CMIP7 Planning
WIP options for the future
CMIP6+/EXT (Covid-MIP), CMIP7

Matthew Mizielinski, Paul Durack & Karl Taylor

With contributions from the WGCM infrastructure panel

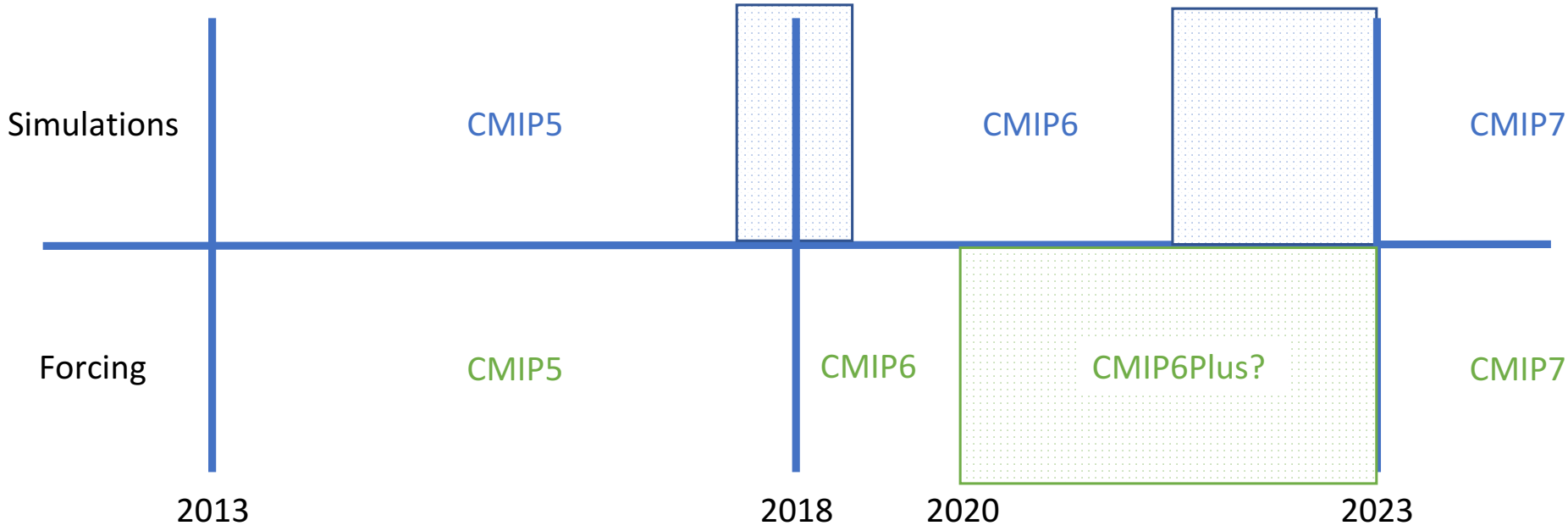
WGCM23 meeting, 9th December 2020

Contents

- Motivations for CMIP6+
  - Forcing development timeline
  - Covid-MIP experience
- Reminder of CMIP6 infrastructure
- CMIP6+ challenges & benefits
  - Infrastructure components
- CMIP6+ summary
- CMIP7

