https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

Geoscientific
Model Development
Discussions
Open Access

Designing and Documenting Experiments in CMIP6

Charlotte Pascoe[1,2], Bryan N Lawrence[2,3], Eric Guilyardi[2,4], Martin Juckes[1,2], and Karl E Taylor[5]
[1]STFC Rutherford Appleton Laboratory, UK.
[2]National Centre of Atmospheric Science, UK.
[3]Departments of Meteorology and Computer Science, University of Reading, UK.
[4]LOCEAN/IPSL, Sorbonne Université/CNRS/IRD/MNHN, Paris, France
[5]PCMDI, Lawrence Livermore National Laboratory, Livermore, CA, USA
Correspondence to: Charlotte Pascoe (charlotte.pascoe@ncas.ac.uk)

Abstract. Earth system modelling relies on contributions from groups who develop models and from those involved in devising, executing, and exploiting numerical experiments. Often these people work in different institutions, and they may communicate primarily via published information (whether journal papers, technical notes, or websites). The complexity of the models, experiments, and methodologies, along with the diversity (and sometimes inexact nature) of information sources can easily lead to misinterpretation of what was actually intended or done. In this paper we introduce a taxonomy of terms for more clearly defining numerical experiments, put it in the context of previous work on experimental ontologies, and describe how we have used it to document the CMIP6 experiments. We describe how this process involved iteration with a range of CMIP6 stakeholders to rationalise multiple sources of information and add clarity to experimental definitions. We demonstrate how this process has added value to CMIP6 itself by a) helping those devising experiments to be clear about their goals and expected methodology, b) making it easier for those executing experiments to know what was intended, c) exposing inter-relationships between experiments, and d) making it clearer for third parties (data users) to understand the CMIP6 experiments. We conclude with some lessons learned, and how these may be applied for any modelling campaign as well as future CMIP phases.

1 Introduction

Climate modelling involves the use of models to carry out simulations of the real world, usually as part of an experiment aimed at understanding processes, hypothesis testing, or projecting some future climate system behaviour. Such numerical experiments can be organized into “Model Intercomparison Projects” (MIPs) in which participants execute common experiments and share results. Perhaps the best known of these are the CMIP series of Climate Model Intercomparison Projects, of which the latest is CMIP6 (Eyring et al., 2016).

The design, documentation and accompanying protocols have all evolved over time, reflecting both an increasing scope and wider-spread interest, and two important new constituencies: (1) Those who have organised “Diagnostic MIPs”, which do not require new experiments, but rather request specific output from existing planned experiments to address specific interests; and (2) an even wider group of downstream users who use the CMIP data opportunistically, having little or no direct contact with the modelling groups.

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

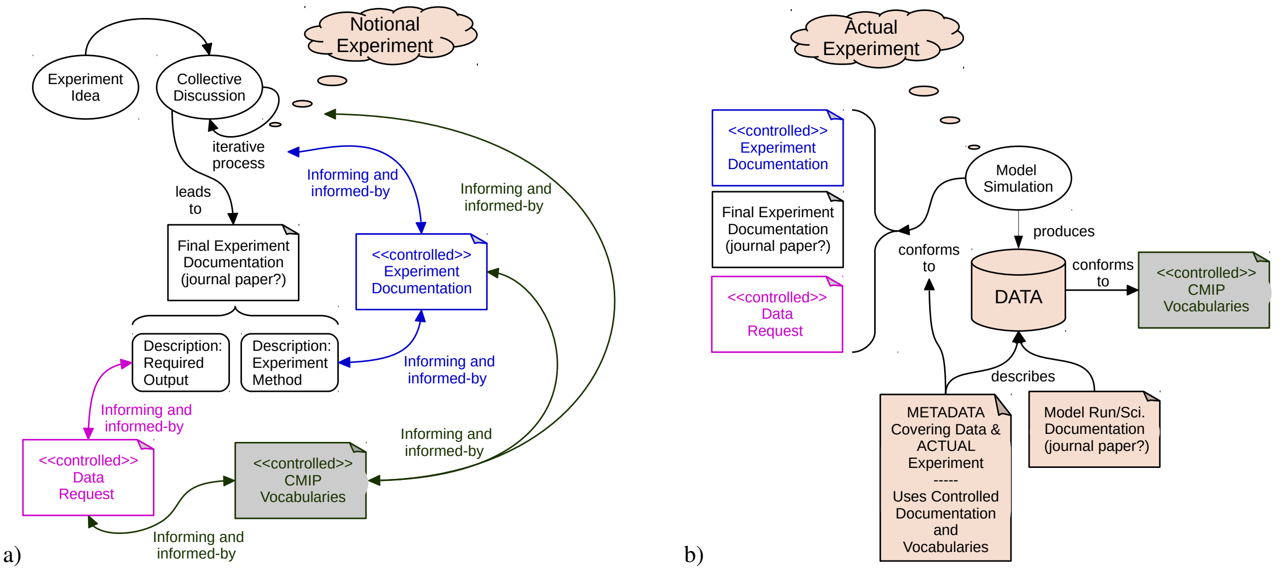

Figure 1. The process of defining an "experiment" involves multiple steps, interactions, and component descriptions. In a) we can see the various representations of an experiment during the design phase, from the initial idea to a final description (e.g., published in a journal paper) reached through an iterative process. In the case of CMIP6 this is accompanied by co-design of various essential components (the data request, the experiment documentation, and the CMIP vocabularies). In b) we see the realisation of these experiment descriptions in the form of a simulation, which may in practice deviate somewhat from the experiment as defined. Describing why the various components in (a) are needed, how competing approaches are reconciled, and how a final design is arrived at is a key goal of this paper. We will publish the methodology and experience with (b) elsewhere.

Over the years, the increase in complexity, size, and scope of CMIP has led to a requirement to improve in each phase of CMIP the documentation of the activity, from experiment specification to data output. CMIP5 addressed this in three ways: by documenting the experiment design in a detailed specification paper (Taylor et al., 2011); by improving documentation of metadata requirements and data layout to improve access to, and interpretation of, simulation output; and by requiring model participants to exploit the (then) "Metafor" system (Lawrence et al., 2012; Guilyardi et al., 2013; Moine et al., 2014) to describe their models and simulations.

Metafor was a qualified success; useful information was collected, but the tools were not able to be fully tested before use and were found to be difficult to use by those providing the documentation content. Such difficulties resulted in documentation generally arriving too late to be of use to the target audience: scientists analysing the data. The Metafor project has been superseded by the ES-DOC project (https://es-doc.org) which provides a much improved tool chain, and ES-DOC use is now required for the documentation of CMIP6 simulations (Balaji et al., 2018).

In this paper we describe how ES-DOC concepts have been applied in the design phase of CMIP6 to improve not only the eventual documentation of CMIP6 simulations, but also of the experiments themselves. As a consequence, we believe it will

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

be easier for both the MIP designers and participants to be confident that they have requested, understood, and/or executed experiments that will meet their scientific objectives.

We begin by introducing a vocabulary for describing the experiments and the simulations and put it in the context of other work. We then use the vocabulary to provide a high level summary of CMIP6 itself. We proceed to a description of how the CMIP6 MIPs were designed and linked to the fundamental CMIP core experiments - the so-called "DECK." We provide some examples of ES-DOC experiment descriptions, and then present some of the experiment linkages which can be understood from the use of our canonical experiment descriptions. Our experiences in gathering information and the linkages (and some of the missing links) required to define and document CMIP6 experiments expose opportunities for improving future MIP designs, which we present in the summary.

2 Experiment Definitions

The process of experimental definition is potentially complex (figure 1a). It begins with an idea and often entails an iterative community discussion which results in the final experimental documentation, which needs to cover at least the imposed experimental conditions and the required output. With the experiments defined, modelling groups carry out simulations which conform as best they can to the specifications of the experiment and which produce the desired output.

In the case of CMIP6, the iterative discussion includes input from the ES-DOC community aiming to get a formal experiment description, from the Data Request coordinator, and the CMIP6 central team at the Program for Climate Model Diagnosis and Intercomparison (PCMDI) responsible for cross-experimental common CMIP vocabularies. These extra activities result in additional documentation which can be used by those carrying out the actual experiment (figure 1b) in an attempt to minimise the burden of interpreting and carrying out many experiments.

2.1 Key concepts

To discuss numerical experiments it is necessary to have a vocabulary which clearly identifies the actions and artifacts of the workflow summarized in figure 2. Referring to the figure, we outline such a vocabulary: Model Intercomparison Projects design *Numerical Experiments* and define their *NumericalRequirements*. These experiment definitions are adopted by modelling groups who use a model to run *Simulations*, with *Output Data* requirements (“data requests”) being one of the many experimental requirements. A simulation is run with a *Configured Model*, with a configuration which will include details of *InputData* and may include *Modifications* required to conform to the experiment requirements. Not all of the configuration will be related to the experiment, aspects of the workflow and computing environment may also need to be configured. In some cases a simulation might not conform exactly to the requirements, so a key part of a simulation description is the set of *Conformance* descriptions which indicate how the simulation conforms to the experimental requirements. In this paper focusing on CMIP6, we are limiting our attention to the definition of the Experiment and its Requirements, and the relationship between the MIPs and those requirements.

