CMIP
Coupled Model Intercomparison Project

Priority variables for evaluation and exploitation of WCRP climate simulations

Workshop Report.

June 2022 CMIP IPO Publication No.: 01/2022

Bibliographic information

This report should be cited as: O'Rourke, E. and Turner, B. (eds). 2022:
Priority variables for evaluation and exploitation of WCRP climate simulations workshop report
Top priority variables community workshop - 12 and 17 May 2022
CMIP International Project Office, Harwell, UK.

Contact information

All enquiries regarding this report should be directed to CMIP-IPO@esa.int

Authorship and publisher's notice

This report was authored by Eleanor O'Rourke and Briony Turner, with the Foreword provided by Martin Juckes.

Acknowledgements

The work developing the core variables is being coordinated and led by Martin Juckes, UKRI-STFC. Its implementation is being supported by the CMIP International Project Office and is made possible by funding from IS-ENES3 part of the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084.

Disclaimer

The designations employed in the WCRP Coupled Model Intercomparison Project, or CMIP, publications and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of CMIP, the World Climate Research Programme (WCRP) – including its Sponsor Organizations the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of UNESCO and the International Science Council (ISC) – or the European Space Agency (ESA) in its role as host organisation of the CMIP International Project Office (CMIP IPO), concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The findings, interpretations and conclusions expressed in CMIP publications with named authors are those of the authors alone and do not necessarily reflect those of CMIP, WCRP, its Sponsor Organizations, ESA or of their Members.

This document is not an official publication of the WCRP and has been issued without formal editing. The views expressed herein do not necessarily have the endorsement of WCRP or its Members. Any potential mention of specific companies or products does not imply that they are endorsed or recommended in preference to others of a similar nature which are not mentioned or advertised.

Foreword

Martin Juckes, CMIP Data request lead. WGCM Infrastructure Panel Member

Head of Centre for Environmental Data Analysis for the Atmosphere (CEDA-Atmosphere) and deputy head of CEDA

This report offers a comprehensive summary of two community workshops convened by the CMIP International Project Office, on May 12th and 17th 2022, to discuss top priority variables in the context of the international exchange of climate simulations for evaluation and exploitation. We are in an exciting era where the use of climate simulations in the science community and beyond is evolving and expanding at a rapid pace. The community effort supporting evaluation and exploitation of these simulations is an enormous effort spread across hundreds, if not thousands, of institutions. There are concerns that current ways of working need to evolve, and a good diversity of opinion on ways forward.

As chair of both workshops, I would like to reflect on the key outcomes of the consultation and welcome such a successful start to the process.

The Phase 6 of the Coupled Model Intercomparison Project (CMIP6) has followed in the steps of previous phases in delivering huge successes (Eyring, 2020), even at this stage, when the scientific analysis of the immense archive of data is just starting.

There are, nevertheless, many frustrations about the level of stress experienced by many who signed up to the community effort and found themselves subjected to unexpected pressures.
Many ideas are circulating within the climate modelling and exploitation community about future directions of efforts to exchange, evaluate and exploit climate model data. The emergence of a scientific community based on exploitation of climate model data is driving many of these ideas. The concept of "enforcement" often emerges in various forms, but it is important to remember that the success of CMIP is, in a very fundamental sense, grounded in voluntary participation. There are many reasons to consider a change to an operational system with enforceable standards, but, in the meantime, we need to explore how to make best use of the current approach.

One of the recurrent themes in the discussions leading up to this workshop was the idea that one sector of the CMIP community had unreasonable expectations of another, such as those on the modelling side suggesting, perhaps, that scientists are requesting data without taking due account of the effort needed to produce and process it and those on the data exploitation side suggesting, perhaps, that the data providers do not understand the obstacles and inconvenience created by avoidable irregularities in the data. As implied by the caveats, the evidence for these problems is not on the record. These workshops and the effort to take steps both to enhance the level of consistency in the archive and to reduce the stress imposed on data providers is rooted in the assumption that all

parties have real problems that need to be addressed and that there is a strong desire on both sides to address these problems.

I was particularly encouraged by the quality of the discussions in both workshops and by the clear willingness of delegates from a wide range of specialisms to work together to support a strong transdisciplinary approach (i.e., an approach which not only combines the interests and resources but also reaches beyond the scientific disciplines). I look forward to working with the appointed author team to build on this strong start so that we can both address many of the challenges faced to date and start the process of building a robust framework to face the data-sharing challenges which will come with the next cycle of WCRP climate simulations.

Contents

I. Executive Summary

The Coupled Model Intercomparison Project (CMIP) has grown considerably and now serves a wide range of communities, all with their own specialised requirements for data. The CMIP Data Request function, led by Martin Juckes of UKRI-STFC, is establishing a process to address the challenges presented by having too many variables listed as top priority while meeting the needs of both data providers and users. It is envisaged that a core set of variables can form a baseline for exchange of climate model data, in any intercomparison project, in accordance with FAIR data and Open Science principles. Establishing this baseline will address the community intention discussed at WGCM 2019 in Barcelona of giving more authority and meaning to variable prioritisation. The intention is to publish these as a Geoscientific Model Development (GMD) paper.

In spring 2022, members of the CMIP community were engaged in a survey and two workshops on the proposed methodological approach and a paper publication process to devise an agreed list of core variables. The two workshops, held in May 2022, were based upon the 32 responses to the survey issued by the CMIP IPO in April 2022 to Modelling Centre and Data Request Leads, and the MIP Chairs.

The objectives of the workshop were to:

1. Review objectives of the paper.
2. Agree on a process of expert elicitation to identify a list of top priority variables which reflects community demands for a substantial reduction relative to CMIP6.
3. Identify an initial list of authors and reviewers.
4. Agree on a time frame and process for drafting the paper.

There were 31 contributors to the workshop, available to participate online and in their own time. The key outcomes were:

- Broad support for the proposed objective to establish an agreed list of core variables to be archived for exchange of climate model data.
- Early engagement is vital for the planned expert elicitation process. Further consideration required as to whether to take a top-down or bottom-up approach to expert elicitation.
- Inclusion of an additional first task for authors to refine the objective, particularly in establishing clarity on purpose, function and intended (current and future) users of the core set of variables.
- Authors will need to consider and make recommendations on proposed implementation, in close collaboration with key stakeholders (WCRP, WGCM, modelling centres, MIP chairs etc.).
- Variable selection criteria are contentious and will require further iterative community engagement. Suggestions made of consideration of a matrix rather than list approach and further thought regarding the prioritisation approach regarding issues such as potential bias and choice of metrics for variable selection criteria and how to serve a wide range of users.

The proposed next steps and publication timetable was acceptable, and a short list of authors has been established.

II. Introduction

1.1 Context

The Coupled Model Intercomparison Project (CMIP) coordinates the design and distribution of global climate model simulations of the past, current, and future climate (Eyring et al., 2016). There is a centrally coordinated data request which for CMIP6 defines all the quantities from CMIP6 simulations that should be archived (Juckes et al., 2020). This includes data request requirements from all the CMIP endorsed MIPs.

There are over 300 experiments within CMIP6 with distinct sets of highly tailored variables.
All Model Intercomparison Projects (MIPs) define output streams in the centrally coordinated CMIP6 data request for each of their own experiments as well as the DECK and CMIP6 historical simulations so that that the required variables are stored at the frequency and resolution required to address the specific science questions and evaluation needs of each MIP and to enable a broad characterization of the performance of the CMIP6 models (Eyring et al., 2016).

Martin Juckes highlighted the issue of consistency in variable output using the example of the CMIP historical simulation. In Figure 1 the rank 1 model (IPSL-CM6A-LR) has 673 archived priority 1 variables, and the rank 20 model (AWI-ESM-1-1-LR) has 346 archived priority 1 variables; the intersection of variables output by the first 20 models is only 53. The level of consistency is even further reduced when considering multiple experiments. By identifying a set of core variables and associated metadata this will enable consistent and efficient comparison of simulations across multiple intercomparison projects, reducing the workload for data providers and users by providing a reusable basic set of variables. The aim is for 90% of models to provide 90% of variables. In comparison 28% of models provide 28% of CMIP6 priority 1 variables, 90% of models provide 7.8% or more, and just two models provide 50% or more.

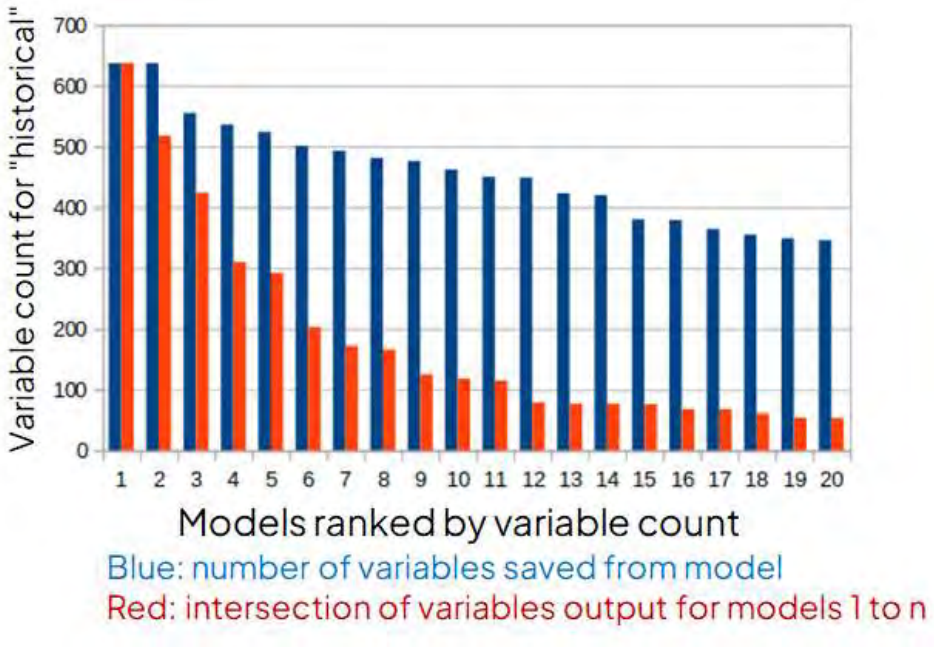


Figure 1: CMIP variable counts per model

Prioritisation

The priority of a variable is an indication of the importance of that output for that variable from a simulation and for a specific objective. A single variable may have different levels of importance for different simulations and different objectives (Juckes. 2020). The prioritisation is intended to guide contributing modelling centres with the intention that all modelling centres should provide the highest priority variables so that data users can benefit from a uniform selection of variables.

In CMIP5, each requested parameter was assigned a priority from 1 (high) to 3 (low), and this priority applied to requests for that variable from all CMIP5 experiments (Juckes et al., 2020). During CMIP6, there was greater flexibility which caused confusion about the interpretation of request priorities; modelling groups had the choice of which MIPs to support. Consequently, modelling groups needed to know how important the variables requested were for the MIPs that they elected to support (Juckes. 2020).