Forcing development timeline

A forcing evolution following the continuous CMIP DECK paradigm



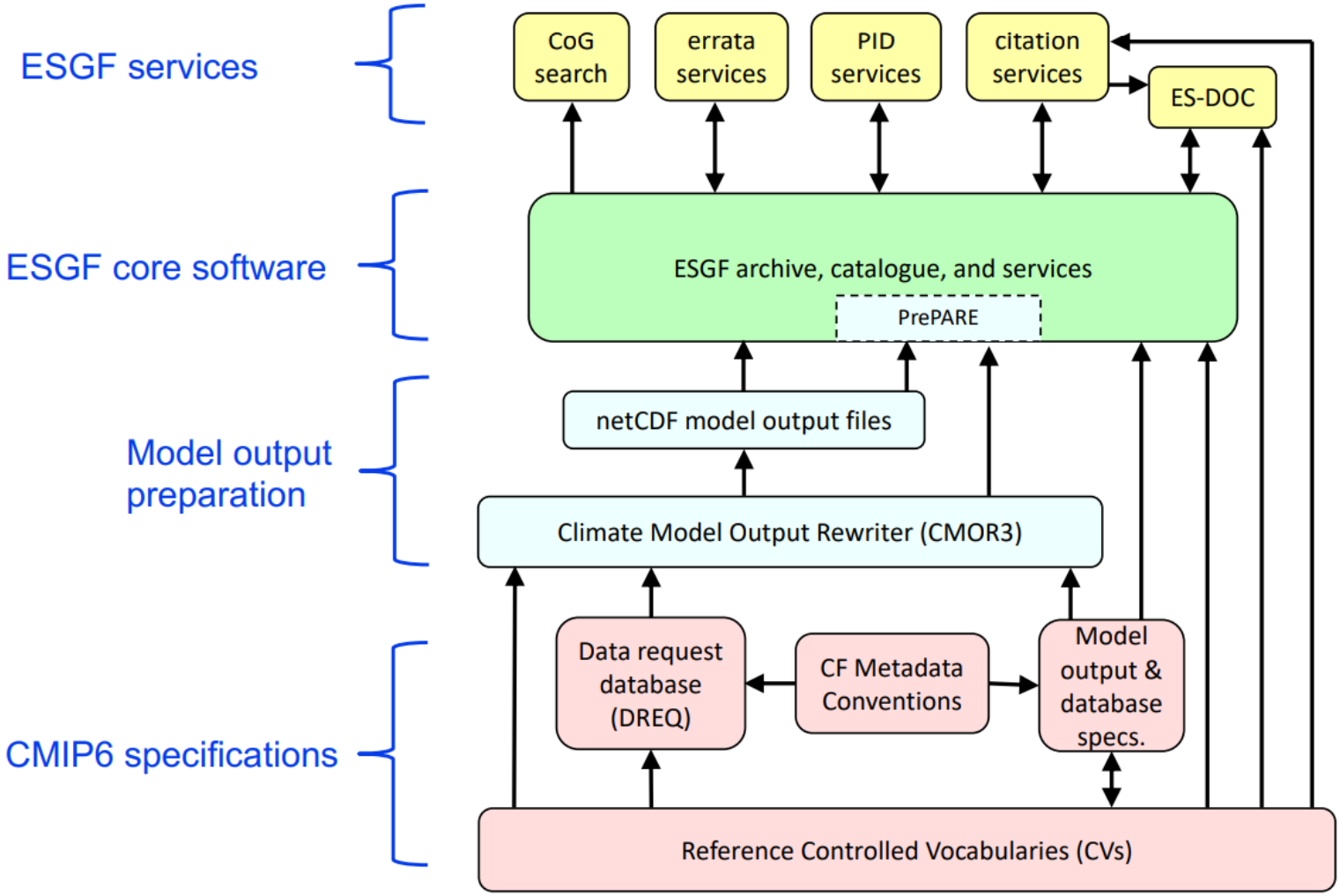
Transition period between MIP-era model simulations

Transition period between MIP-era forcing datasets (much broader, as prototype datasets need iteration before "formal" model simulations are contributed)

Adding a new MIP: The Covid-MIP experience

- Agreement reached on adding 6 new “Covid-MIP” experiments to DAMIP
- Controlled vocabularies updated with the new entries
- Data request updated with new experiments set up as “clones” of existing ones (same variables as DAMIP experiment hist-aer)
- ES-Doc experiment descriptions added based on pre-existing information for other experiments
- Updates pulled through into CMOR tables and ESGF configuration files
- Citation system updated where new institution, model, MIP combinations
- Data published to ESGF

Infrastructure components and dependencies



CMIP6+: a new MIP era?