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

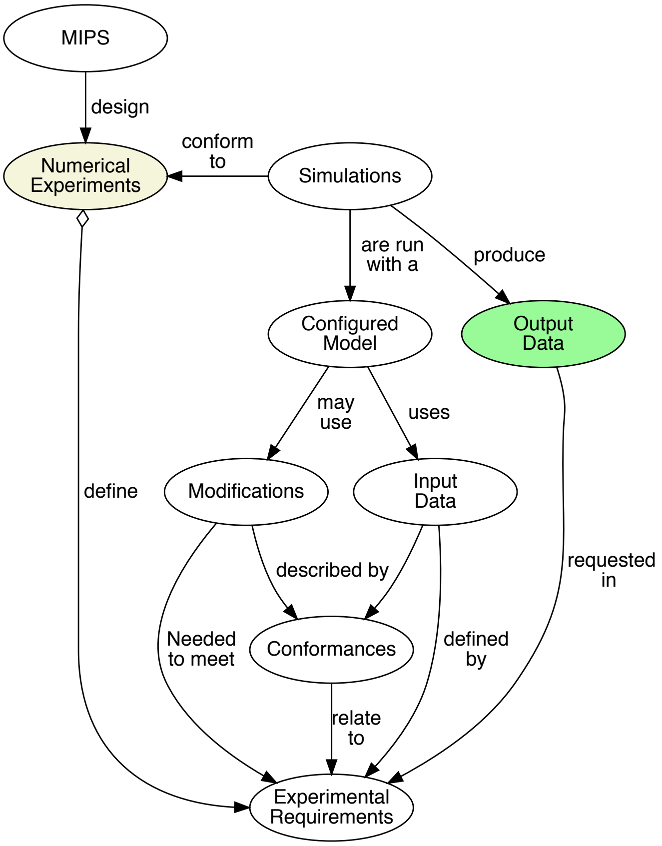

Figure 2. Simulation workflow in which experimental requirements (termed "Numerical Requirements") play a central role.

As noted above, a Project (formed by groups of individuals and commonly referred to generically as a "MIP") has certain scientific objectives that lead it to define one or more *NumericalExperiment*s. We describe the rules for performing the numerical experiments as *NumericalRequirement*s (figure 3). Both *NumericalExperiment*s and *NumericalRequirement*s may be nested and the former may also explicitly identify specific related experiments which may provide dependencies or other scientific context such as heritage. Nested requirements are used to bundle requirements together for easy re-use across experiments. An example nested requirement can be seen in table 5 (appendix) where all the components which go into pre-industrial solar particle forcing such as electron and cosmic-ray forcing and others are bundled together. We will see later that many implicit relationships arise from common requirements.

The experiment description itself includes attributes covering the scientific objective and the experiment rationale addressing the questions: What is this experiment for? Why is it being done?

The concept of "experiment" has shifted slightly in terms of the start-date ensembles used for the decadal hindcast experiments of the DCPP (Decadal Climate Prediction Project) MIP. For example, whereas experiments such as *decadal1995* and *decadal2000* were two distinct experiments in CMIP5, in CMIP6 they are ensemble members of the single experiment *dcppA-*

hindcast which has multiple realisations for each start date. Thus the DCPP experiments in CMIP6 are distinguished only by the science question they address.

2.2 Requirements

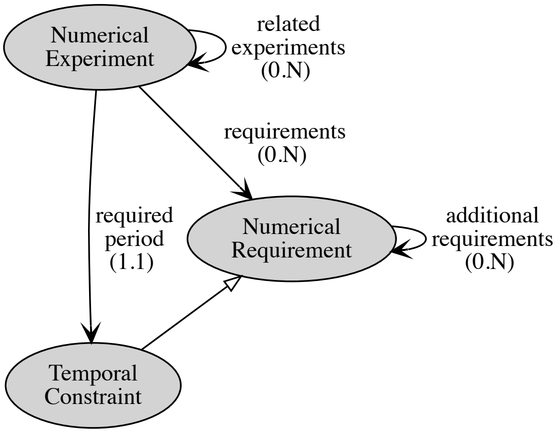

Figure 3. *Numerical Experiment*s are designed and governed by intercomparison projects. Each numerical experiment is defined by *Numerical Requirements*, including a mandatory constraint setting out the required period of the numerical experiment. Numerical requirements may have complicated internal structures (see fig4).

The *NumericalRequirement*s are the set of instructions required to configure a model and provide prescribed input in preparation for executing a simulation that conforms to a *NumericalExperiment*. These instructions include (figure 4) specifications such as the start date, simulation period, ensemble size and structure (if required), any forcings (e.g. external boundary conditions such as the requirement to impose a one percent increase in carbon dioxide over 100 years), initialisation requirements (e.g should the model be "spun-up" or initialised from the output of a simulation from another experiment), and domain requirements (for limited area models). A scope keyword from a controlled vocabulary can be used to indicate whether the requirement is re-used elsewhere, e.g. in the specifications for related experiments.

Each requirement carries a number of optional attributes and may contain mandatory attributes, as shown in tables 1 and 2 for a *ForcingConstraint*.

2.3 Related Work

The ES-DOC vocabulary is an instance of an ontology ("a formal specification of a shared conceptualisation", Borst 1997). There is considerable literature outlining the importance of such ontologies in establishing common workflow patterns with

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

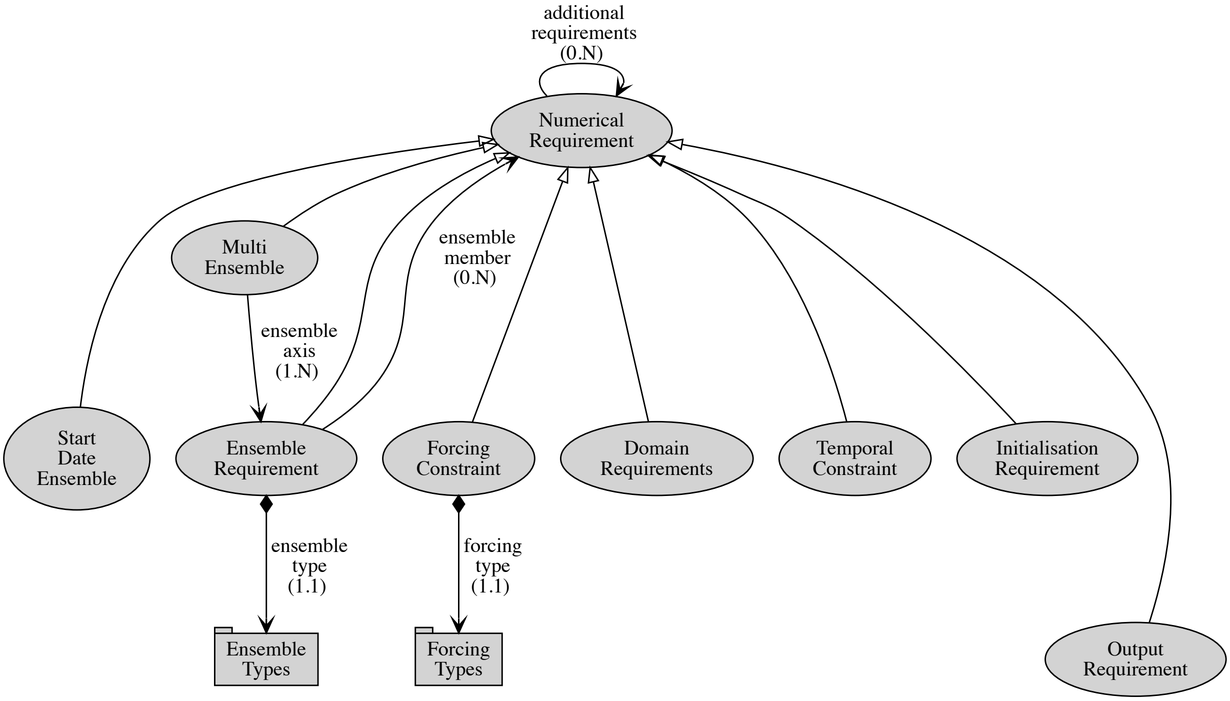

Figure 4. *NumericalRequirements* govern the structure of a numerical experiment covering constraints on duration (*TemporalConstraint*), the domain covered (*DomainRequirement*, e.g. global or a regional bounding box), any forcings (*ForcingConstraint*, such as particular greenhouse gas concentrations), output requirements (e.g. the CMIP6 data request), and a complicated interplay of potential *EnsembleRequirements* (see text). Controlled vocabularies are necessary for *EnsembleTypes*, *ForcingTypes*, and *NumericalRequirementScopes*.

Forcing Constraint			
attribute	**type**	**cardinality**	**description**
category	str	0.1	Category to which this belongs (from a CV, e.g. GASES).
code	str	0.1	Programme wide code from a controlled vocabulary (e.g. N2O).
data_link	*data.dataset*	0.1	A data record used by the forcing
forcing_type	*designing.forcing_types*	1.1	Type of integration
group	str	0.1	Sub-Category (e.g. GHG)
origin	*shared.citation*	0.1	Pointer to origin, e.g. CMIP6 RCP database.

Table 1. ES-DOC controlled structure for describing a forcing constraint: each attribute has a name, a type (those in italics are other ES-DOC types), a cardinality (0.1 means either zero or one, 1.1 means one is required) and a description.

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

Forcing Types	
keyword	**definition**
historical	Best estimates of actual state (included synthesized)
idealised	Simplified and/or exemplar, e.g. 1%C02
scenario	Intended to represent a possible future, e.g. RCP4.5
driven	Driven with data output from another simulation

Table 2. ES-DOC Forcing types controlled vocabulary, provides context for a forcing constraint.

the goal of improving reproduction of results and reuse of techniques, whether they be traditional laboratory experiments or *in silico*, explicitly calling out the failure of papers to provide the necessary information (e.g. Vanschoren et al. 2012 in the context of reproducible machine learning).

The description of ontologies is often presented in the context of establishing provenance for specific workflows, and often retrospectively. Little attention has been paid to the composition phase of workflows (Mattoso et al., 2010), let alone more abstract goals.