The recent CMIP Community Survey, carried out in early 2022, for which over 300 responses were received, did contain comments suggesting there were too many core, or priority 1, variables. However, responses also suggested a need for additional variables including:

- Increased temporal resolution.
- More ocean variables.
- Variables relevant to extremes.
- Variables required for CORDEX/regional downscaling.

The challenge

There is widespread agreement that there are too many variables being listed as top priority; however, there is no clear consensus regarding which, and how many, should be considered as highest priority. Given the ever-wider range of CMIP data users, each with their own specialised requirements, a transparent and documented approach is required.

As discussed at WGCM 2019 in Barcelona, there is community intention to reduce the number of variables at priority 1 from around 50% to a significantly smaller number, perhaps starting with those prioritised by AR6 WG1 (Juckes, 2020).

The CMIP Data Request function wishes to address the immediate challenge of establishing an agreed variable prioritisation methodology from the CMIP modelling community and an approach for giving authority to "priority = 1" statements. It is envisaged that these prioritised variables can form a baseline set of variables for exchange of climate model data, in following FAIR data and Open Science principles. The intention is to publish these methods as a Geoscientific Model Development (GMD) paper. By defining the baseline set in advance of the next CMIP phase, it should then be easier for MIPs to think about what they need in addition to a core set of variables.

1.2 Community engagement

The CMIP-IPO is supporting the CMIP Data Request function to establish and publish an appropriate methodology for prioritising variables that could be considered as a baseline set of variables for exchange of climate model data, in any intercomparison project, in accordance with FAIR data and Open Science principles. The activity is being led by Martin Juckes STFC-UKRI[1].

Survey

A community survey was issued by CMIP IPO in April 2022 to the Modelling Centre leads, Data Request Leads and the MIP Chairs. The survey invited community members to express interest in being a paper author or reviewer, express interest in participating in the workshop and provided an opportunity for reflections and input on the proposed methodological approach which was produced by Martin Juckes and made available at: https://bit.ly/MIPVariables.

There were 32 respondents. A summary of the survey responses can be found in Annex 3. Highlights were presented to the workshop attendees.

[1]Contact support@ceda.ac.uk with "WCRP Core Variables" in the subject line.

Workshops

The workshop was available in two live format sessions which took place on the 12th and 17th May. Notification and registration for the workshops was circulated to all survey participants, to the Modelling Centre leads, Data Request Leads and the MIP Chairs and advertised on the WCRP community calendar.

A Miro board was used during the workshop to capture discussion points. An edited recording of the first workshop was made available, along with access to the Miro board, for participants wishing to contribute but unable to join the live workshops. These participants were requested to add their name to the 'who is contributing' section on the Board if they wished to be acknowledged within the workshop report as a contributor.

The objectives of the workshops were to:

1. Review objectives of the paper.
2. Agree on a process of expert elicitation to identify a list of top priority variables which reflects community demands for a substantial reduction relative to CMIP6.
3. Identify an initial list of authors and reviewers.
4. Agree on a time frame and process for drafting the paper.

Each workshop was chaired and led by Martin Juckes and facilitated by the CMIP IPO.
The workshop programme comprised of context setting providing an overview of the challenge, the approach proposed to initiate community discussion, the list of Expressions of Interest received for authors and reviewers, and a summary of the findings of the survey on the methodology as well as relevant points from the CMIP Community Survey. The workshop slides are available in Annex 4.

Workshop agenda

1. Chair's welcome and workshop introduction - Martin Juckes
2. Summary of community response to proposed paper methodology & summary of relevant CMIP Community Survey responses - Eleanor O'Rourke, Director, CMIP IPO
3. Plenary Session 1 - to gather thoughts and identify community agreement on the objectives of the paper.
4. Chair's reflections
5. Plenary Session 2 - Meeting attendees' direction for authors.
6. Review of meeting outcomes and closing remarks by Chair.

III. Workshop outcomes

1.3 Paper objectives and direction for authors

In the workshop introduction, Martin Juckes presented the proposed paper objectives and the paper tasks. Key outcomes are recorded below, the full list of contributions to the Miro board are available in Annex 5.

Paper objectives

The proposed objective is to establish an agreed list of core variables to be archived for exchange of climate model data. This includes associated metadata to enable consistent and efficient comparison of simulations across multiple intercomparison projects, reducing the workload for data providers and users by providing a reusable basic set of variables. It was also welcomed in terms of the consistency it would bring. There was general agreement with the objective and the overall aim of the activity. However, some participants suggested more clarity around the scope and potentially an alternative shortlisting approach.

Direction for authors

The suggested areas for the paper authors to work on regarding further refinement of the objective focused in the following areas:

- **Clarity on the purpose of this exercise** - is this a core set to ensure that basic variables are there for every run, as well as provide templates for other experiments or would core variables be restricted to a core set of experiments? Greater clarity on the purpose of the overall list rather than particular variables may require further community engagement and buy in.
- **Identification and requirements of the potential users** of these variables as this is wider than CMIP; CMIP has traditionally been focused on model development, evaluation, and science. It was suggested that identifying a list of both science areas and applications to be addressed early in the process would be useful and to consider wider WCRP requirements and work with facilitators of downstream users.
- **Clarity on function of the core variables**, this was raised by a couple of participants with specific mention of MIPs, for whom it would act as a starting point, allowing them to be consistent if they wanted to, but not limit the data they choose to produce. There will potentially be further guidance if part of, or endorsed by, CMIP7. Whilst it was recognised that the approach could help rationalise process with ability for advance variable specification, one participant raised concern that the approach for MIPs needs to be considered in parallel with how this maps on to what the modelling centres have to do.

Authors will also need to consider:

- **The implementation approach**: participants were asking whether it would be enforceable and, if so, how? Participants suggested authors will first need to ask modelling centres "can you

produce ~90% of this list regularly (where appropriate)"...and if not, they will need to establish if there is some support WCRP can give for their production.

- **Think about the future not just the present**
    - Users are not just the present users, this exercise should have regard for who the future users might be, who the community would like future users to be, including but not limited to those in the impacts community and addressing wider societal needs.
    - In addition to community review as part of the publication process, it would be important to seek endorsement from WGCM, which might require enlarging/reducing the list.
    - Will need to consider the demands on the data infrastructure and storage particularly with regards to inclusion of higher resolution (space and time) variables e.g., hourly data required for energy or hydrologically modelling.
- **Domain-specific technical considerations**:
    - Atmospheric - for 3D atmospheric variables, important to consider the information about pressure levels on which the variables would be output.
    - Surface - importance of high impact surface variables for the impacts community.
    - Marine - revising the output regarding ocean biogeochemistry could be useful.

Discussion included reflection on why the idea of a core set of variables in CMIP6 had not materialised due to time pressures and therefore that this workstream, to establish an approach with community input in advance of the next iteration of CMIP, is welcomed. It was remarked that it could potentially support process rationalisation for modelling centres but dealing with MIP specific requirements could remain challenging.

Several positive references were made to the IPCC AR6 WG1 list of variables produced by the WGCM Infrastructure Panel (WIP). Workshop participants were reminded that Martin had included this as a column in the variable list arising from the proposed methodology, and that this was an internal WIP document which had not been formally published and carried no IPCC endorsement.

Out of scope

Martin Juckes set out what he believed should be outside of the paper scope, specifying that the paper would not:

- Deal with procedures needed in CMIP7 to incorporate requirements from multiple MIPs into a consolidated request.
- Deal with technical details of metadata implementation which need to be agreed with other elements of the CMIP technical infrastructure (CVs, ESGF, ES-DOC, CMOR, Citation, PID etc.).

The exclusion of these areas was broadly supported and further out of scope areas were highlighted including the future of CMIP and any science results. The full list of contributions to this topic on the Miro board is available in Annex 5.

1.4 Paper tasks

Martin Juckes proposed five key tasks for discussion.

- **Task 1:** Describe the role of the core variables as baseline set of variables for curated data from WCRP endorsed climate modelling projects.
- **Task 2:** To define (around 120?) the core variables, set out the recommendations for preparation of data, and explain how the metadata links to the objectives of interoperability and the FAIR principles.
- **Task 3:** Describe prioritisation process and methodology.
- **Task 4:** Provide simple validation tools covering the variable definition element of metadata.
- **Task 5:** Describe process for review and update of the list (e.g., every 2 years?).

The full list of contributions to this topic on the Miro board Is available in Annex 5. Additional suggestions to authors from participants included:

- Inclusion of a first task ('Task 0'), building on from the objective discussion, on what the core list will be used for, what is in scope, and what is out of scope.
- Consider how to continue engagement with the community throughout the process with suggestions of an initial dynamic collaborative document.

A number of points related to the approach for variable selection which is dealt with in the next section.

1.5 Variable selection

A proposed methodology devised by Martin Juckes was made available for comment during the survey and as a resource on the Miro board. Most survey respondents agreed in principle with the approach. Their comments focused on the prioritisation process, potential bias, and metrics. This was mirrored in the workshop discussion and contributions on the Miro board. The full list of contributions to this topic on the Miro board is available in Annex 5.

Additional suggestions from participants included:

- Consideration of a matrix rather than list approach
    - Develop a set of prioritised core variables needed for different communities or purposes. Modelling groups and MIPs could then choose which communities they wish to serve.
    - Map variables to MIPs rather than experiments. In the workshop, Martin Juckes responded to this suggestion indicating that there will need to be additional requests from specific MIPs, building on the core variable list.
- More clarity is needed about how this core variable list will fit into plans for the CMIP7 Data Request.
- Map high priority variables to observations - consider how the proposed high priority variables map onto the GCOS ECVs and observing systems and availability of high quality observations as part of the criteria.

- Suggestion that the 53 common variables (intersection of variables in top ranking models) could be a good starting point.
- Variable metrics as proxies for user relevance - discussion on volume of data download vs number of files as proxy. Importance of expert elicitation was raised regarding the following concerns discussed:
    - Common variables that may not have been available to download yet or important papers have not been written.
    - The use of file count for different temporal resolutions.
    - Accounting for emerging/future non-academic users who are interested in a much smaller set of the most common variables (perhaps around 10-15), but who are not yet reflected in downloads statistics.
- Use of existing lists such as:
    - The CMIP6 dataset lists provided to the IPCC DDC by WGI TSU, who collected the information from the IPCC WGI authors.
    - The variables used in the IPCC atlas - should these all be in the top priority list? Some disagreement between participants with some advocating use of the full WG1 list (TSU list collected retrospectively from the authors for CMIP and CORDEX) instead of just the Atlas variables.
- Listening to the modelling groups - what can they provide, what would they wish to provide? Could this help with reducing the long list of 120 to more manageable shortlist of 50?

Some suggestions related to expert elicitation, which is covered in the next section.

