to consider incorporating REF add-ons for enhancements in funding proposals.

# Timeline

# Milestones for REF development

- **Milestone 1** – Community engagement to finalize the implemented metrics and diagnostics in a minimal version of the REF – **completed, will be launched 4 November**
  - **Milestone 2** – Provide recommendations for enhanced QA/QC package(s)
  - **Milestone 3** – Prototype workflow across at least two participating ESGF nodes and test containerised version with at least 3 modelling centres
  - **Milestone 4**– Governance and operational structure for future evolution presented to and approved by WCRP JSC
  - **Milestone 5**– Publication (peer reviewed) after the REF is in place and tested
- **March 2025** Prototype operational and undergoes stress testing
  - **May 2025**– Beta portal available for early adopter modelling centres
  - **October 2025**– Public release of the metrics portal and REF for community evolution.



# Work packages

# Work packages

- CMIP IPO
- DOE
- NLeSC (Netherlands eScience Center)
- CR (Climate Resource)
- MBTT (Model Benchmarking Task Team)

**RACI**  
 R: Responsible (Does the work)  
 A: Accountable (Ultimately responsible)  
 C: Consulted (Provides input)  
 I: Informed (Kept in the loop)

In the MVP, WP1 would develop the engine to execute metric calculations

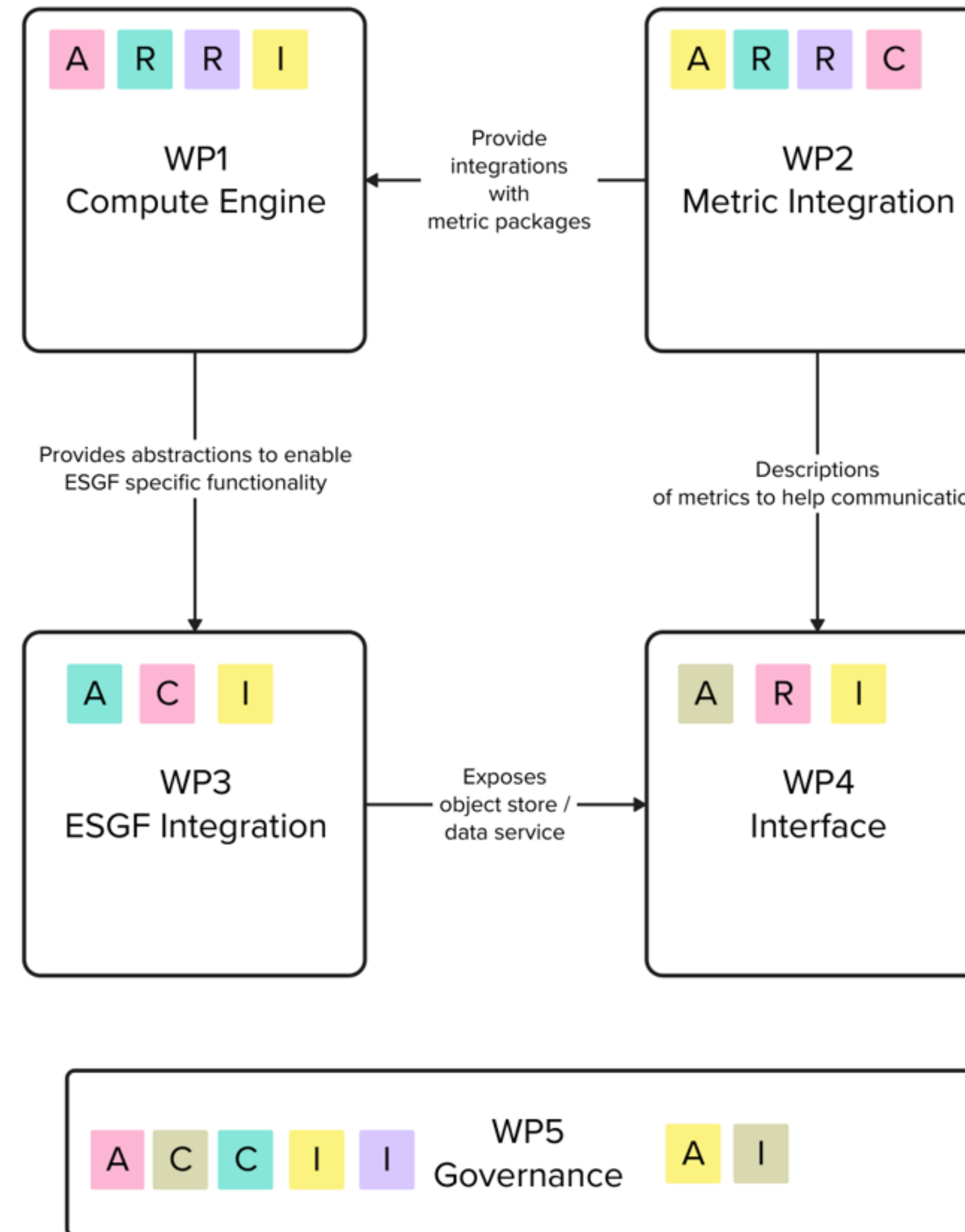
- Development of Open Source package that can update
- Provides an abstraction for running tasks that enables multiple target environments
- Includes a PoC local runner for testing/modelling centres

In the MVP, WP3 would createan ESGF deployment of the REF

- Development of required configuration for deployment to ESGF (K8 Helm chart/Ansible deployment)
- Production deployment of workers, event queue, object store and job database

Technical Director: Jared Lewis

Oversee the technical implementation of the REF  
 Define specification document that describes the metric calculation  
 Reports to the Model Benchmarking TT



In the MVP, WP2 would define howthe metrics are calculated

- Develop metric integrations for key packages using the CMEC driver
- Develop integration for timeseries calculations
- Define the CMIP AR7 Fast Track configuration
- Provide assistance packaging metrics into job environment

In the MVP, WP4 would build an portal to interrogate the results

- Benchmark portal for a select set of metrics
- Built on top of results served from ESGF

Science: Birgit Hassler/Forrest Hoffman in coordination with the CMIP MBTT

- Define of the user requirements of the REF (done)
- Decide on the list of metrics to implement as part of CMIP AR7 Fast Track (In progress)

# Communicating evaluation outcomes

Sign up to the Model benchmarking mailing list

<https://wcrp-cmip.org/get-involved/>



# Q&A

# Thank You



@wcrpcmip



wcrp-cmip



cmip-ipo@esa.int

<https://bit.ly/CMIP-event-feedback>

<https://wcrp-cmip.org/get-involved>