The full ES-DOC ontology has been significantly updated from that introduced in Lawrence et al. (2012), and will be fully described elsewhere. Here we concentrate on work relevant to the experiment descriptions discussed in this paper; these map directly onto part of the “conception phase” of workflow design, introduced in Mattoso et al. as part of a proposed description of “experiment life cycles”. In their view, the conception phase potentially consists of an abstract workflow, describing what should be done (but without specifying how), and a concrete workflow, binding abstract workflows to specific resources (models, algorithms, platforms, etc). ES-DOC respects that split with an explicit separation of design (experiment descriptions) and simulation (the act of using a configured model to produce data conforming to the constraints of an experiment).

The notion of “an experiment” also needs attention, since the experiments described here are even more abstract than the notion of “a workflow”, and cover a wider scope than that normally attributed to an experiment. Dictionary definitions of “scientific experiment” generally emphasise the relationship between hypothesis and experiment (e.g. “An experiment is a procedure carried out to support, refute, or validate a hypothesis. Experiments provide insight into cause-and-effect by demonstrating what outcome occurs when a particular factor is manipulated.”, Wikipedia contributors 2018). In this context “factor” has a special meaning, a factor generally being one of a few input variables; but in numerical modelling there can be a multiplicity of such factors, leading to difficulties in formal experimental definition and consistency of results (Zocholl et al. 2018 in the context of big data experiments).

The first formal attempt to define a generic ontology of experiments (as opposed to workflows), appears to be that of Soldatova and King (2007) (who also expressly identify the limitations of natural language alone for precision and disambiguation).

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

Key components of their ontology include the notions of experimental classification, design, results and their relationships, but it is not obvious how this ontology can be used to guide either conception or implementation. da Cruz et al. (2012) build on Mattoso et al. (2010) to specify more fully the abstract conception phase of workflow with more generic experiment concepts with much the same aim as Soldatova and King, however, they introduce many elements in common with ES-DOC and one could imagine some future mapping between these ontologies (although there is not yet any clear use case for this).

With the advent of simulation, another type of experiment (beyond those defined earlier) is possible: the simulation (and analysis) of events which cannot be measured, such as predictions of the state of a system influenced by factors which cannot be replicated (or which may be hypothetical, such as the climate on a planet with no continents). For climate science, the most important of these is of course the future; experiments can be used to predict possible futures (scenarios).

In this form of experiment, ES-DOC implicitly defines two classes of “controllable factor” being those controlled by the experiment design (and defined in *NumericalRequirement*s, in particular, by constraints) and those which are controlled by experiment implementation (the actual modelling system). Only the former are discussed here. Possibly because most of the existing work does not directly address this class of experiment, there is no similar clear split along these lines in the literature we have seen.

3 CMIP6

Global model intercomparison projects have a long history with pioneering efforts beginning in the late 1980’s (e.g., Cess et al. 1989 and Gates et al. 1999). The first phase of CMIP was initiated in the mid 1990s (Meehl et al., 1997). CMIP1 involved only a handful of modelling groups, but with each succeeding phase of CMIP, participation grew. The sixth phase (CMIP6), underway now, is expected to involve dozens of institutions, including all the major centres and many smaller modelling groups. Throughout the CMIP history, there has been a heavy reliance on CMIP results in the preparation of IPCC reports — CMIP1 diagnostics were linked to IPCC diagnostics and the timing of CMIP phases has been driven by the IPCC timelines.

With each phase, more complexity has been introduced. CMIP1 had three relatively simple goals: to document mean model climate errors, assess the ability of models to simulate variability, and assess flux adjustment (Sausen et al., 1988). CMIP6 continues to address the first two of these objectives (flux adjustment being rarely used in modern models), but with a broader emphasis on past, present and future climate in a variety of contexts covering process understanding, suitability for impacts and adaptation, and climate change mitigation.

In CMIP5 and again in CMIP6, there was a substantial increase in the number and scope of experiments. This has led to a new organizational framework in CMIP6 involving the distributed management of a collection of quasi-independently constructed Model Intercomparison Projects, which were required to meet requirements and expectations set by the overall coordinating body (the CMIP Panel) before they were "endorsed" as part of CMIP6.

At the heart of the current CMIP process is a central suite of experiments known as the DECK (Diagnosis, Evaluation, and Characterization of Klima (Eyring et al., 2016). The DECK includes a pre-industrial control under 1850 conditions, an atmosphere-only AMIP simulation with imposed historical sea surface temperatures, and two idealised CO2 forcing experi-

<table>
<tr><th colspan="2">DECK (CMIP6)</th></tr>
<tr><td colspan="2"><i>Diagnosis, Evaluation, and Characterization of Klima (Climate)</i></td></tr>
<tr><td colspan="2">Description: Core simulations for climate model intercomparison.</td></tr>
<tr><td colspan="2">Rationale: To maintain continuity and help document basic characteristics of models across different phases of CMIP.</td></tr>
<tr><td colspan="2">Experiments</td></tr>
<tr><td>esm-piControl: A pre-industrial control simulation with non-evolving pre-industrial conditions and atmospheric CO2 calculated. Conditions chosen to be representative of the period prior to the onset of large-scale industrialization, with 1850 being the reference year. The piControl starts after an initial climate spin-up, during which the climate begins to come into balance with the forcing. The recommended minimum length for the piControl is 500 years. To be performed with an Earth System Model (ESM) that can calculate atmospheric CO2 concentration and account for the fluxes of CO2 between the atmosphere, the ocean, and biosphere.</td><td>esm-piControl-spinup: A pre-industrial control spin-up simulation with non-evolving pre-industrial forcing and atmospheric CO2 calculated. Conditions chosen to be representative of the period prior to the onset of large-scale industrialization, with 1850 being the reference year. This experiment describes an initial climate spin-up, during which the climate begins to come into balance with the forcing. To be performed with an Earth System Model (ESM) that can calculate atmospheric CO2 concentration and account for the fluxes of CO2 between the atmosphere, the ocean, and biosphere. Run until Earth System reaches equilibrium.</td></tr>
<tr><td>piControl-spinup: A pre-industrial control spin-up simulation with non-evolving pre-industrial forcing. Forcing conditions are chosen to be representative of the period prior to the onset of large-scale industrialization, with 1850 being the reference year. This experiment describes an initial climate spin-up, during which the climate begins to come into balance with the forcing. Run until at least the surface climate reaches equilibrium.</td><td>piControl: A pre-industrial control simulation with non-evolving pre-industrial conditions. Conditions chosen to be representative of the period prior to the onset of large-scale industrialization, with 1850 being the reference year. The piControl starts after an initial climate spin-up, during which the climate begins to come into balance with the forcing. The recommended minimum length for the piControl is 500 years.</td></tr>
<tr><td>1pctCO2: Increase atmospheric CO2 concentration gradually at a rate of 1 percent per year. The concentration of atmospheric carbon dioxide is increased from the global annual mean 1850 value until quadrupling.</td><td>amip: An atmosphere only climate simulation using prescribed sea surface temperature and sea ice concentrations but with other conditions as in the Historical simulation.</td></tr>
<tr><td>abrupt-4xCO2: Impose an instantaneous quadrupling of the concentration of atmospheric carbon dioxide from the global annual mean 1850 value, then hold fixed.</td><td></td></tr>
</table>

Table 3. The experiments within the DECK, as described in ES-DOC. The content of this table, like all the ES-DOC tables in this paper, was generated directly from the online documentation using a python script (details in the appendix). The choice of content to display was made in the python code; other choices could be made (e.g., see https://documentation.es-doc.org/cmip6/mips/deck).

ments where in one CO2 is either increased by 1 percent per year until reaching 4 times the original concentration, while in the other CO2 is abruptly increased to 4 times the original concentration. Variants of most of these fundamental experiments have been core to CMIP since the beginning, and now within the DECK there are two variants of the pre-industrial control designed to test the relatively new earth system models which respond to internally calculated CO2 concentrations as opposed to responding to externally imposed CO2 concentration (table 3). Completion of the suite of DECK experiments is intended to serve as an entry card for model participation in the CMIP exercise. The CMIP project leaders are responsible for DECK design and definition, which should evolve only slowly over future phases of CMIP and will enable cross-generational model comparisons. CMIP is also responsible for the "historical" experiments, but the definition of these will change as better forcing data becomes available and as the historical period extends forward in time.

Table 4 provides a summary of most of the CMIP6 "endorsed" MIPs as of December 2018, with the DECK incorporated in CMIP as discussed above. It does not include the Coordinated Regional Climate Downscaling Experiment (CORDEX, Gutowski Jr. et al. 2016 or the three diagnostic MIPs - DYnVarMIP (Dynamics and Variability MIP, Gerber and Manzini 2016), SIMIP (Sea Ice MIP, Notz et al. 2016) and VIACSAB (Vulnerability, Impacts, Adaptation and Climate Services Advisory Board, Ruane et al. 2016), as these are not yet included in ES-DOC. There are of course other "non endorsed" MIPs such as ISA-MIP (the Interactive Stratospheric Aerosol MIP Timmreck et al. 2018) which could also be documented by ES-DOC at some future time.

3.1 Documentation and the MIP Design Process

An overview of the MIP design process is given in figure 1, which refers to the co-design process which involved the MIP teams, the CMIP6 team (both the CMIP panel[1] and the PCMDI support group[2]), the ES-DOC team, and the development of the data request[3]. The data request was an integral part of the process, since some MIPs are dependent on data produced in other MIPs, and the data is the key interface between the aspirations of the MIP and the community who exploit the MIP to deliver the science.

The semantic structure of the data request was developed in parallel to the development of the CMIP6 version of ES-DOC each had to deal with a distinctive range of complex expectations and requirements. Hence ES-DOC has not yet fully defined or populated the *OutputRequirement* shown in figure 4. Similarly, the Data Request was not able to fully exploit ES-DOC experiment descriptions. A future development will bring these together, and make use of the relationships between MIPs and between their output requirements and objectives. However, despite some semantic differences, there was communication between all parties throughout the definition phase.