1.6 Expert elicitation process

The full list of contributions to this topic on the Miro board is available in Annex 5. Additional suggestions to authors from participants included:

- **Early engagement –** participants felt this should be done as early as possible with suggestions that it could be carried out at the same time as a community elicitation for the purpose of this exercise e.g., on the science outcomes that this list is to support.
- **Stakeholders to engage** - several recommendations including the MIP Chairs, VIACS Advisory Board, and IPCC WGI CLAs as well as IPCC WGII sector and regional leads. Consistent engagement with non-academic communities was flagged (e.g., climate services), with an offer made by Copernicus Climate Change Service (C3S) to assist with this.
- **Top-down vs bottom up** -both approaches were discussed:
    - **Top-down** e.g., ask the experts to rank the top 53 variables, then select the common top ones for the core list.
    - **Bottom-up** e.g., ask experts what variables they think are most important to include in all experiments early in the process and what criteria is important for them. Also suggested the need to check with MIPs for instances where low usage is due to inconsistency of variables across experiments.

- **Compromise –** could be to look at scientific and application requirements and establish those in common as highest priority.

- **Consult on format and usability of the variables** -establish the key needs. One discussed in the workshop was the need for large scale post processing of model output (i.e., downscaling efforts, translation to more standard formats, derived quantities etc.) and how to minimise the effort to produce and ensure required variables are included in the core.

1.7 Authors and reviewers

Martin Juckes reported on the Expressions of Interest received through the survey (see Annex 3). He highlighted that the initial author list lacks representation from the Southern Hemisphere and is weak on Asian representation. The full list of comments on this topic on the Miro board is available in Annex 5.

Additional suggestions from participants regarding representation focused on consideration of CMIP community function and domain as well as geography e.g., MIP leads representing specialist areas as well as wider communities, and domain representation covering oceans, clouds, atmosphere, land surface etc. Other issues of diversity such as gender and ethnicity may be challenging to address.

1.8 Next steps

Workshop participants were taken through the proposed process and timeframe (see Figure 2). Authors will be selected and supported with monthly meetings and a final decision meeting/workshop in October prior to submission for publication with the aim of having a published paper in advance of preparation for CMIP7 in 2023 (subject to confirmation by the authors).

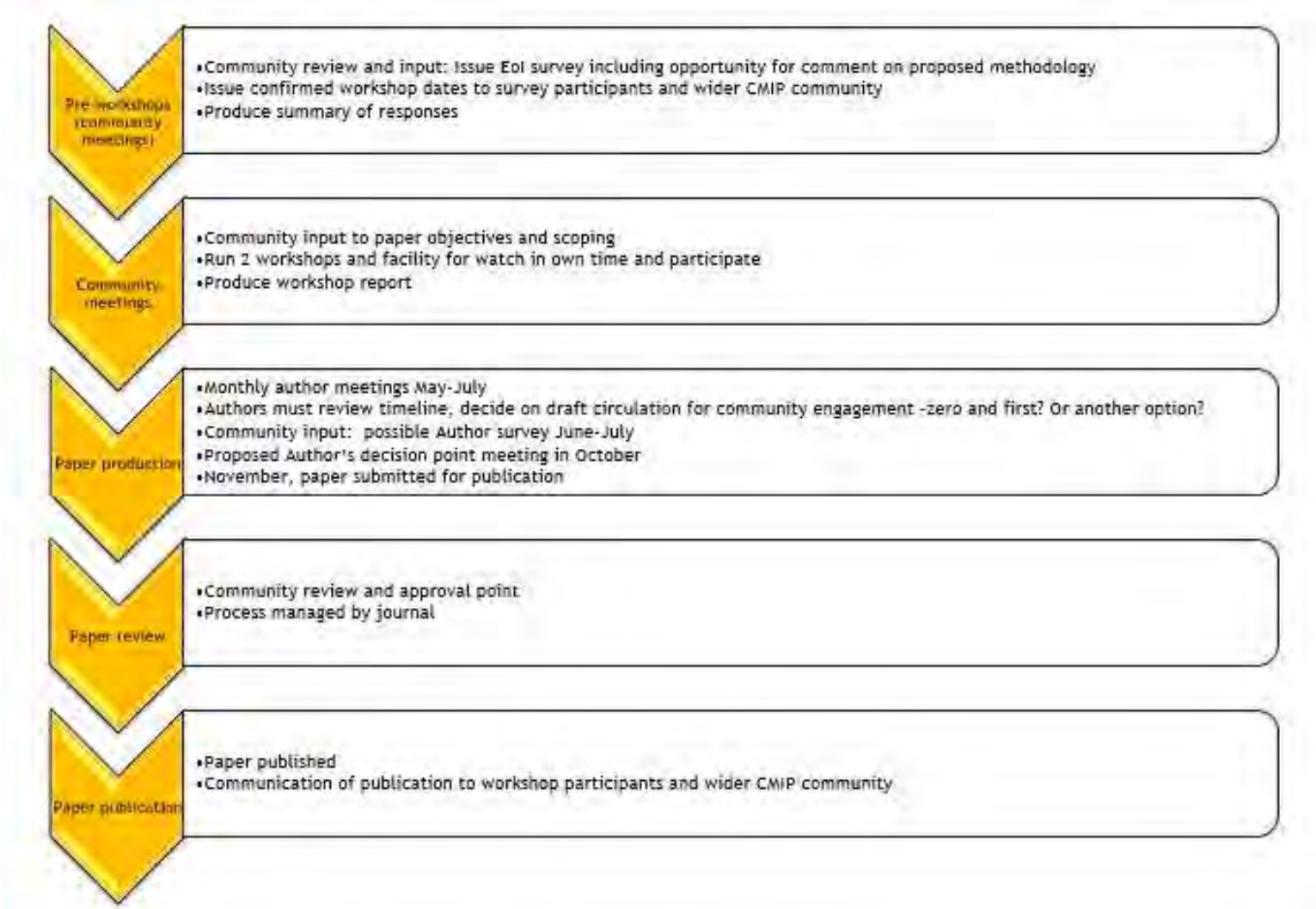


Figure 2: Proposed paper production process and timeframe.

IV. References

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R.J. Stouffer, and Taylor, K. E. 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Eyring, V., 2020: CMIP6 Status: highlights, lessons learned, challenges. Presentation at WGCM 23rd Session, 9th December 2020, https://www.wcrp-climate.org/images/modelling/WGCM/WGCM23/Presentations/2a_Eyring_CMIP6Overview_WGCMMeeting_201209.pdf.

Juckes, M., K. E. Taylor, P. J. Durack, B. Lawrence, M.S. Mizielinski, A. Pamment, J.-Y. Peterschmitt, M. Rixen, and S. Sénési. 2020: The CMIP6 Data Request (DREQ, version 01.00.31), *Geosci. Model Dev.*, 13, 201–224.

Juckes, M. 2020: IS-ENES3 Milestone M10.2 CMIP Data Request Schema 2.0

V. Annex 1: Survey participants

There were 32 survey responses. 3 survey participants chose not to be listed in this workshop report.

Name	Institution	Country
Brayshaw, David	University of Reading	United Kingdom
Brient, Florent	Sorbonne Université / Laboratoire de Météorologie Dynamique	France
Brookshaw, Anca	European Centre for Medium-Range Weather Forecasts	United Kingdom
Buontempo, Carlo	ECMWF / C3S	United Kingdom / Germany / Italy
Cagnazzo, Chiara	European Centre for Medium Range Weather Forecast	Germany
Cao	Nanjing University of Information Science and Technology	China
Horanyi, Andras	ECMWF	United Kingdom / Germany / Italy
Jones, Colin	National Centre for Atmospheric Science and University of Leeds	United Kingdom
Kawamiya, Michio	JAMSTEC	
Kim, Hyungjun	Korea Advanced Institute of Science and Technology	South Korea
Kim, YoungHo	Pukyong National University	South Korea
Koshiro, Tsuyoshi	Meteorological Research Institute	Japan
Lamarque, Jean-Francois	National Center for Atmospheric Research	United States
Lovato, Tomas	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, CMCC	Italy
Mizielinski, Matthew	Met Office	United Kingdom
Moine,Marie-Pierre	CERFACS (Centre Européen de Recherche et Formation Avancée en Calcul Scientifique)	France
Nikulin, Grigory	Swedish Meteorological and Hydrological Institute (SMHI)	Sweden
Orr, James	Laboratoire des Sciences du Climat et de l'Environnement / Institut Pierre Simon Laplace, CEA-CNRS	France
Pinnock, Simon	European Space Agency - Climate Office	United Kingdom
Roberts, Malcolm	UK Met Office	United Kingdom
Ruane, Alex C.	NASA Goddard Institute for Space Studies	United States
Schmidt, Gavin	NASA Goddard Institute for Space Studies (GISS)	United States
Schulz, Michael		Norway
Stockhause, Martina	German Climate Computing Center / IPCC Data Distribution Centre	
Strand, Gary	National Center for Atmospheric Research	United States
Taylor, Karl E.	Lawrence Livermore National Laboratory	USA
Visioni, Daniele	Cornell University	United States
Walton, Jeremy	Met Office	United Kingdom
Webb, Mark	Met Office Hadley Centre	United Kingdom

VI. Annex 2: Workshop Contributors

Surname	Name	Position	Institution
Brayshaw	David	Professor of Climate Science & Energy Meteorology	University of Reading
Brient	Florent	Lecturer	Sorbonne Université
Buontempo	Carlo	Director	ECMWF/C3S
Horányi	András	TBC	ECMWF
Jones	Colin	Head, UKESM project	Met Office
Juckes	Martin	Head of Atmospheric Science and Research and deputy head of CEDA	STFC
Kawamiya	Michio	Director, CEMA	JAMSTEC
Kim	Hyungjun	Associate Professor	Korea Advanced Institute of Science and Technology
Koshiro	Tsuyoshi	Researcher	Japan Meteorological Agency
Lovato	Tomas	Post-doctoral researcher	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, CMCC
Marchand	Roger	Professor	University of Washington
Mizielinski	Matthew	Climate Data Delivery Manager	Met Office
Moine	Marie-Pierre	Research Engineer	CERFACS
Nikulin	Grigory	leading scientist	Swedish Meteorological and Hydrological Institute (SMHI)
O'Rourke	Eleanor	Director	WCRP CMIP-IPO
Pamment	Alison	Environmental Data Scientist	UKRI/NCAS
Pascoe	Charlotte	Senior Data Scientist	CEDA (Centre for Environmental Data Analysis)
Pinnock	Simon	Earth Observation Applications Engineer	European Space Agency - Climate Office
Roberts	Malcolm	Global high resolution climate modelling	UK Met Office
Schmidt	Gavin	Director	NASA Goddard Institute for Space Studies (GISS)
Semmler	Tido	Senior scientist	Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research
Stephens	Ag	Head of Partnerships	STFC CEDA
Stockhause	Martina	IPCC DDC Manager	German Climate Computing Center (DKRZ)
Taylor	Karl	scientist	LLNL
Turner	Briony	Programme Manager	WCRP CMIP-IPO
Visioni	Daniele	Research Associate	Cornell University
Walton	Jeremy	Scientific Systems Manager	Met Office
Webb	Mark	Research Scientist	Met Office Hadley Centre
Zhang	Chengzhu Jill	Research Scientist	Lawrence Livermore National Lab

VII. Annex 3: Summary of survey responses

There were 32 responses representing 11 countries, see Table 1. Many of the participants had multiple positions of responsibility within the data request, modelling centres and MIP Chair elements of the CMIP community, see Figure 3. Several respondents had other roles, summarised in Table 2.