Challenges

- New funding required for some pieces of existing infrastructure
- Keeping systems relevant and running requires continual support
- Many single points of responsibility leading to little redundancy; fragile resiliency, potential bottlenecks and/or single points of failure

Benefits

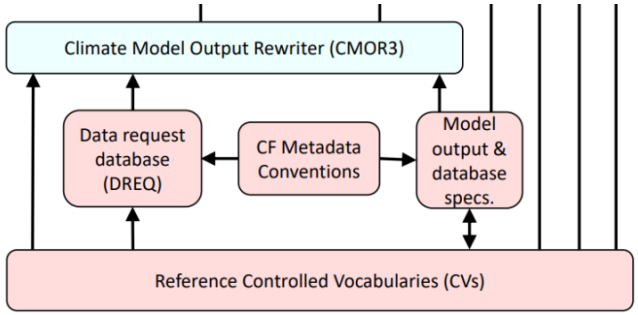
- Exploit investment
- Facilitate faster model and forcing development
- Adaptive infrastructure can enable new projects and experiments
- Allow smooth progression towards CMIP7, with opportunity for continuous integration testing

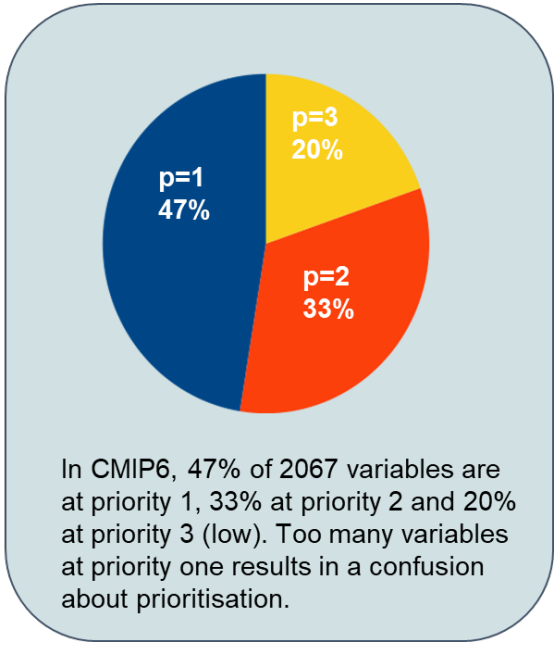
CMIP6+: Infrastructure

- New activities will need to agree to meet resource requirements with data hosting institutions / ESGF nodes
- Include a data management plan – how long does this data need to be available for; 5 years?
- Clear, well documented procedure and requirements;
  - Sufficient models/groups involved to be worthwhile hosting?
- Reliant on long term support and development of ESGF and the services it depends on
- More integration between different components and automation where possible

Infrastructure components I

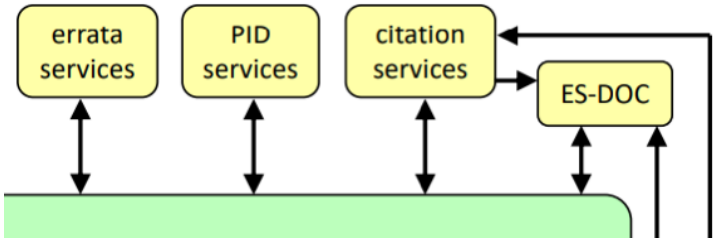
- CVs: Parallel set of CMIP6+ CVs based on current structure
- Data request:
  - Simplified structure with more templating
  - Retain existing variable definitions
  - Address grade inflation in prioritisation of variables
- CMOR & MIP tables
  - Build on existing tools
  - PrePARE tools are key for current QC (used by ESGF)
  - Consider structural harmonisation with other projects (CORDEX, Copernicus?)