The initial documentation was generated from a range of sources, and then iterated through the co-design phase which provided both challenges and opportunities. An example of the challenge was keeping track of material through the changing nomenclature — experiment names were changed through the process, some experiments were discarded and new experiments were added. In one case an experiment ensemble was formed from a set of hitherto separate experiments. Conversely, a key opportunity was the ability to influence MIP design to add focus and clarity, including influencing those very names. For example, the names of experiments which applied SST anomalies for positive and negative phases of ocean oscillation states were changed from "plus" to "pos" and "minus" to "neg" to better reflect the nature of the forcing and the relationship between experiment objectives and names).

---

[1] https://www.wcrp-climate.org/wgcm-cmip/cmip-panel

[2] see https://pcmdi.llnl.gov/CMIP6/

[3] https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest

CMIP6 (core MIPS recorded by ES-DOC)
AerChemMIP: Aerosols and Chemistry MIP - Collins et al. (2016)
C4MIP: Coupled Climate Carbon Cycle MIP - Jones et al. (2016)
CDRMIP: The Carbon Dioxide Removal Model Intercomparison Project - Keller et al. (2018)
CFMIP: Cloud Feedback Model Intercomparison Project - Webb et al. (2016)
CMIP: Climate Model Intercomparison Project - Eyring et al. (2016)
DAMIP: Detection and Attribution Model Intercomparison Project - Gillett et al. (2016)
DCPP: Decadal Climate Prediction Project - Boer et al. (2016)
FAFMIP: Flux-Anomaly-Forced Model Intercomparison Project - Gregory et al. (2016)
GMMIP: Global Monsoons Modeling Inter-comparison Project - Zhou et al. (2016)
GeoMIP: The Geoengineering Model intercomparison Project - Kravitz et al. (2015)
HighResMIP: High Resolution Model Intercomparison Project - Haarsma et al. (2016)
ISMIP6: Ice Sheet Model Intercomparison Project for CMIP6 - Nowicki et al. (2016)
LS3MIP: Land Surface, Snow and Soil Moisture MIP - van den Hurk et al. (2016)
LUMIP: Land-Use Model Intercomparison Project - Lawrence et al. (2016)
OMIP: Ocean Model Inter-comparison Project - Griffies et al. (2016)
PAMIP: Polar Amplification Model Intercomparison Project - Smith et al. (2018)
PMIP: Paleoclimate Modeling Intercomparison Project - Kageyama et al. (2018)
RFMIP: Radiative Forcing Model Intercomparison Project - Pincus et al. (2016)
ScenarioMIP: Scenario Model Intercomparison Project - O'Neill et al. (2016)
VolMIP: Model Intercomparison Project on the climatic response to Volcanic forcing - Zanchettin et al. (2016)

Table 4. The modelling CMIP6 experiments as introduced in Eyring et al. (2016) — except for CDRMIP and PAMIP which arrived later. This list does not include CORDEX or the diagnostic MIPs, which are not currently included in the ES-DOC MIP documentation.

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

The ES-DOC documentation process also raised a number of discrepancies and duplications, which were sorted out by conversations mediated by PCMDI. Many of the latter arose from independent development within MIPs of what eventually became shared experiments between those MIPs. For example, not all shared experiment opportunities were identified as such by the MIP teams, and it was the co-design process and the consolidated ES-DOC information which exposed the potential for shared experimental design (and significant savings in computational resources).

A specific example of such a saving occurred with ScenarioMIP and CDRMIP, which both included climate change overshoot scenario experiments that examine the influence of CO2 removal (negative net emissions) from 2040-2100 following unmitigated baseline scenarios through to 2040. As originally conceived, the ScenarioMIP experiment (ssp534-over) utilised year 2040 from the CMIP6 updated RCP8.5 for initialisation, but the CDRMIP equvalent (esm-ssp534-over) requested initialisation in 2015 fom the esm-historical experiment. In developing the ES-DOC descriptions of these experiment it was apparent that CDRMIP could follow the ScenarioMIP example and initialise from the C4MIP experiment esm-ssp585 in 2040 and avoid 25 years of unnecessary simulation (by multiple groups). This is now the recommended protocol.

Discrepancies also arose from the parallel nature of the workflow. For example, specifications could vary between what was published in the GMDD paper, and what had been agreed by the MIP authors with the Data Request and/or the PCMDI team with the controlled vocabulary. On occasion ES-DOC publication exposed such issues, resulting in revisions all round. This process required the sustained attention of representatives of each of these groups, and eventually resulted in a notification system which exploits Slack[4] so that all involved are notified of updates (but in most cases it requires initiation by a human identifying an issue). However, synchronicity was and is a problem, with quite different timescales involved in each of the processes. For example, the formal literature itself was evolving, and so version control has been important — all current ES-DOC documents cite the literature as it was during the co-design phase, and will be updated as necessary. Names too were a problem, with experiment names evolving, or specified differently within a MIP than in the wider CMIP6, leading to issues in both documentation and specification. A rather late addition to the taxonomies supported by both ES-DOC and PCMDI is support for aliases, to try and minimise this issue.

The co-design process had other outcomes too: LUMIP originally had a set of experiments that were envisaged to address the impact of particular behaviours such as “grass crops with human fire management”. Some of these morphed to became entirely the opposite of their original incarnation, such as “land_noFIRE” where the experiment requires no human fire management (see table 6). This sort of change prompted discussions about how to describe experiments that are built around the concept of missing out one or more processes. If you have a suite of experiments that require that the land scheme is run without specific process or phenomena w, x, y and z, can we define the individual experiments in the suite with the form “not w but with x, y and z” and “not x but with w, y and z”? To which the answer was no, they should simply be described as “not w” and “not x”. It turns out that as yet there isn’t much uniformity about how land models are set up, each is very different, so it only makes sense for LUMIP to constrain the experiments in this suite in terms of the phenomena that are not included.

---

[4] https://slack.com/

3.2 Forcing Contraints in Practice

Somewhat naively, the initial concepts for *ForcingConstraints* anticipated the description of forcing in terms of specific input boundary conditions, or perhaps, specific modifications needed to models — this was how they were described for the CMIP5 documentation. The ES-DOC semantics introduced for CIMP6 are more inclusive and allowed a wider range of possible forcing constraints. For example, in CMIP5 the infamous ES-DOC questionnaire asked modellers to describe how they implemented solar forcing. In CMIP6, the approach to solar forcing requirements were outlined in the literature (Matthes et al., 2017), and the resulting requirements are found in rather more precise forcing constraints (with additional related requirements) — an example of which appears in table 5. The ES-DOC documentation now provides a checklist of important requirements and a route to the literature for both those implementing the experiments and those interpreting their results.

Increasing precision is evident throughout CMIP6 and in the documentation. In some cases, rather than ask how it is done in a model post fact, the experiment definition describes what is expected, as in the GeoMIP experiment G7SST1-cirrus (table 7) where explicit modelling instructions are provided. However, where appropriate, experiments still leave it open to modelling groups to choose their own methods of implementing constraints, e.g. the reduction in aerosol forcing described in GeoMIP experiment G6sulpher (table 8).

Unexpected constraints also included the “anti-pattern” forcing constraint introduced in the the example of land_noFire above: an experimental constraint emphasizing the lack of a specific phenomenon (or in this case, parametrised behaviour).

4 Experimental Relationships

CMIP6 is more than just an assemblage of unrelated MIPs. One of the beneficial outcomes of the formal documentation of CMIP6 within ES-DOC has been a clearer understanding of the dependencies of MIPs on each other, and of experiments on shared forcing constraints.

4.1 Common Experiments

Figure 5 shows the sharing of experiments between MIPs. The importance of *piControl*, *historical*, *AMIP*, key scenario experiments (*ssp245* and *ssp585*), and the idealised experiments (*1pctCO2* and *abrupt-4xCO2*) is clear. These seven experiments form part of the protocol for many of the CMIP6 MIPs (figure 7). The scope of the *historical* and *piControl* experiments is demonstrated by their connections to MIPs on the far edges of the plot in all directions.

There are other shared experiments too, which bring MIPs together around shared scientific goals: *land-hist* jointly defined and shared by LUMIP and L3SMIP; *past1000* defined by PMIP forms part of VolMIP; *piClim-control* defined by RFMIP forms part of AerChemMIP; and *dcppC-forecast-addPinatubo* defined by DCPP forms part of VolMIP. By contrast, OMIP stands alone, sharing no experiments with other MIPs.

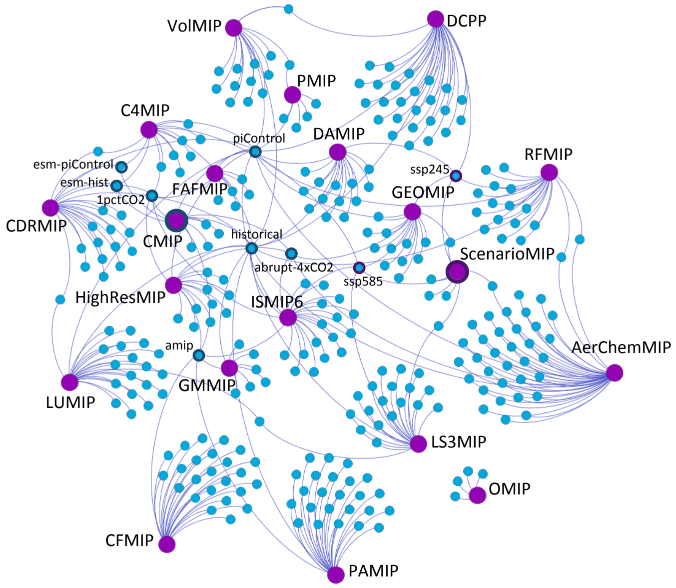

Figure 5. *CMIP6 MIPs and experiments.* Individual MIPs are represented by large purple dots. Lines connect each MIP to the experiments that are related to it, which are shown as smaller blue dots. Some widely used experiments are labelled, such as the piControl, historical, amip, ssp245 and ssp585, which are used by numerous MIPs within CMIP6.