Table 1: Survey respondents' geographic representation

Country	Number of respondents
China	1
Italy	1
Norway	1
Sweden	1
Germany	2
Germany, Italy, UK	2
Japan	2
South Korea	2
France	4
USA	6
UK	10

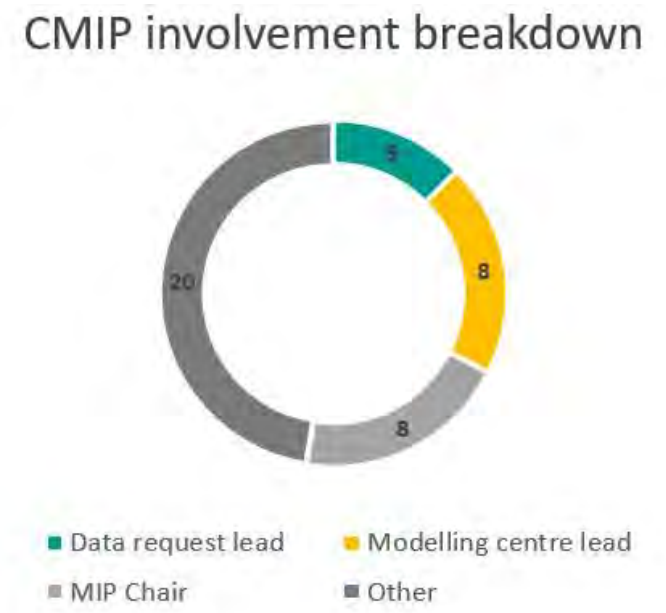


Figure 3: Survey participant CMIP involvement (many declared multiple roles).

Table 2: Survey respondents' other roles within the CMIP community.

Other (roles summarised):
Data Centre Management
CFMIP Representatives
CMIP data users/Climate service providers
CMIP Service Lead
CMIP5 Data Request Lead
Data Delivery System Manager
dr2xml -python tool based on the Data Request - developer
ESM Development and Application Lead
Former CMIP Panel members
IPCC Data Distribution Centre
MIP (diagnostic): VIACS Advisory Board & CORDEX
Modelling Centre Computation Scientist (CMIP6)
Modelling Centre Data Engineer (CMIP5 and CMIP6)
Observations (Obs4MIPs)
WGCM Infrastructure (WIP)

1.9 Authors and reviewers

15 survey participants expressed interest in being an author, of which 10 were happy instead to be considered as a reviewer. Seven participants indicated they would be willing to be a reviewer.

Name	EoI	Country
Jian Cao	Author or reviewer	China
James Orr	Author or reviewer	France
Chiara Cagnazzo	Author or reviewer	Germany
Martina Stockhause	Author or reviewer	Germany
Carlo Buontempo	Author or reviewer	Multiple
Tomas Lovato	Author or reviewer	Italy
Michael Schulz	Author	Norway
Hyungjun Kim	Author	S. Korea
Colin Jones	Author or reviewer	UK
David Brayshaw	Author or reviewer	UK
Jeremy Walton	Author	UK
Martin Juckes	Author	UK
Matt Mizielinski	Author or reviewer	UK
Daniele Visioni	Author	USA
Karl Taylor	Author or reviewer	USA

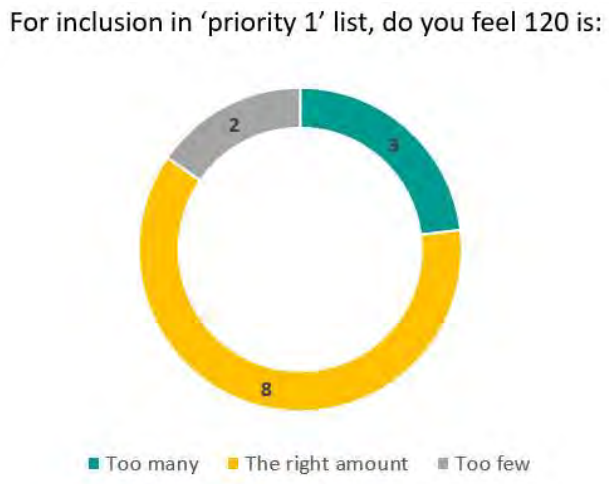
The 7 participants indicating happy to be a reviewer

Reviewer	
Michio Kawamiya	Japan
Malcolm Roberts	UK
Gavin Schmidt	US
Giorgio Graffino	UK
YoungHo Kim	S. Korea
Mark Webb	UK
Florent BRIENT	France

1.10 Methodological feedback and suggestions

Detailed feedback was provided by 15 of the 32 respondents on the proposed methodology and is included in full below.

Number of variables



The majority responding to this question felt 120 was the right amount. Two respondents elected not to comment.

Those who listed the right amount raised:

- Upgrading of variables for domain specific requirements e.g., Omon/thkcello is something needed for ocean models with variable vertical levels as Ofx/thkcello cannot be produced. Similarly models that use conservative potential temperature in the ocean may need Omon/bigthetao rather than Omon/thetao.
- The number should be determined by user need rather than pre-determined as a fraction of all available variables. Suggestion of following similar process as was done with VIACS for CMIP 6

to review the previous CMIP variables and interests in additional variables or temporal resolutions.
- Prioritise data/variables that have been extensively used in earlier CMIP cycles.
  Ensure data is saved for driving offline impact models (e.g., as in ISIMIP). Seeking clarity on why 120 as the cut off - is there a pronounced step down at 120th place?
- Commend use of DL metrics
- The number of 120 is nice but a clustering of the final criteria would help in selecting a group of variables based on their usefulness and not cutting at a subjective rank.

Those who listed too many referenced:
- Greater progress made when focused on a list of 80, 120 should be a maximum
- Most variables on model levels (atmosphere, ocean, soil) should be downgraded. These are not directly comparable across models.
- We should be favouring diagnostics on fixed pressure levels or depths. Especially variables for which there is no observational counterpart such as 'cloud fraction' - these should only be diagnosed using a robust simulator such as the ISCCP or COSP packages. Otherwise, we are not comparing like with like.
- Assuming priority 1 variables are provided by all CMIP participants, concern that this number too high for every participant to provide.

Those who listed too few referenced:
- Listed variables almost entirely physical
- Context not clear for number of P1 variables -is this for science needs or manageable/sustainable volumes?
- Confusion regarding objective of the work

Methodological approach

The majority who answered agreed in principle with the approach. Further comments focused on the prioritisation process, potential bias and metrics (see Figure 4 for a summary):

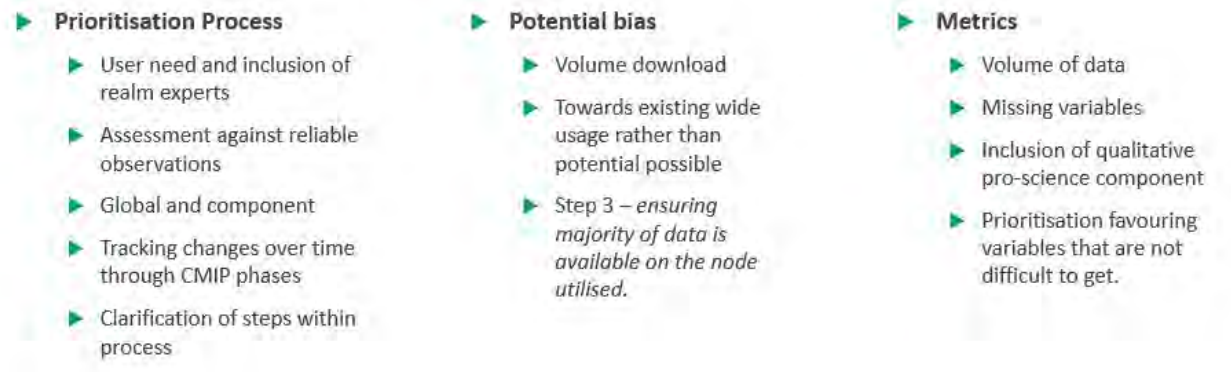


Figure 4: Summary of survey respondents' reflections on the proposed methodological approach.

Prioritisation

- User need
    - How priorities from particular groups get taken into account, e.g. how aerosol chemistry variables would be prioritised for experiments that that community are particularly interested in.
    - The number should be determined by user need rather than pre-determined as a fraction of all available variables.
    - Realm experts should only review the result of the objective ranking and point out to inconsistencies of the objective method to improve it. It is the only way to select variables in a fair way in between realms. There could be some reasons to promote a specific realm, in this case, we could think of a factor for this realm.
- Suggestion that any variable that can be assessed against reliable observations should be included at highest priority.
- Variables needed to account for the global and component (atmosphere, ocean, surface) budgets of energy, momentum, water, and certain trace constituents to be included at highest priority.
- Enabling tracking changes in models over time -consideration should be given to variables requested in phases of CMIP prior to CMIP5.
- Clarity needed on steps for prioritisation within methodology.

Bias (unintentional)

- Volume download -potential rigging as the surface (2D or 3D) variables are smaller in size than 4D variables + some variables that are downloaded at smaller rates may be very important for niche applications with strong stakeholder interests.
- Will tend to prioritise variables that have *already* been widely used, rather than addressing what *could* be done with additional/new output. Example provided: many impact modelling groups make a lot of use of reanalysis surface wind/solar data for renewables modelling - essentially at hourly resolution. CMIP is not widely used because it typically does not have the same capabilities (mainly in temporal resolution, but also missing variables etc), hence the demand for this data will not be visible in historic records of what have been downloaded so far.
- At step 3, check that the majority of data is available on this data node (or doing this check on two nodes to avoid lack of replication of some models).

Metrics

- A method to compare fairly data with different shapes and frequencies is needed.
- Volume
    - Volume downloaded can be used but cautiously. File count is better. However, download volumes and file number were referred to by another respondent as indicatives but polluted (by data replication operations) and not systematically reflecting the importance of the considered variable.
    - The distribution of the volume of data download (and data count for what it matters) it is highly skewed. Suggestion: consider as category 1 only those variables whose

download data volume is within 3 orders of magnitude of the data volume of the most downloaded variable.

- The volume criteria advantage a lot of high frequency data and 3D data whereas monthly 2D data are used largely, easy to produce and share. Suggestion: Before ranking, all volumes should be scaled to monthly 2D data, i.e., daily data volume should be divided by 30, 6hr data volume by 120, 3D monthly data by the number of vertical levels, 3D daily fields volume would be divided by 30*number of vertical levels.

- Looking for the variables that are missing, instead of only looking at the statistics for those up/downloaded
- The process needs a non-statistic, non-automatic, pro-science component.
- Questioning of whether the "same" variables in different frequencies, resolutions, domains be prioritised over difficult to get otherwise variables.

Additional quantitative criteria

Respondents also had additional suggestions regarding considerations for quantitative criteria for variable selection and prioritisation. These focused on criteria specific to various user communities, specific points on the variables and a few other suggestions (see Figure 5 for a summary).