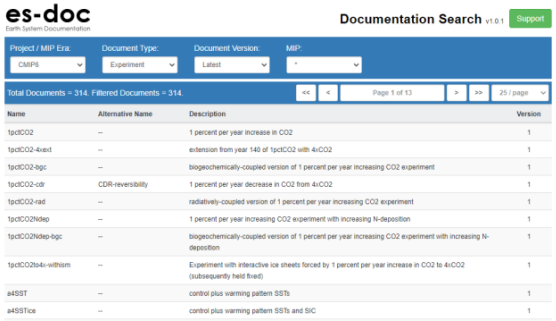




Infrastructure components II

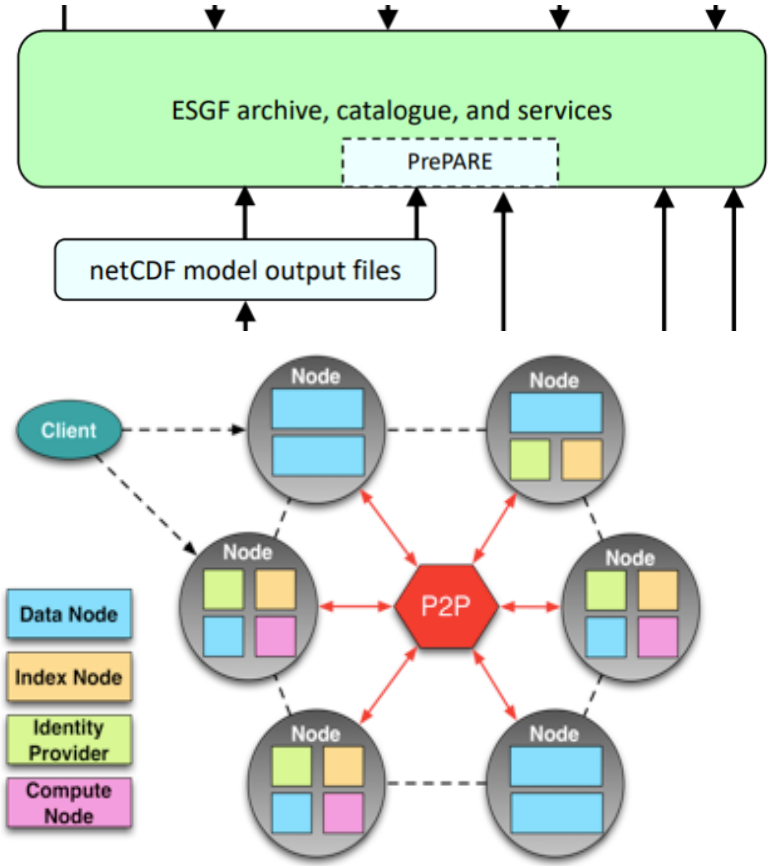
- Citation and PID services
  - Integrate citation and publication services
  - Bring more information directly to the user via integration with ESGF search portal
- ES-Doc & Errata
  - Experiment definitions; further integration with change process for CVs
  - Errata system critical for science quality assurance
    - Can we develop greater integration with ESGF?
    - User reported errata
  - Model descriptions – what is the best way of collecting and maintaining this information





Infrastructure components III

- ESGF
  - New web frontend: “MetaGrid” in testing
  - Centralised search interface
  - Common tools for ESGF compute and expand user base
  - A host of technical improvements and integration with new technologies (e.g. containerisation) to enhance scalability and stability



ESGF Future Architecture report, July 2020, doi:10.5281/zenodo.3928223

CMIP6/6+ Summary

- CMIP6
  - Data to grow until at least December 2022
  - Potential to exceed existing ESGF federation storage pool (> 30 PB?)
- CMIP6+
  - Extend CMIP6 approach to become an “operational” service
  - Leverages existing investments (for infrastructure and modelling groups)
  - Foster forcing dataset development alongside models
  - Accommodate new science (MIPs)
- Principle limitation is insecurity of resources and funding
  - “Institutional memory” is hard to maintain
  - Loss of key projects, people and resources continues to occur
  - Research funding paradigm limits continuity

CMIP7 concerns & requirements

- What is the CMIP7 science plan?
  - Are modeling groups supportive of a CMIP6-scale project?
  - Another order of magnitude increase in data volume would require major changes in infrastructure.
- The infrastructure's single points of failure need shoring up.
- The current infrastructure has become essential to the climate research community:
  - It underpins CMIP, input4MIPs, obs4MIPs, etc.
  - Modeling and analysis groups have invested in the current infrastructure rely on it beyond CMIP6. Continuity is important.
- Underlying data technologies must continually evolve, which requires long term support

Aspirations for CMIP7

- Maintenance of effort/FTEs across existing CMIP6 partners
  - Load-balancing across partnering institutions
  - Anticipate funding “holes” (IS-ENES ends 2022)
- Build on flexibility in place (presumably) for CMIP6+
  - A more nimble infrastructure requires long-term/operational support, not characteristic of the “research funding” paradigm
- A CMIP “cloud”?
  - Private cloud (not Amazon/Google) - costs are defined
  - Storage for most data - federated clouds (DOE, DKRZ, CEDA, IPSL, ..)
  - Compute for analysts alongside the data
  - No download needed
- Common tools
  - Regridding, computing climatologies
  - Model evaluation