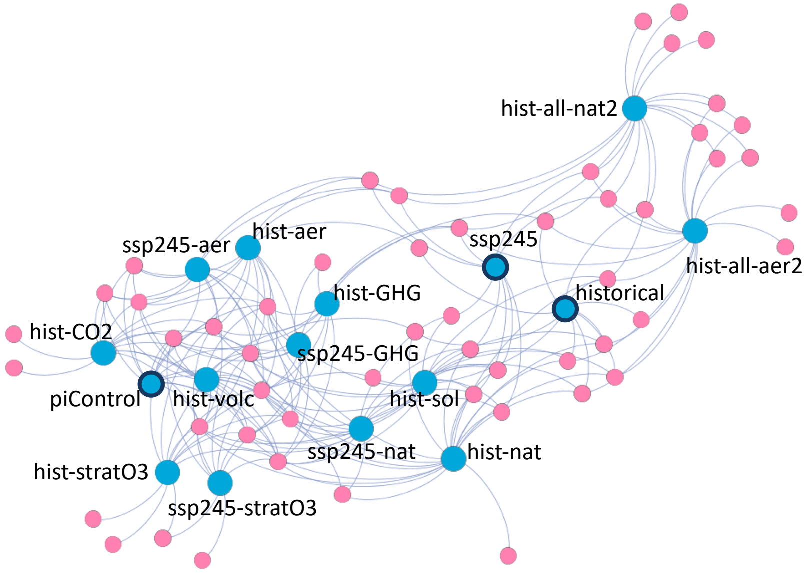

Figure 6. *DAMIP experiments and forcing constraints.* Individual experiments are represented by large blue dots. Lines connect each experiment to the forcing constraints that are related to it. The three experiments with dark blue borders (piControl, historical and ssp245) are required by DAMIP but not defined by it. The forcing constraints for these three external experiments are used extensively by the DAMIP experiment suite.

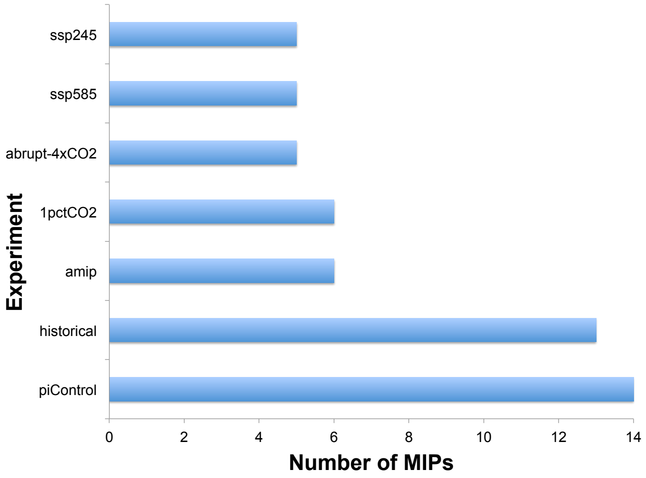

Figure 7. The most-used CMIP6 experiments in terms of the number of Model Intercomparison Projects (MIPs) to which they contribute.

4.2 Common Forcing

Experiments share forcing constraints, just as MIPs share experiments. Figure 6 shows the interdependence of the DAMIP experiments on common forcing constraints. Experiments are grouped near each other when they share forcing constraints. The dense network shown reflects the similarity of experiments within DAMIP, and arises from a common design pattern/protocol in numerical experiment construction: a new experiment is a variation on a previous experiment with one (or a few) forcing changes. It is of course this "perturbation experiment" pattern which provides much of the strength of simulation in exposing causes and effects in the real world.

Unique modifications appear in figure 6 as forcing constraint nodes that are only connected to one or two experiments which is also why the alternative forcing experiments *hist-all-nat2* and *hist-all-aer2* are placed further from the main body of the DAMIP network — they share fewer forcing constraints with the other experiments. However, they themselves are similar to each other as between them they share a number of unique forcing constraints.

The importance of the perturbation experiment pattern is further emphasised in DAMIP by noting that the three external experiments (piControl, historical and ssp245) account for 62 percent of the DAMIP forcing constraints; five of the DAMIP experiments can be completely described by forcing constraints associated with these external experiments — being different assemblies of the same "forcing building blocks". The key role of these building blocks is exposed by placing the DAMIP experiments into sets according to which of those external experiments is used for forcing constraints (figure 8).

This framing of shared forcing constraints exposes some apparent anomalies. Why, for example, is *hist-CO2* not in the "historical" set)? The reasons for these apparent anomalies expose the framing of the experiments. In the *historical* experiment, greenhouse gas forcing is a single constraint which includes CO2 and other well mixed greenhouse gases. By contrast, *hist-*

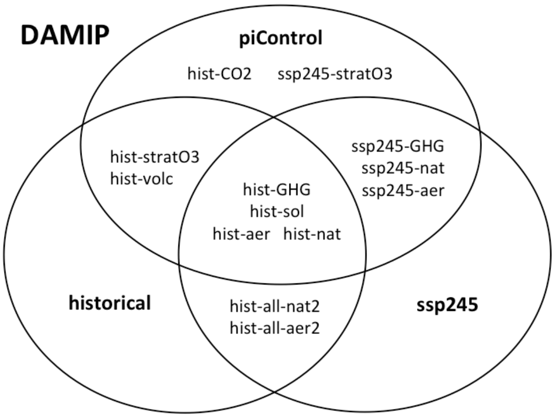

Figure 8. A view of *DAMIP* with experiments placed in sets according to the forcing constraints they share with the external experiments: piControl, historical and ssp245.

CO2 varies only CO2, with the other well mixed greenhouse gases constrained to pre-industrial levels (and hence uses the *piControl* forcing constraints for those, with it's own CO2 forcing constraint).

It would have been possible to avoid this sort of anomaly by constructing finer constraints in the case of *historical*, but this would have been at the cost of simplicity of understanding (and greater multiplicity in reporting as discussed below). There is a necessary balance between clear guidance on experimental requirements, and re-use of such constraints to expose relationships between experiments.

4.3 Forcing Constraint Conformance

One of the goals of the constraint formalism is to minimise the burden on modelling groups of both executing the CMIP6 experiments, and documenting how the experiments were carried out (that is, populating the concrete part of the experiment definition, using the language of Mattoso et al. 2010, as discussed in section 2.3). By clearly identifying commonalities between experiments, modelling groups can implement constraints once, and reuse both the implementation and documentation across experiments.

Constraint "conformance" documentation is intended to provide clear targets for interpreting the differences between simulations carried out with different models. Given that differing constraints often define differing experiments, understanding why models give different results can be aided by understanding differences in constraint implementation (in those cases where there is implementation flexibility). Section 3.2 discussed some aspects of this from a constraint definition perspective.

One can then ask, how much re-use of constraints is possible? Figure 9, shows some re-use, but unfortunately most are not re-used. Of the 476 forcing constraints identified during the documentation of CMIP6, 265 are only used once by a single experiment.

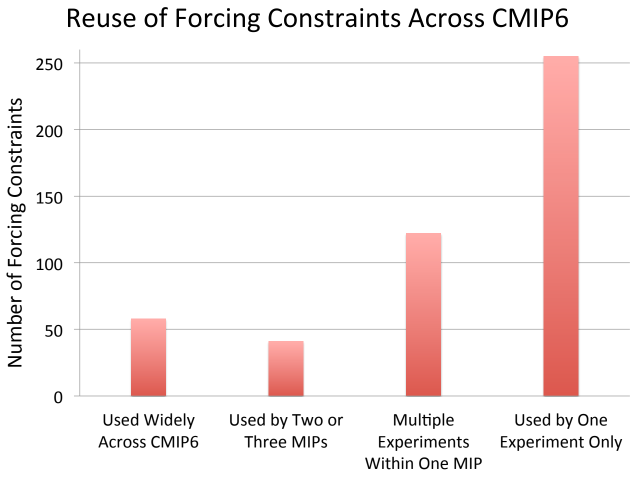

Figure 9. *Distribution of forcing constraint reuse across CMIP6.* Forcing constraints are categorised in terms of how widely they are used. Widely used forcing constraints are used by experiments in four or more MIPs.

4.4 Temporal Constraints

Temporal constraints are mandatory in the ES-DOC design, as history suggests that there has been — and continues to be — unnecessary ambiguity in expected simulation duration. This often manifests itself by delivering “off by one” differences in the number of years of simulation expected (arising from start and/or end date ambiguity), which could be an expensive proposition in computer time (and the associated energy costs). However, despite being mandatory, there is very little re-use of temporal constraints within CMIP6 (indeed, the duration and start dates vary considerably across the experiments, even though some standardisation might have been possible).

5 Summary and Further Work

We have introduced a formal taxonomy for experimental definition based around collections of climate modelling projects (MIPs), experiments, and numerical requirements and, in particular, constraints of one form or another. These provide structure for the formal definition of the experiment goals, design and method. The conformance, model and simulation definitions (to be fully defined elsewhere) will provide the concrete expression of how the experiments were executed.

The co-design of CMIP6, involving MIP teams, with coordination provided at various stages by the CMIP panel and PCMDI, has improved on previous MIP exercises, albeit with a larger increase in process and still with opportunities for imprecision, duplication of design effort, and unnecessary requirements on participants. The ES-DOC experiment definitions provided another route to internal review of the co-design, and aided in identifying and removing some of the imprecision, duplication of effort, and simulation requirements. However, there is still scope for improving the design phase. Earlier involvement of formal documentation, would have facilitated more interaction between the MIP design teams by requiring more information to be shared earlier. Doing so in the future might allow more common design patterns, and perhaps more experiment and simulation re-use between MIPs, reducing the burden on both carrying out the simulations, and on storing the results. This potential gain would need to be evaluated and tensioned against the potential process burden, but it can be seen that the ES-DOC experiment/requirement/constraint definitions are relatively lightweight, yet communicate significant precision of objective and method.