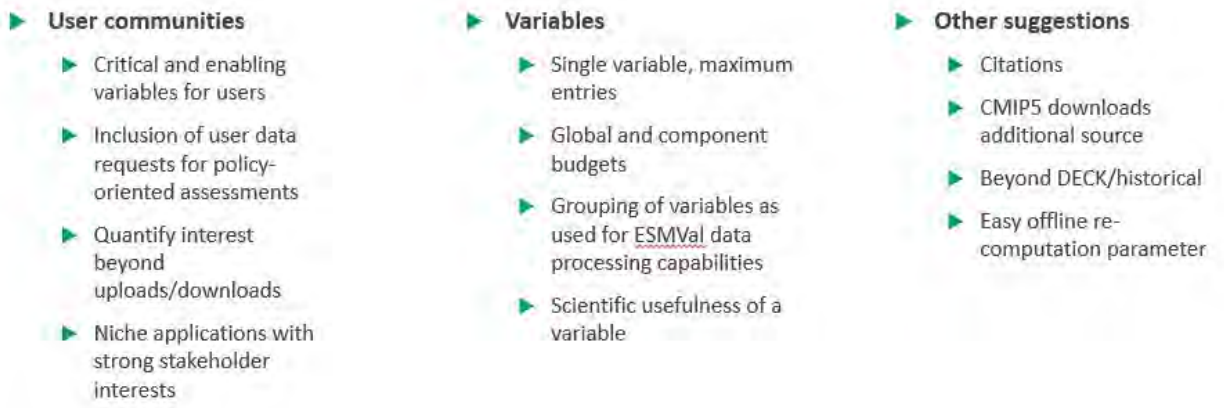


Figure 5: Summary of survey respondents' suggestions relating to quantitative criteria within the methodological approach.

User communities

- User communities need inclusion - how outputs are used, what are critical and enabling variables + suggestion for inclusion in prioritisation process of specific data requests to support key international policy-oriented assessments (e.g., IPCC, UNEP etc)
- Quantify interest in user communities (e.g., VIACS) beyond uploads/downloads. How are outputs used? What are critical and enabling variables? Some variables that are downloaded at smaller rates may be very important for niche applications with strong stakeholder interests.

Variables

- Maximum number of entries for a single variable
- Variables needed to account for the global and component budgets of energy, momentum, water, and certain trace constituents be included at highest priority.

- Variable groups used as prioritisation method when developing data processing capabilities for ESMVal.
- Instead of volume and files, a more objective criteria would be the scientific usefulness of a variable (e.g., measured by the number of scientific publications it has been used for).

Other suggestions

- Number of published papers and their quality (number of citations).
- Additional source for CMIP5 downloads to be considered: IPCC DDC AR5 for CMIP5
- The existence of observational counterparts, and the ability to compare across models.
- Go beyond DECK/historical -look at the CMIP6 priority attribution according to the MIP/experiment.
- Introduce a parameter indicating the degree of easy offline re-computation (if feasible).

Science/impact-based prioritisation

Several respondents had suggested consideration of the list that IPCC AR6 WG1 had circulated during CMIP6. This was added into the proposed methodology output list (column K) to provide a comparison.

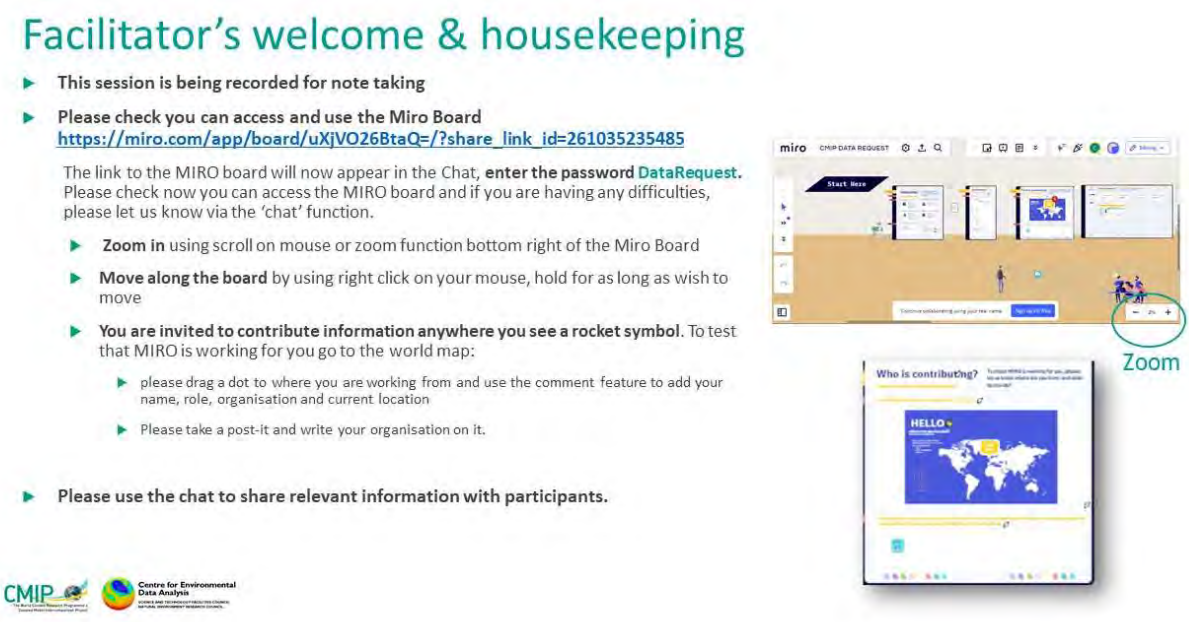
Concerns were raised about the need for expert elicitation to identify variables critical to climate science that do not necessarily come through prioritisation as proposed. Respondents raised the following points:

- Suggestion of comparing this list to the IPCC WG1 list that was circulated during CMIP6 to see if there are many differences.
- A limited number of high frequency variables are required by CORDEX, etc.; we were able to accommodate these after we had processed the monthlies, etc. This prioritisation seemed to work well.
- Variable usage in publication provided by PCMDI's pubsite for CMIP5: https://cmip-publications.llnl.gov/
- Variable subset selection from ETH Zurich for CMIP5 datasets provided for IPCC AR5 WGI
- There may be some obscure field that's critical to climate science, but it's not yet recognised as such. On the other hand, lots of good science is being extracted out of a small part of the total data.
- For variables that are subject to conservation constraints, we should be asking for the intrinsic variable (i.e., ocean heat content, rather than temperature), ozone column amount vs. level concentrations, etc.
- It seems to have already been considered via C3S statistics.
- The model limitation itself, because not computing some diagnostics (or even diagnostics not able to be computed offline due to basic variables not being output).

VIII. Annex 4: Workshop slide deck

Slide 1

Slide 2



Slide 3

Agenda

- 1. Chair's welcome and workshop introduction
- 2. Summary of community response to proposed paper methodology & summary of relevant CMIP Community Survey responses
- 3. Plenary Session 1 - to gather thoughts and identify community agreement on the objectives of the paper.
- 4. Chair's reflections
- 5. Plenary Session 2 - Meeting attendees' direction for authors
- 6. Review of meeting outcomes and closing remarks by Chair

CMIP Centre for Environmental Data Analysis

Slide 4

Workshop Objectives

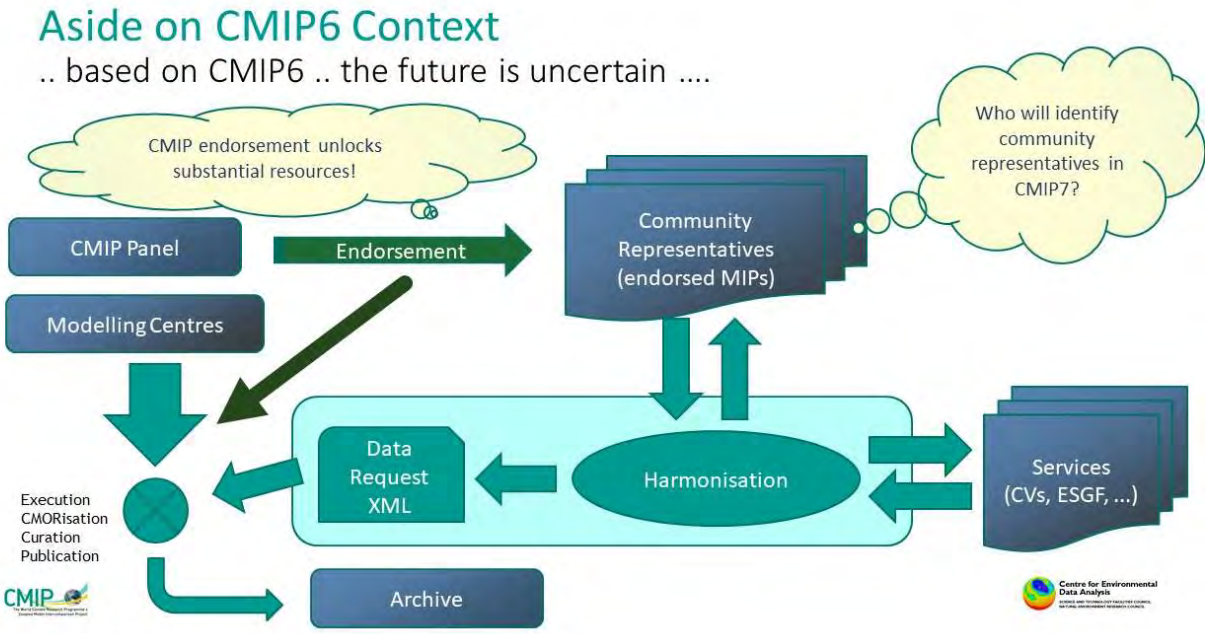
Initiate the drafting of a paper on WCRP core variables:

1. Identify an initial list of authors and reviewers;
2. Agree on a process of expert elicitation to identify a list of top priority variables which reflects community demands for a substantial reduction relative to CMIP6;
3. Agree on a time frame and process for drafting the paper;
4. Review objectives of the paper.