Although sharing could be improved (particularly of temporal constraints) sharing of experiments and constraints is clearly common within CMIP6. Section 2.3 outlines a set of important relationships between the MIPs, and MIP dependency on key experiments — most of which are in the CMIP (and DECK) sub-project. Such sharing introduces extra problems of governance: who owns the shared experiment definition? In the case of the dependencies on the DECK, this is clear, it is the CMIP panel, but for other cases it is not so clear. For example, both LS3MIP and LUMIP needed a historical land experiment, and it was obvious it should be shared. In this case (and hopefully most cases) the solution was amicable, but not really ideal for downstream users (e.g.”Start year either 1850 or 1700 depending on standard practice for particular model. This experiment is shared with the LS3MIP, note that LS3MIP expects the start year to be 1850.”). If sharing is to be enhanced in future CMIP exercises, then their early identification (and resolution of any related governance issues) will be necessary.

The sharing and visualisation of constraint dependencies (section 4.2) provides a route to both efficient execution and better understanding of experimental structure. In the case of DAMIP there is clear value to the interpretation of the MIP goals in terms of the forcing constraints, and this sort of analysis could both be extended to other MIPs, and used during future design phases. While there is a trade-off between granularity of forcing and the burden of conformance documentation, with CMIP6 this trade-off was never explicitly considered. In the future it is possible that such consideration may in fact improve experimental design.

ES-DOC remains a work in progress. It is fair to say that there was not wide community acceptance of the burden of documentation for CMIP5, but this was in part because of the tooling available then. With the advent of CMIP6, the tooling is much enhanced, and available much earlier in the cycle — but both the underlying semantic structure and tooling can and will be improved. There is clearly opportunity of convergence between the Data Request and ES-DOC and there will undoubtedly be much community feedback to take on board!

ES-DOC is not intended to apply only to CMIP exercises. We believe the experiment design and methodology, as well as the publication of experimental methods, should be of use even when only one or a few models generate related simulations — one such target will be the sharing of national resources to deliver the larger expensive simulations (in time, resource, and energy) where individuals and small communities could not justify the expense without sharing goals and outputs. Realising such sharing opportunities is often impaired by insufficient communication and documentation. We believe the ES-DOC methodology can go some way towards to alleviating these missed opportunities, and will become essential as we contemplate using significant portions of future exascale machines.

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

6 Code Availability

All the underlying ES-DOC code is publicly available at https://github.com/es-doc. The full CMIP6 documentation is available online at https://search.es-doc.org/. The code to extract and produce the ES-DOC tables in this paper is available online at https://github.com/bnlawrence/esdoc4scientists (Lawrence, 2019). Figures 5 and 6 were produced using triples generated from ES-DOC and imported into gephi (https://gephi.org/) with manual annotations.

Author contributions. CP represented ES-DOC in the experiment co-design, collecting information and influencing design. MJ was responsible for the data request. KT led the PCMDI involvement in experiment co-design. EG and BL led various aspects of ES-DOC at different times. BL and CP wrote the bulk of this paper, with contributions from the other authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Clearly the CMIP6 design really depends on the many scientists involved in designing and specifying the experiments under the purview of the CMIP6 panel. The use of ES-DOC to describe experiments depends heavily on the tool chain, much of which was designed and implemented by Mark Morgan under the direction of Sébastien Denvil (CNRS/IPSL). Paul Durack was instrumental in the support for CMIP6 vocabularies at PCMDI. Most of the ES-DOC work described here has been funded by national capability contributions to NCAS from the UK Natural Environment Research Council (NERC) and by the European Commission under FW7 grant agreement 312979 for IS-ENES2. Work by Karl E. Taylor was performed under the auspices of the US Department of Energy (USDOE) by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344 with support from the Regional and Global Modeling Analysis Program of the USDOE's Office of Science.

References

Balaji, V., Taylor, K. E., Juckes, M., Lawrence, B. N., Durack, P. J., Lautenschlager, M., Blanton, C., Cinquini, L., Denvil, S., Elkington, M., Guglielmo, F., Guilyardi, E., Hassell, D., Kharin, S., Kindermann, S., Nikonov, S., Radhakrishnan, A., Stockhause, M., Weigel, T., and Williams, D.: Requirements for a Global Data Infrastructure in Support of CMIP6, Geoscientific Model Development, 11, 3659–3680, doi:https://doi.org/10.5194/gmd-11-3659-2018, 2018.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The Decadal Climate Prediction Project (DCPP) Contribution to CMIP6, Geosci. Model Dev., 9, 3751–3777, doi:10.5194/gmd-9-3751-2016, 2016.

Borst, W. N.: Construction of Engineering Ontologies for Knowledge Sharing and Reuse, Ph.D. thesis, University of Twente, Centre for Telematics and Information Technology (CTIT), Enschede, 1997.

Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Ghan, S. J., Kiehl, J. T., Treut, H. L., Li, Z.-X., Liang, X.-Z., Mitchell, J. F. B., Morcrette, J.-J., Randall, D. A., Riches, M. R., Roeckner, E., Schlese, U., Slingo, A., Taylor, K. E., Washington, W. M., Wetherald, R. T., and Yagai, I.: Interpretation of Cloud-Climate Feedback as Produced by 14 Atmospheric General Circulation Models, Science, 245, 513–516, doi:10.1126/science.245.4917.513, 1989.

Collins, W. J., Lamarque, J.-F., Schulz, M., Boucher, O., Eyring, V., Hegglin, M. I., Maycock, A., Myhre, G., Prather, M., Shindell, D., and Smith, S. J.: AerChemMIP: Quantifying the Effects of Chemistry and Aerosols in CMIP6, Geosci. Model Dev., 10, 585–607, doi:10.5194/gmd-10-585-2017, 2017.

da Cruz, S. M. S., Campos, M. L. M., and Mattoso, M.: A Foundational Ontology to Support Scientific Experiments, in: Proceedings of the VII International Workshop on Metadata, Ontologies and Semantic Technology (ONTOBRAS-MOST), edited by Malucelli, A. and Bax, M., Recife, Brazil, 2012.

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

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., Fiorino, M., Gleckler, P. J., Hnilo, J. J., Marlais, S. M., Phillips, T. J., Potter, G. L., Santer, B. D., Sperber, K. R., Taylor, K. E., and Williams, D. N.: An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I), Bulletin of the American Meteorological Society, 80, 29–56, doi:10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2, 1999.

Gerber, E. P. and Manzini, E.: The Dynamics and Variability Model Intercomparison Project (DynVarMIP) for CMIP6: Assessing the Stratosphere–Troposphere System, Geosci. Model Dev., 9, 3413–3425, doi:10.5194/gmd-9-3413-2016, 2016.

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B. D., Stone, D., and Tebaldi, C.: The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) Contribution to CMIP6, Geosci. Model Dev., 9, 3685–3697, doi:10.5194/gmd-9-3685-2016, 2016.

Gregory, J. M., Bouttes, N., Griffies, S. M., Haak, H., Hurlin, W. J., Jungclaus, J., Kelley, M., Lee, W. G., Marshall, J., Romanou, A., Saenko, O. A., Stammer, D., and Winton, M.: The Flux-Anomaly-Forced Model Intercomparison Project (FAFMIP) Contribution to CMIP6: Investigation of Sea-Level and Ocean Climate Change in Response to CO2 Forcing, Geosci. Model Dev., 9, 3993–4017, doi:10.5194/gmd-9-3993-2016, 2016.

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., Chassignet, E. P., Curchitser, E., Deshayes, J., Drange, H., Fox-Kemper, B., Gleckler, P. J., Gregory, J. M., Haak, H., Hallberg, R. W., Heimbach, P., Hewitt, H. T., Holland, D. M., Ilyina, T., Jungclaus, J. H., Komuro, Y., Krasting, J. P., Large, W. G., Marsland, S. J., Masina, S., McDougall, T. J., Nurser, A. J. G., Orr, J. C., Pirani, A., Qiao, F., Stouffer, R. J., Taylor, K. E., Treguier, A. M., Tsujino, H., Uotila, P., Valdivieso, M., Wang, Q., Winton, M., and Yeager, S. G.: OMIP Contribution to CMIP6: Experimental and Diagnostic Protocol for the Physical Component of the Ocean Model Intercomparison Project, Geosci. Model Dev., 9, 3231–3296, doi:10.5194/gmd-9-3231-2016, 2016.

Guilyardi, E., Balaji, V., Lawrence, B., Callaghan, S., Deluca, C., Denvil, S., Lautenschlager, M., Morgan, M., Murphy, S., and Taylor, K. E.: Documenting Climate Models and Their Simulations, Bulletin of the American Meteorological Society, 94, 623–627, doi:10.1175/BAMS-D-11-00035.1, 2013.

Gutowski Jr., W. J., Giorgi, F., Timbal, B., Frigon, A., Jacob, D., Kang, H.-S., Raghavan, K., Lee, B., Lennard, C., Nikulin, G., O'Rourke, E., Rixen, M., Solman, S., Stephenson, T., and Tangang, F.: WCRP COordinated Regional Downscaling EXperiment (CORDEX): A Diagnostic MIP for CMIP6, Geosci. Model Dev., 9, 4087–4095, doi:10.5194/gmd-9-4087-2016, 2016.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., Chang, P., Corti, S., Fučkar, N. S., Guemas, V., von Hardenberg, J., Hazeleger, W., Kodama, C., Koenigk, T., Leung, L. R., Lu, J., Luo, J.-J., Mao, J., Mizielinski, M. S., Mizuta, R., Nobre, P., Satoh, M., Scoccimarro, E., Semmler, T., Small, J., and von Storch, J.-S.: High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6, Geosci. Model Dev., 9, 4185–4208, doi:10.5194/gmd-9-4185-2016, 2016.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., and Zaehle, S.: C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: Experimental Protocol for CMIP6, Geosci. Model Dev., 9, 2853–2880, doi:10.5194/gmd-9-2853-2016, 2016.

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Peterschmitt, J.-Y., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Brierley, C., Crucifix, M., Dolan, A., Fernandez-Donado, L., Fischer, H., Hopcroft, P. O., Ivanovic, R. F., Lambert, F., Lunt, D. J., Mahowald, N. M., Peltier, W. R., Phipps, S. J., Roche, D. M., Schmidt, G. A., Tarasov, L., Valdes, P. J., Zhang, Q., and Zhou, T.: The PMIP4 Contribution to CMIP6 – Part 1: Overview and over-Arching Analysis Plan, Geosci. Model Dev., 11, 1033–1057, doi:10.5194/gmd-11-1033-2018, 2018.

Keller, D. P., Lenton, A., Scott, V., Vaughan, N. E., Bauer, N., Ji, D., Jones, C. D., Kravitz, B., Muri, H., and Zickfeld, K.: The Carbon Dioxide Removal Model Intercomparison Project (CDRMIP): Rationale and Experimental Protocol for CMIP6, Geoscientific Model Development, 11, 1133–1160, doi:10.5194/gmd-11-1133-2018, 2018.

Kravitz, B., Robock, A., Tilmes, S., Boucher, O., English, J. M., Irvine, P. J., Jones, A., Lawrence, M. G., MacCracken, M., Muri, H., Moore, J. C., Niemeier, U., Phipps, S. J., Sillmann, J., Storelvmo, T., Wang, H., and Watanabe, S.: The Geoengineering Model Intercomparison Project Phase 6 (GeoMIP6): Simulation Design and Preliminary Results, Geosci. Model Dev., 8, 3379–3392, doi:10.5194/gmd-8-3379-2015, 2015.

Lawrence, B.: Bnlawrence/Esdoc4scientists: Tools for Creating Publication Ready Tables from ESDOC Content, doi:https://doi.org/10.5281/zenodo.2593445, 2019.

Lawrence, B., Balaji, V., Bentley, P., Callaghan, S., DeLuca, C., Denvil, S., Devine, G., Elkington, M., Ford, R. W., Guilyardi, E., Lautenschlager, M., Morgan, M., Moine, M.-P., Murphy, S., Pascoe, C., Ramthun, H., Slavin, P., Steenman-Clark, L., Toussaint, F., Treshansky, A., and Valcke, S.: Describing Earth System Simulations with the Metafor CIM, Geoscientific Model Development, 5, 1493–1500, doi:10.5194/gmd-5-1493-2012, 2012.

Lawrence, D. M., Hurtt, G. C., Arneth, A., Brovkin, V., Calvin, K. V., Jones, A. D., Jones, C. D., Lawrence, P. J., de Noblet-Ducoudré, N., Pongratz, J., Seneviratne, S. I., and Shevliakova, E.: The Land Use Model Intercomparison Project (LUMIP) Contribution to CMIP6: Rationale and Experimental Design, Geosci. Model Dev., 9, 2973–2998, doi:10.5194/gmd-9-2973-2016, 2016.

Matthes, K., Funke, B., Andersson, M. E., Barnard, L., Beer, J., Charbonneau, P., Clilverd, M. A., de Wit, T. D., Haberreiter, M., Hendry, A., Jackman, C. H., Kretzschmar, M., Kruschke, T., Kunze, M., Langematz, U., Marsh, D. R., Maycock, A. C., Misios, S., Rodger, C. J., Scaife, A. A., Seppälä, A., Shangguan, M., Sinnhuber, M., Tourpali, K., Usoskin, I., van de Kamp, M., Verronen, P. T., and Versick, S.: Solar Forcing for CMIP6 (v3.2), Geosci. Model Dev., p. 56, 2017.

Mattoso, M., Werner, C., Travassos, G. H., Braganholo, V., Ogasawara, E., Oliveira, D. D., Cruz, S. M. S. D., Martinho, W., and Murta, L.: Towards Supporting the Life Cycle of Large Scale Scientific Experiments, International Journal of Business Process Integration and Management, 5, 79, doi:10.1504/IJBPIM.2010.033176, 2010.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J.: Intercomparison Makes for a Better Climate Model, Eos, Transactions American Geophysical Union, 78, 445–451, doi:10.1029/97EO00276, 1997.

Moine, M.-P., Valcke, S., Lawrence, B. N., Pascoe, C., Ford, R. W., Alias, A., Balaji, V., Bentley, P., Devine, G., Callaghan, S. A., and Guilyardi, E.: Development and Exploitation of a Controlled Vocabulary in Support of Climate Modelling, Geoscientific Model Development, 7, 479–493, doi:10.5194/gmd-7-479-2014, 2014.

Notz, D., Jahn, A., Holland, M., Hunke, E., Massonnet, F., Stroeve, J., Tremblay, B., and Vancoppenolle, M.: The CMIP6 Sea-Ice Model Intercomparison Project (SIMIP): Understanding Sea Ice through Climate-Model Simulations, Geosci. Model Dev., 9, 3427–3446, doi:10.5194/gmd-9-3427-2016, 2016.

Nowicki, S. M. J., Payne, A., Larour, E., Seroussi, H., Goelzer, H., Lipscomb, W., Gregory, J., Abe-Ouchi, A., and Shepherd, A.: Ice Sheet Model Intercomparison Project (ISMIP6) Contribution to CMIP6, Geosci. Model Dev., 9, 4521–4545, doi:10.5194/gmd-9-4521-2016, 2016.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geosci. Model Dev., 9, 3461–3482, doi:10.5194/gmd-9-3461-2016, 2016.

Pincus, R., Forster, P. M., and Stevens, B.: The Radiative Forcing Model Intercomparison Project (RFMIP): Experimental Protocol for CMIP6, Geosci. Model Dev., 9, 3447–3460, doi:10.5194/gmd-9-3447-2016, 2016.

Ruane, A. C., Teichmann, C., Arnell, N. W., Carter, T. R., Ebi, K. L., Frieler, K., Goodess, C. M., Hewitson, B., Horton, R., Kovats, R. S., Lotze, H. K., Mearns, L. O., Navarra, A., Ojima, D. S., Riahi, K., Rosenzweig, C., Themessl, M., and Vincent, K.: The Vulnerability, Impacts, Adaptation and Climate Services Advisory Board (VIACS AB v1.0) Contribution to CMIP6, Geosci. Model Dev., 9, 3493–3515, doi:10.5194/gmd-9-3493-2016, 2016.

Sausen, R., Barthel, K., and Hasselmann, K.: Coupled Ocean-Atmosphere Models with Flux Correction, Climate Dynamics, 2, 145–163, doi:10.1007/BF01053472, 1988.

Smith, D. M., Screen, J. A., Deser, C., Cohen, J., Fyfe, J. C., García-Serrano, J., Jung, T., Kattsov, V., Matei, D., Msadek, R., Peings, Y., Sigmond, M., Ukita, J., Yoon, J.-H., and Zhang, X.: The Polar Amplification Model Intercomparison Project (PAMIP) Contribution to CMIP6: Investigating the Causes and Consequences of Polar Amplification, Geoscientific Model Development Discussions, pp. 1–42, doi:https://doi.org/10.5194/gmd-2018-82, 2018.

Soldatova, L. and King, R. D.: An Ontology of Scientific Experiments, Journal of the Royal Society, Interface / the Royal Society, 3, 795–803, doi:10.1098/rsif.2006.0134, 2007.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: A Summary of the CMIP5 Experiment Design, Tech. rep., PCMI, Livermore, 2011.

Timmreck, C., Mann, G. W., Aquila, V., Hommel, R., Lee, L. A., Schmidt, A., Brühl, C., Carn, S., Chin, M., Dhomse, S. S., Diehl, T., English, J. M., Mills, M. J., Neely, R., Sheng, J., Toohey, M., and Weisenstein, D.: The Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP): Motivation and Experimental Design, Geoscientific Model Development, 11, 2581–2608, doi:10.5194/gmd-11-2581-2018, 2018.

van den Hurk, B., Kim, H., Krinner, G., Seneviratne, S. I., Derksen, C., Oki, T., Douville, H., Colin, J., Ducharne, A., Cheruy, F., Viovy, N., Puma, M. J., Wada, Y., Li, W., Jia, B., Alessandri, A., Lawrence, D. M., Weedon, G. P., Ellis, R., Hagemann, S., Mao, J., Flanner, M. G., Zampieri, M., Materia, S., Law, R. M., and Sheffield, J.: LS3MIP (v1.0) Contribution to CMIP6: The Land Surface, Snow and Soil Moisture Model Intercomparison Project – Aims, Setup and Expected Outcome, Geosci. Model Dev., 9, 2809–2832, doi:10.5194/gmd-9-2809-2016, 2016.

Vanschoren, J., Blockeel, H., Pfahringer, B., and Holmes, G.: Experiment Databases, Machine Learning, 87, doi:10.1007/s10994-011-5277-0, 2012.

Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., Chepfer, H., Douville, H., Good, P., Kay, J. E., Klein, S. A., Marchand, R., Medeiros, B., Siebesma, A. P., Skinner, C. B., Stevens, B., Tselioudis, G., Tsushima, Y., and Watanabe, M.: The Cloud Feedback Model Intercomparison Project (CFMIP) Contribution to CMIP6, Geosci. Model Dev., 10, 359–384, doi:10.5194/gmd-10-359-2017, 2017.

Wikipedia contributors: Experiment — Wikipedia, The Free Encyclopedia, 2018.