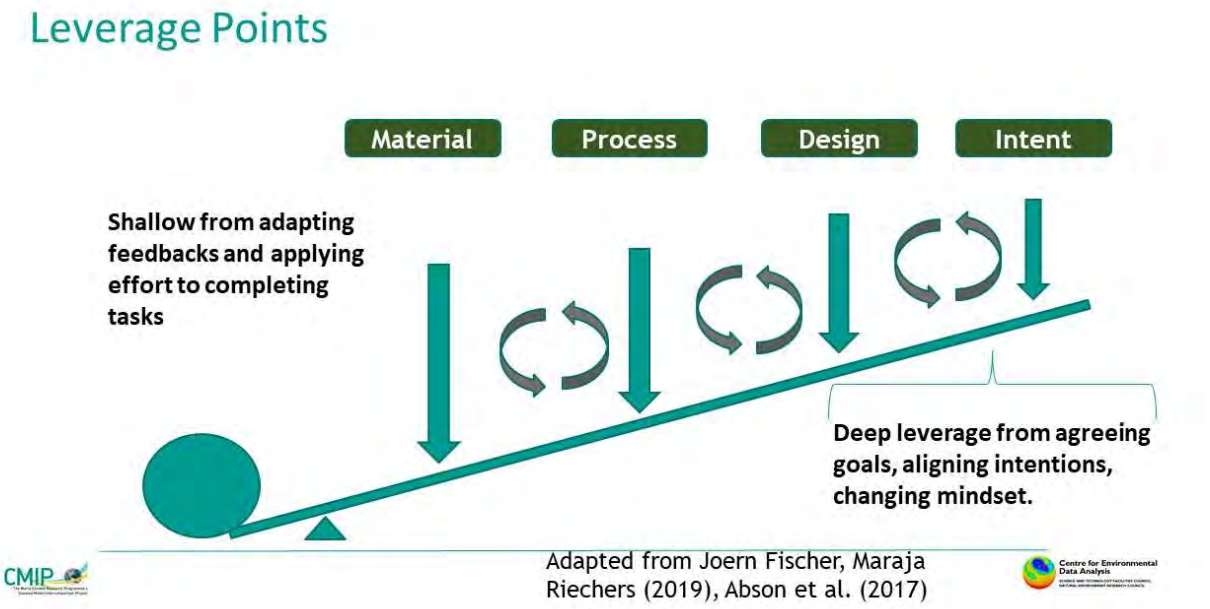
CMIP Centre for Environmental Data Analysis

1.11

Slide 5



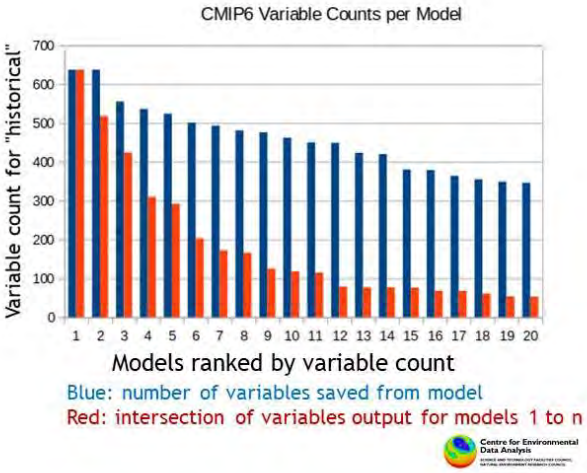
Slide 6



Slide 7

Variable output by Model for CMIP6 Historical

- There are 673 variables archived from the `IPSL-CM6A-LR` model `historical` simulation.
- 20th ranked `AWI-ESM-1-1-LR` has 346 variables archived.
- The intersection of the variables output by the first 20 models is only 53.
- The level of consistency is reduced if multiple experiments are considered.



CMIP

Centre for Environmental Data Analysis

Slide 8

Paper objectives and boundaries

Identify a set of core variables and associated metadata to enable consistent and efficient comparison of simulations across multiple intercomparison projects, reducing the workload for data providers and users by providing a reusable basic set of variables.

Aiming to enable consistency with respect to the variables published in the data archive.

Aim for 90% of models providing 90% of variables. cf. 28% of models providing 28% of CMIP6 priority 1 variables; 90% of models provide 7.8% or more; just two models provide 50% or more.

The paper will not:

- Deal with procedures needed in CMIP7 to incorporate requirements from multiple MIPs into a consolidated request;
- Deal with technical details of metadata implementation which need to be agreed with other elements of the CMIP technical infrastructure (CVs, ESGF, ES-DOC, CMOR, Citation, PID, ..).

CMIP Centre for Environmental Data Analysis

Slide 9

Paper tasks

1. **Describe the role of the core variables** as baseline set of variables for curated data from WCRP endorsed climate modelling projects.
2. To **define** (around 120?) **the core variables**, set out the recommendations for preparation of data, and explain how the metadata links to the objectives of interoperability and the FAIR principles.
3. **Describe prioritisation process and methodology**.
4. **Provide simple validation tools** covering the variable definition element of metadata.
5. **Describe process** for review and update of the list (e.g. every 2 years?).

CMIP Centre for Environmental Data Analysis

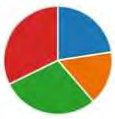
Slide 10

Authors & Reviewers

9. **Would you like to express interest in being a paper author/reviewer?**

More Details

- Yes -reviewer 7
- Yes - author 5
- Yes - happy to be considered fo... 9
- No 10



Reviewer	
Michio Kawamiya	Japan
Malcolm Roberts	UK
Gavin Schmidt	US
Giorgio Graffino	UK
YoungHo Kim	S. Korea
Mark Webb	UK
Florent BRIENT	France

CMIP Centre for Environmental Data Analysis

Name	EoI	Country
Martin Juckes	Author	UK
Matt Mizielinski	Author or reviewer	UK
Daniele Visioni	Author	USA
Karl Taylor	Author or reviewer	USA
Jeremy Walton	Author	UK
Martina Stockhause	Author or reviewer	Germany
Hyungjun Kim	Author	S. Korea
Colin Jones	Author or reviewer	UK
Michael Schulz	Author	Norway
Tomas Lovato	Author or reviewer	Italy
James Orr	Author or reviewer	France
Jian Cao	Author or reviewer	China
David Brayshaw	Author or reviewer	UK
Carlo Buontempo	Author or reviewer	Int

Initial author list lacks representation from S. Hemisphere and is weak on Asian representation.

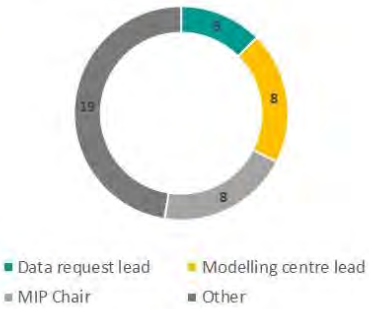
Slide 11 (Please note that one respondent completed the survey after the snapshot was taken for the workshop. This respondent did not elect to comment on the methodological approach.)

Pre-workshop survey results

31 respondents

Country	Number of respondents
China	1
Germany	1
Italy	1
Norway	1
Sweden	1
Germany, Italy, UK	2
Japan	2
South Korea	2
France	4
USA	6
UK	10

CMIP involvement breakdown



Other (roles summarised):
Data Centre Management
CFMIP Representatives
CMIP data users/Climate service providers
CMIP Service Lead
CMIP5 Data Request Lead
Data Delivery System Manager
dr2xml - python tool based on the Data Request - developer
ESM Development and Application Lead
Former CMIP Panel members
IPCC Data Distribution Centre
MIP (diagnositic): VIACS Advisory Board & CORDEX
Modelling Centre Computation Scientist (CMIP6)
Modelling Centre Data Engineer (CMIP5 and CMIP6)
Observations (Obs4MIPs)
WGCM Infrastructure (WIP)

CMIP Centre for Environmental Data Analysis

Slide 12

Methodology feedback –number of variables

For inclusion in 'priority 1' list, do you feel 120 is:

- The Right Amount
  - Upgrading of variables for domain specific requirements
  - The number should be determined by user need rather than pre-determined as a fraction of all available variables.
  - Prioritise data/variables that have been extensively used in earlier CMIP cycles.
  - Seeking clarity on why 120 as the cut off - is there a pronounced step down at 120th place?
  - Commend use of DL metrics
  - Clustering of the final criteria would help in selecting a group of variables based on their usefulness

CMIP Centre for Environmental Data Analysis

Slide 13

Methodology feedback - approach

- **Prioritisation Process**
  - User need and inclusion of realm experts
  - Assessment against reliable observations
  - Global and component
  - Tracking changes over time through CMIP phases
  - Clarification of steps within process
- **Potential bias**
  - Volume download
  - Towards existing wide usage rather than potential possible
  - Step 3 – *ensuring majority of data is available on the node utilised.*
- **Metrics**
  - Volume of data
  - Missing variables
  - Inclusion of qualitative pro-science component
  - Prioritisation favouring variables that are not difficult to get.

CMIP Centre for Environmental Data Analysis

Slide14

Methodology feedback – additional quantitative criteria (not already raised previously)

- **User communities**
  - Critical and enabling variables for users
  - Inclusion of user data requests for policy-oriented assessments
  - Quantify interest beyond uploads/downloads
  - Niche applications with strong stakeholder interests
- **Variables**
  - Single variable, maximum entries
  - Global and component budgets
  - Grouping of variables as used for ESMVal data processing capabilities
  - Scientific usefulness of a variable
- **Other suggestions**
  - Citations
  - CMIP5 downloads additional source
  - Beyond DECK/historical
  - Easy offline re-computation parameter

CMIP Centre for Environmental Data Analysis

Slide 15

Methodology feedback – science/impact based prioritization (not already raised previously)

- Suggestion of comparing this list to the IPCC AR6 WG1 list circulated during CMIP6 to see if there are many differences *(Martin has added this as column K in the variable prioritisation sheet)*
- The model limitation itself, because not computing some diagnostics (or even diagnostics not recomputable offline because basic variables not output)
- There may be some obscure field that's critical to climate science but not yet recognized.
- For variables that are subject to conservation constraints, we should be asking for the intrinsic variable (e.g.,ocean heat content, rather than temperature)
- A limited number of high frequency variables are required by CORDEX, etc.; we were able to accommodate these after we'd processed the monthlies, etc. This prioritisation seemed to work well.
- Variable usage in publications provided by PCMDI's pubsite for CMIP5: https://cmip-publications.llnl.gov/

Slide 16

Summary of relevant CMIP Community Survey responses

- The CMIP community survey was carried out in early 2022 and over 300 responses were received.
- 78% were satisfied or very satisfied with core variables included in the DECK reducing to 64% satisfaction with core variables included across the MIPs they interacted with.
- Many respondents' comments suggested there were too many core or priority 1 variables.

However.....

- There were also many comments suggesting a need for:
  - Increased temporal resolution.
  - More ocean variables.
  - Variables relevant to extremes.
  - Variables required for CORDEX/regional downscaling.
  - A variety of other variables.

Slide 17

Agenda

- 1. Chair's welcome and workshop introduction
- 2. Summary of community response to proposed paper methodology & summary of relevant CMIP Community Survey responses
- 3. Plenary Session 1 - to gather thoughts and identify community agreement on the objectives of the paper.
- 4. Chair's reflections
- 5. Plenary Session 2 - Meeting attendees' direction for authors
- 6. Review of meeting outcomes and closing remarks by Chair

CMIP Centre for Environmental Data Analysis

Slide 18

Plenary discussion – objectives of the paper

Please be clear if your discussion point is focused on that variable type. In the second breakout session you can choose to stay or move to the other room.

Use the Miro Board to capture your points –remember people contributing offline will only be able to see the contents of the Miro Board.

- AIM: to gather thoughts and identify community agreement on the objectives of the paper.
- **Key discussion points**
  - **Paper objectives**
  - **Approach for initial variable selection**
  - **Process of expert elicitation that should be incorporated into devising the core list**
  - **If time:** ***How do we do expert elicitation to get reasonable representation across communities for adding and removing variables***

(note we need to avoid building a wishlist as this happens later in the CMIP process)

CMIP Centre for Environmental Data Analysis

Slide 19

Agenda

- 1. Chair's welcome and workshop introduction
- 2. Summary of community response to proposed paper methodology & summary of relevant CMIP Community Survey responses
- 3. Plenary Session 1 - to gather thoughts and identify community agreement on the objectives of the paper.
- 4. Chair's reflections
- 5. Plenary Session 2 - Meeting attendees' direction for authors
- 6. Review of meeting outcomes and closing remarks by Chair

CMIP Centre for Environmental Data Analysis

Slide 20

Agenda

- 1. Chair's welcome and workshop introduction
- 2. Summary of community response to proposed paper methodology & summary of relevant CMIP Community Survey responses
- 3. Plenary Session 1 - to gather thoughts and identify community agreement on the objectives of the paper.
- 4. Chair's reflections
- 5. Plenary Session 2 - Meeting attendees' direction for authors
- 6. Review of meeting outcomes and closing remarks by Chair

CMIP Centre for Environmental Data Analysis

Slide 21

Breakout session 3

You will shortly be sent to one of two breakout rooms (randomly assigned)
Use the Miro Board to capture your points -remember people contributing offline will only be able to see the contents of the Miro Board.
AIM: to gather thoughts and identify community agreement on the proposed paper tasks and author representation

Discussion point: Paper tasks

- **Do you agree with the proposed tasks? If not, what is missing/needs to be changed?**
  - **Task** 1 Describe the role of the core variables as baseline set of variables for curated data from WCRP endorsed climate modelling projects.
  - **Task 2** To define (around 120?) the core variables, set out the recommendations for preparation of data, and explain how the metadata links to the objectives of interoperability and the FAIR principles.
  - **Task 3** Describe prioritisation process and methodology.
  - **Task 4** Provide simple validation tools covering the variable definition element of metadata.
  - **Task 5** Describe process for review and update of the list (e.g. every 2 years?).

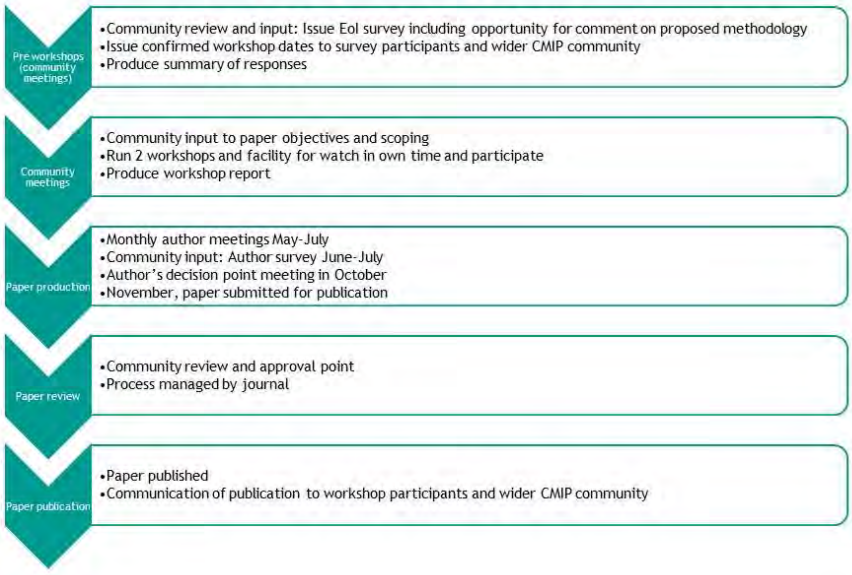
Discussion point: Representation

- **Do we need broader representation in the paper authors or if authors survey the community is this considered appropriate?**
- **Are there community members that should be invited to act as reviewers, not currently listed?**
- **Should we schedule a meeting for authors with reviewers prior to submission?**

CMIP Centre for Environmental Data Analysis

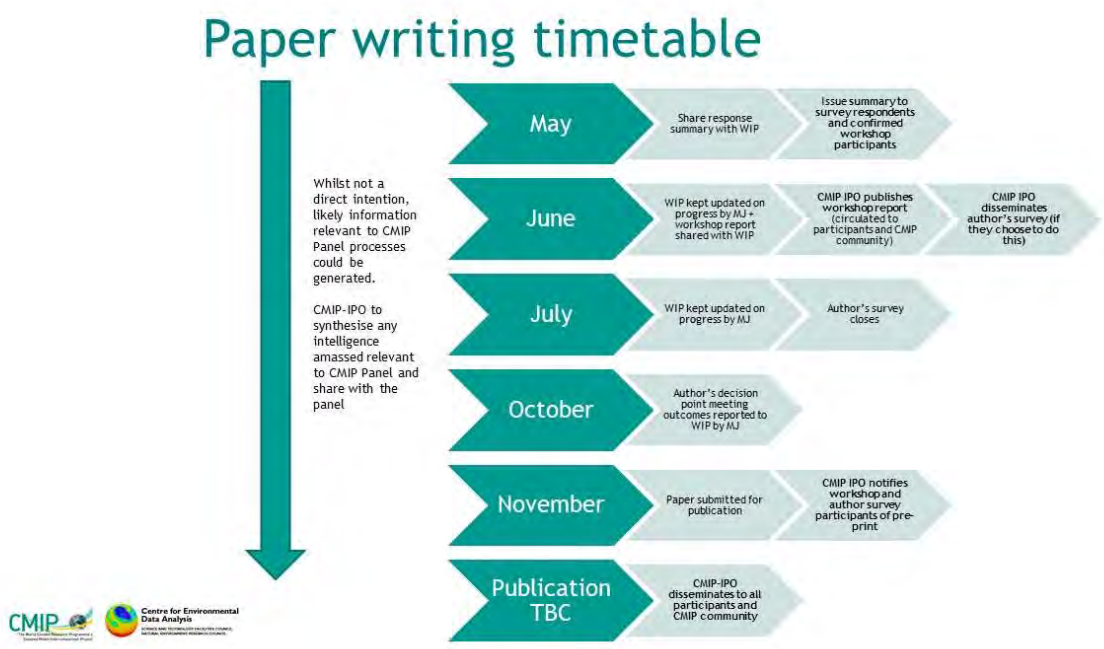
Slide 22

Overview of proposed process



CMIP Centre for Environmental Data Analysis

Slide 23



Slide 24

Breakout session 3

You will shortly be sent to one of two breakout rooms (randomly assigned)
Use the Miro Board to capture your points -remember people contributing offline will only be able to see the contents of the Miro Board.
AIM: to gather thoughts and identify community agreement on the proposed paper tasks and author representation

Discussion point: Paper tasks

- **Do you agree with the proposed tasks? If not, what is missing/needs to be changed?**
  - **Task** 1 Describe the role of the core variables as baseline set of variables for curated data from WCRP endorsed climate modelling projects.
  - **Task 2** To define (around 120?) the core variables, set out the recommendations for preparation of data, and explain how the metadata links to the objectives of interoperability and the FAIR principles.
  - **Task 3** Describe prioritisation process and methodology.
  - **Task 4** Provide simple validation tools covering the variable definition element of metadata.
  - **Task 5** Describe process for review and update of the list (e.g. every 2 years?).

Discussion point: Representation

- **Do we need broader representation in the paper authors or if authors survey the community is this considered appropriate?**
- **Are there community members that should be invited to act as reviewers, not currently listed?**
- **Should we schedule a meeting for authors with reviewers prior to submission?**

CMIP Centre for Environmental Data Analysis

Slide 25

Agenda

- 1. Chair's welcome and workshop introduction
- 2. Summary of community response to proposed paper methodology & summary of relevant CMIP Community Survey responses
- 3. Plenary Session 1 - to gather thoughts and identify community agreement on the objectives of the paper.
- 4. Chair's reflections
- 5. Plenary Session 2 - Meeting attendees' direction for authors
- 6. Review of meeting outcomes and closing remarks by Chair

CMIP Centre for Environmental Data Analysis

Slide 26

Please give us feedback

CMIP

Slide 27


CMIP
The World Climate Research Programme's Coupled Model Intercomparison Project
Centre for Environmental Data Analysis
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL
Thank you for your contributions today.
CMIP IPO contact details
Director: eleanor.orourke@ext.esa.int
Programme Manager: briony.turner@ext.esa.int
Science & Communications Officer: annia.domenech@ext.esa.int

IX. Annex 5: Workshop Miro board content

This content is an export from the Miro board used in both workshops and was made available for all participants, including those electing to participate in their own time. The board was closed on the 31 May 2022. Where a note was added in response to a point, this has been indicated in italics.

Objective response and community engagement considerations for authors

- Who are the users of this data? CMIP focused on model development and research
- Define clearly what the short list is for.
- A core set to ensure that basic variables are there for every run, plus templates for other experiments.
- Could core variables be restricted to core experiments?
- How to decide who are user community is? What do we want the future user community? High impact surface variables for impacts communities
- Is the idea of this list to provide a core set of variables that should be archived in all CMIP7 experiments? If so, what analysis do we want it to be possible to do on all experiments? What scientific areas do these cover? A specific list of applications would be useful, and this should be consulted on and agreed early on. These could be prioritised in terms of impact on IPCC.
- Core set of variables a MIP would be encouraged to look at as a starting point. It is up to them what data they produce. If part of an organised effort, e.g. CMIP7 there might be CMIP7 guidance
- What happened to the idea of a core set of variables in CMIP6? How did that fail? *Response: There was discussion about how to do it but it came too late.*
- Discussed solution to include: Get endorsement from WGCM -might need to enlarge/reduce list to get endorsement. Need also the community endorsement we get through a publication process.
- For 3D atmospheric variables, information about pressure levels on which the variables would be output is also important.
- Revising the output regarding ocean BGC could be useful.
- I agree on the objectives too.
- Defining community: needs to be useful to data users (who often want the consistency - particularly important in derived products)
- The IPCC list was a big hit with us as regards prioritisation, filtering etc.
- The IPCC list went into scenario MIP, ad hoc process to get out (sic) a more manageable (sic) list
- Could the list of variables be "enforced" the way participation in the DECK was for a model?
- Shift from mitigation to more impacts focused?
- The data request was very different from MIP to MIP - reducing DR to core variables would be useful and then add MIP specific (perhaps would reduce the amount of data to be reduced)
- Clarity: enabling data sharing, not about requesting data. By agreeing a core set, MIPs have a starting point and enable MIPs to be consistent if they would like to be.
- Need a way to allow each experiment to set priority. As is, centers seem to have not paid much attention to what different experiments needed. If this is done, then the "Core set" can be very small as long as MIPS/experiments can expand on this. ... my point here is this didn't work in CMIP6. If there isn't some way to tailor then the "core" set will be large.
- Looking forward to future data demands e.g. hourly resolution for energy modelling
- CMIP is now used by a community which goes way beyond academia. Accounting for the non-academic use of the data will be particularly important.

- Address variables to support impact/economic research connected to overall societal objectives, e.g. Sustainable Development Goals.
- Scenarios and historical mostly used for impacts community (one of set of high priority) while others focused on model evaluation.
- Yes -needs to be part of the elicitation, circulated to the modelling centres + Near surface humidity field in CMIP which is regularly produced and difficult to use -some of the variables persistent in data requests of different CMIP cycles have issues we need to revisit and think about carefully -strange features in sea surface variables which occur when have landlocked seas
- Centers are very limited in how much they can customise output for different experiments. Easy to add variables as a function of included modules (i.e. atmos. chemistry (sic), ocean, dynamic icesheets, carbon cycle etc.)
- Need to ask modelling centres "can you produce ~90% of this list regularly (where appropriate)"...and if not, is there some support WCRP can give you to produce them?
- There is a complicated workflow being conducted in the centres -having these variables specified in advance could provide a small step towards rationalising (sic) the process BUT a concern about how we deal with the different MIPs and how this maps on to what the modelling centres have to do.
- A variable set that can support the production of colloquial variables such as rainfall (mm/day).
- The current process is very focused at a granular level on particular variables. I think that getting more clarity on what this list is for at the start is important. Community engagement / buy in on this activity would be beneficial.

Out of scope: What the paper should not deal with

- Post processing tools to reduce the data output and tailor.
- future of CMIP
- Science results
- Common grids and (re)gridding
- Difficult to find a generalised tool for all model but would be very useful.
- Venn type structure with intersections with more specific requirements - what are the 20-30 variables to be used by all. [report authors suspect this was a consideration for paper authors, placed in wrong area]
- Rather than having a single central list of variables to all MIPS (which have very different (sic) science objectives) try develop a set of list of variables needed for different communities/purposes (e.g., what vars does community X need, what vars do community Y need). Then produce prioritised sets of variables that 'serve' each community. Modelling groups/MIPs can then choose which communities they wish to serve. [report
- authors suspect this was a consideration for paper authors, placed in wrong area].

PAPER TASKS: Do you agree with the proposed tasks? If not, what is missing/needs to be changed?

- Task 0? Clarify what the core list will be used for, what is in scope, what is out of scope. The idea is that this is to support analysis that makes sense across all experiments -in the process it might be valuable to have more upfront work on deciding what this list is for and community engagement on that before diving into the granular process of prioritising individual variables -suggestion for discussion.

- Why not work on a shared document where we can put everyone's opinions on the table and allow us to react and interact. By starting to feed this document with the suggestions collected by the questionnaire (before sharing with the WIP).
- This should be discussed at the core authors' first meeting.
- Authors will need to engage continually to keep paper on track with community expectations.

TASK 1 Describe the role of the core variables as baseline set of variables for curated data from WCRP endorsed climate modelling projects.

- Wondering about the role of core variables - is there really a core set. Have packages of diagnostics useful to each community that modelling centres choose to serve?
- Have a list that everyone uses rather than selective - dependent on CMIP7 definition or wider WGCM endorsed MIPs [contained red sticky dot indicating someone also supported this point].
- Need a core set of experiments for a core set of variables that all modelling centres produce. MIPs then do 'legwork' to get modelling centres to output.
- Small set of core variables to be produced by all. [contained red sticky dot indicating someone also supported this point].

TASK 2 Define (120?) the core variables, set out recommendations for preparation of data & explain how metadata links to the objectives of interoperability & FAIR principles.

- Link between metadata & FAIR is via I and R
- How will the 120 variables be distributed between realms? I.e. would 90% of them be atmospheric? Maybe better to go bottom-up, define core variables per realm and then sum up rather than have an overall set number [contained red sticky dot indicating someone also supported this point].
- Each variable should have a clearly definition rather than just obscure name [contained red sticky dot indicating someone also supported this point].
- Can include a prototype json table? [contained red sticky dot indicating someone also supported this point].
- Colin: Is 6hourly, multi-pressure level (say 30 levels) temperature potentially a single variable in a core list and thus equal to (say) monthly mean precipitation?

TASK 3 Describe prioritisation process and methodology.

- Have a level of categorisation within the list for high volume data.
- Define criteria for prioritising variables - volume, downloads, popularity, feasibility (sic)
- How many modelling groups need to be involved to make it worthwhile? Answer: MJ No cut-off but can't set a target | MM this will provide clarity to the modelling centres as to what the core variables are for comparing models. Need to have a list though that we can reasonably expect other groups to do.
- How many centres are needed to respond -- to make this work/be seen as authoritative?

TASK 4 Provide simple validation tools covering the variable definition element of metadata.

- Not sure what the validation tool would do - check for CMOR compliance?

TASK 5 Describe process for review and update of the list (e.g. every 2 years?).

- CORDEX data request 1. what they request from CMIP6 and then 2. the data requirements from the CORDEX simulations. What they want from the global models is specific to a small number of experiments and should not be part of the core list. Cordex-CMIP6 data request: **https://cordex.org/experiment-guidelines/cordex-cmip6/data-request/** *In response MJ clarified context - Global not regional community focus.*

Approach for initial variable selection

- How to prioritise both variable and temporal resolution - is the 120 monthly mean or hourly? To consider data volume [contained red sticky dot indicating someone also supported this point]
- Mapping high priorities to observations -currently in GMD papers rather than technical metadata. It is something that could be done in this paper, to think about how they map onto GCOS and the observing systems.
- We are not talking about losing variables. Some discussion on what the MIPs would like to see given don't know how CFMIP will fit in CMIP7. MJ -at start of CMIP6 many MIPs wanted a core variable list which is part of the motivation of this exercise, also due to feedback that people want to see a core set of variables.
- Surface variables are vital for impacts - highest priority and then the model development priorities.
- MIPs have very specific requirements so an initial list for the DECK/historical then for MIPs have their own essential variables plus control run.
- Will this affect what goes in the DECK but people in MIPs will be at liberty to choose what they'd like? (assumption) but in CFMIP, relied on diagnostics being in the experiments. Important that these diagnostics are in the DECK. Worried if prioritised from view of everyone, MIPs reliant on diagnostics in DECK will get lost as minorities. Answer MJ: The MIPs are free to request what they want. An additional list was parachuted into everyone's request during CMIP6 process to ensure a baseline set of variables requested from all experiments. Now trying to do it the other way around, define the baseline set in advance to make it easier for MIPs to think about what they need in addition to a core set. Aim is to provide a baseline for high level experiments and for impact studies.
- First priority should be those that can be used to evaluate how well the model represents the earth system to allow for it to be useful.
- Would suggest asking MIP chairs and IPCC CLAs what variables they think are most important to include in all experiments early in the process. Expert engagement will be more effective if it happens sooner rather than later.
- What variables haven't been downloaded/important papers haven't been written because not enough models had that common variable(s) available to download?
- Look at volume of data download / number of files (as proxy for user relevance)
- Are available observations used as a prioritisation criterion - e.g. the GCOS ECVs?
- Use CMIP6 dataset lists provided to the IPCC DDC by WGI TSU, who collected the information from the IPCC WGI authors. [contained red sticky dot indicating someone also supported this point]
- Verify whether variables used in the IPCC atlas are all in the top priority list https://interactive-atlas.ipcc.ch/
- The 53 common variables could be a good starting point for the "priority 0" [contained red sticky dot indicating someone also supported this point].
- What is the list of 53 common variables across the top models? *Response 53 comes from intersection of variables in top ranking models.*
- Discussion on taking out the model level data (in due to high volume) -can we deal with that through expert elicitation and get some transparency Discussion suggestion modelling groups need to be able to say -this is what we can give you e.g. 120 candidate variables and modelling groups voting on the ones they like/can deliver. 1200 variables is too long for people to look through -MJ made shortlist of 120, a shortlist could be cut down to 50
- There may be more files in daily and sub-daily data than monthly data even for the same variable. Is this considered for counting the number of file?
- Hard to evaluate if dataset is not available.

- The non-academic/modelling users are disproportionally interested in a small (<<120) set of variables.
- Good strategy to communicate to the wider community and how to participate - may not be able to access the data. *Response: May not necessarily be the case - very few people download winds from idealised but are important from historical and scenario.*
- The adaptation community was included in CMIP6 via VIACs advisory board. Did add variables to the DATA REQUEST -the advantage is that this is transparent and modelling centres know that collection of variables is only needed if their modelling is relevant to impact studies BUT feedback on having different set of variable requirements for each experiments was difficult. Could look at mapping to MIPs rather than to experiments.
- Look at downloads - set much more downloaded much more?

Process of Expert Elicitation for devising the core list

- Ask experts what timescales we want in the core list. Just monthly? Some daily / instantaneous? Links to question of what the list is for.
- Compile an initial list as workshop output and circle it to experts incl. users from VIACS in a review process
- Present an initial selection of variable to community workshop (like today's) and collect feedbacks.
- Close engagement with VIACS community and their network would be a good place to start.
- The Atlas variables are part of the dataset lists provided to the IPCC DDC at DKRZ. Suggestion is to use the whole list not just the atlas part. Offer is to go into the list and bring the variables out. TSU collect the list after the publication of the report. Discussion concluded could be useful contribution to the process.
- I think it's important to consult community experts on what should be included early on. This could be done at the same time as a community elicitation on the purpose of this exercise - e.g. on science outcomes that this list is to support.
- Should use the full WG1 list (TSU list collected retrospectively from the authors for CMIP and CORDEX) instead of just the Atlas variables
- Can we ask the expert to rank the list of the top 53 and then picking the common tops ones? *Response 53 comes from intersection of variables in top ranking models.*
- Thinned out list for onward use
- Variable list needs to be aligned with historical and Scenarios. For DECK what is used most for other communities. [contained red sticky dot indicating someone also supported this point]
- Need conversations (sic) with MIPs. Some variables have low usage because not consistently produced.
- Consult groups such as VIACS - what variables do they need to be able to derive from CMIP model output?
- IPCC WGII sector lead and regional engagements - to ensure the process is transparent
- Large scale post processing of model output is a key need (i.e. downscaling efforts, translation to more standard formats, derived quantities etc.) and should be greatly enhanced, and so we should be thinking about ways to minimise the effort needed for this (i.e. using standard pressure/depth levels, making sure that the needed variables are subset of the core etc.). *Response Expect this core to have some standardised sets*
- Face to model development evaluation process is more important.
- Would it be possible to engage with the "non-academic" (e.g. climate service community) and systematically collect their feedback on the initial selection? In case C3S would be happy to help.

- Different priorities from different communities. Modelling groups for different purposes - wide or specialist. Maybe a list for all modelling groups but more direction. Don't need all models to provide all variables to do decent intercomparison.
- Look at scientific and application requirements and then looking at the highest priority - common

Representation

Do we need broader representation in the paper authors or if authors survey the community, is that considered appropriate

- Authors should be interested in the process
- MIP leads - the current list emerged from this. Some MIPs are specialists but others represent wider community
- Representation across domains (impacts, oceans, clouds etc) in author team
- Representation of the different realms needs to be considered and transparent
- Open document
- Issue of reducing core list could reduce the use for the MIPs - requires serious consideration to allow for tailoring, must recognise this within the process.

Are there community members that should be asked to act as reviewers, not currently listed?

- Core set of variables for WGCM relevant variables rather than linked directly to CMIP7 specifically.
- MJ: Not something all MIPs need to be directly involved in.

Should we schedule a meeting for authors with reviewers prior to submission?

No comments made