Zanchettin, D., Khodri, M., Timmreck, C., Toohey, M., Schmidt, A., Gerber, E. P., Hegerl, G., Robock, A., Pausata, F. S. R., Ball, W. T., Bauer, S. E., Bekki, S., Dhomse, S. S., LeGrande, A. N., Mann, G. W., Marshall, L., Mills, M., Marchand, M., Niemeier, U., Poulain, V., Rozanov, E., Rubino, A., Stenke, A., Tsigaridis, K., and Tummon, F.: The Model Intercomparison Project on the Climatic Response to Volcanic Forcing (VolMIP): Experimental Design and Forcing Input Data for CMIP6, Geosci. Model Dev., 9, 2701–2719, doi:10.5194/gmd-9-2701-2016, 2016.

Zhou, T., Turner, A. G., Kinter, J. L., Wang, B., Qian, Y., Chen, X., Wu, B., Wang, B., Liu, B., Zou, L., and He, B.: GMMIP (v1.0) Contribution to CMIP6: Global Monsoons Model Inter-Comparison Project, Geosci. Model Dev., 9, 3589–3604, doi:10.5194/gmd-9-3589-2016, 2016.

Zocholl, M., Camossi, E., Jousselme, A.-L., and Ray, C.: Ontology-Based Design of Experiments on Big Data Solutions, in: Proceedings of the Posters and Demos Track of the 14th International Conference on Semantic Systems (SEMPDS), edited by Khalili, A. and Koutraki, M., vol. 2198, p. 4, ceur-ws.org, Vienna, Austria, 2018.

Appendix 1 - Examples

To improve readability, a number of examples are provided in this appendix, rather than where first referenced in the main text.

All these tables are produced by a python script. The ES-DOC pyesdoc[5] library is used to obtain the documents and instantiate them as python objects with access to CIM attributes via instance attributes with CIM property names. These can then be used to populate HTML

---

[5] https://pypi.org/project/pyesdoc/

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

tables described using jinja2[6] templates which are then converted to PDF for inclusion in the document using the weasyprint [7] package. This methodology is more fully described in the code (Lawrence, 2019).

<table>
<tr><th colspan="2">abrupt-4xCO2 (CMIP, DECK, AerChemMIP, GeoMIP, HighResMIP, ISMIP6)</th></tr>
<tr><td colspan="2">*Abrupt quadrupling of the atmospheric concentration of carbon dioxide*</td></tr>
<tr><td colspan="2">**Description:** Impose an instantaneous quadrupling of the concentration of atmospheric carbon dioxide from the global annual mean 1850 value, then hold fixed.</td></tr>
<tr><td colspan="2">**Rationale:** To evaluate the equilibrium climate sensitivity of the model and to diagnose the strength of various feedbacks. To characterise the radiative forcing that arises from an increase in atmospheric CO2 as well as changes that arise indirectly due to the warming.</td></tr>
<tr><td colspan="2">**Requirements**</td></tr>
<tr><td>**Pre-Industrial Solar Particle Forcing**: Pre-Industrial solar particle forcing (1850-1873 mean). For models with interactive stratospheric chemistry. Proton forcing: HOx and NOx production by solar protons. Electron forcing: Kp- or Ap-index to describe ionisation from electron precipitation in the lower thermosphere and upper mesosphere. Cosmic ray forcing: ion-pair production by galactic cosmic rays. CMIP6 models that do not have interactive chemistry should prescribe the CMIP6 recommended ozone forcing data set.
Additional Requirements:
• Pre-Industrial Proton Forcing
• Pre-Industrial Electron Forcing
• Pre-Industrial Cosmic Ray Forcing
• Pre-Industrial Ozone Concentrations</td><td>**Pre-Industrial Forcing Excluding CO2 and Solar**: Pre-Industrial forcing excluding carbon dioxide (CO2) and solar forcing.
Additional Requirements:
• Pre-Industrial Well Mixed Greenhouse Gas (WMGHG) Concentrations excluding CO2
• Pre-Industrial Aerosols
• Pre-Industrial Aerosol Precursors
• Pre-Industrial Ozone Concentrations
• Pre-Industrial Stratospheric Water Vapour Concentrations
• Pre-Industrial Stratospheric Aerosol
• Pre-Industrial Land Use</td></tr>
<tr><td>**Pre-Industrial Solar Irradiance Forcing**: Pre-Industrial solar forcing. The standard solar forcing dataset recommended for usage is the solar reference scenario dataset which includes pre-industrial solar forcing (1850-1873 mean). Includes total solar irradiance, F10.7 cm solar radio flux, and spectral solar irradiance for 10-100000 nm range.</td><td>**Abrupt 4xCO2 Increase**: Impose an instantaneous quadrupling of atmospheric carbon dioxide concentration, then hold fixed.</td></tr>
<tr><td>**PreIndustrialInitialisation**: Initialisation from a January in the pre-industrial control simulation.</td><td>**AOGCM Configuration**: Use a coupled Atmosphere-Ocean general circulation model</td></tr>
<tr><td>**SingleMember**: One ensemble member</td><td>**150yrs**: Run for 150 years.</td></tr>
</table>

Table 5. The abrupt 4XCO2 experiment is integral to a number of MIPs. (Not all properties are shown, see http://documentation.es-doc.org/cmip6/experiments/abrupt-4xCO2 for more details.)

---

[6] http://jinja.pocoo.org/

[7] https://weasyprint.org/

land-noFire (LUMIP)
historical land-only with no human fire land management
Description: Land surface model simulation. Same as land-hist except with fire management maintained at 1850 levels. Start year either 1850 or 1700 depending on standard practice for particular model.
Rationale: To assess the relative impact of land cover and incremental land management change on fluxes of water, energy, and carbon in combination with other LUMIP land experiments.
Requirements
1700-2014 315yrs: Historical, from 1700 to 2014.
1850-2014 165yrs: Historical, pre-Industrial to present
Historical GSWP3 Meteorological Forcing: Apply Global Soil Wetness Project phase three (GSWP3) forcing data for offline land surface models running the LS3MIP historical simulation land-hist is provided by the LS3MIP.
Historical Land Use: Apply the global gridded land-use forcing datasets to link historical land-use data and future projections. This new generation of "land use harmonization" (LUH2) builds upon past work from CMIP5, and includes updated inputs, higher spatial resolution, more detailed land-use transitions, and the addition of important agricultural management layers.
Historical land surface forcings except fire management: Apply all transient historical forcings that are relevant for the land surface model except for fire management.
1850 Fire Management: Maintain 1850 levels of fire management (anthropogenic ignition and suppression of fire). If ignitions are based on population density, maintain constant population density.
SingleMember: One ensemble member
LSM Configuration: Offline land surface model
All Land Management Active: All applicable land management active in the land surface model configuration.

Table 6. This is an experiment that has an anti-forcing "Historical Land Surface Forcings Except Fire Management" (note also two temporal constraint options "Start year either 1850 or 1700 depending on standard practice for particular model."). See https://documentation.es-doc.org/cmip6/experiments/land-NoFire for more information.

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

G7SST1-cirrus (GeoMIP)
SSTs from year 2020 of SSP5-8.5; forcings and other prescribed conditions from year 2020 of SSP5-8.5 + cirrus thinning
Description: Time slice at year 2020 of GeoMIP G7cirrus. Run for 10 years.
Rationale: To assess radiative forcing of G7cirrus at the beginning of the simulation (2020).
Requirements
2020-2029 10yrs: Timeslice, begin in 2020 and run for 10 years.
Increase Cirrus Sedimentation Velocity: Add a local variable that replaces (in all locations where temperature is colder than 235K) the ice mass mixing ratio in the calculation of the sedimentation velocity with a value that is eight times the original ice mass mixing ratio. Cirrus seeding to begin in 2020 and continue through to the year 2100.
SSP5-85 SST 2020: Sea surface temperature climatology calculated from the ScenarioMIP SSP5-85 experiment for the year 2020.
SSP5-85 SIC 2020: Sea ice concentration climatology calculated from the ScenarioMIP SSP5-85 experiment for the year 2020.
RCP85 Forcing: Impose RCP8.5 forcing. Represents the high end of the range of plausible future forcing pathways. *Additional Requirements:* • Representative Concentration Pathway 8.5 Well Mixed Greenhouse Gases • Representative Concentration Pathway 8.5 Short Lived Gas Species • Representative Concentration Pathway 8.5 Aerosols • Representative Concentration Pathway 8.5 Aerosol Precursors • Representative Concentration Pathway 8.5 Land Use
SingleMember: One ensemble member
SSP5-85Initialisation2020: Initialisation is from the beginning of year 2020 of the SSP5-8.5 experiment.
AGCM Configuration: An Atmosphere only general circulation model configuration.

Table 7. The "Increase Cirrus Sedimentation Velocity" forcing constraint is very precise about the change to be made to the "Add a local variable that replaces (in all locations where temperature is colder than 235K) the ice mass mixing ratio in the calculation of the sedimentation velocity with a value that is eight times the original ice mass mixing ratio". See https://documentation.es-doc.org/cmip6/experiments/g7sst1-cirrus for more information.

https://doi.org/10.5194/gmd-2019-98
Preprint. Discussion started: 5 June 2019

G6sulfur (GeoMIP)	
stratospheric sulfate aerosol injection to reduce net forcing from SSP585 to SSP245	
Description: Injection of sulfate aerosol precursors in the equatorial stratosphere to reduce the radiative forcing of the ScenarioMIP high forcing scenario (SSP5-85) to match that of the ScenarioMIP medium forcing scenario (SSP2-45). Geoengineering will be simulated over years 2020 to 2100.	
Rationale: To evaluate a climate in which geoengineering is used to only partially offset climate change in order to reduce the burden of adaptation. Assess the climate effects and inter-model variations of a limited amount of geoengineering as part of a portfolio of responses to climate change. Results to be compared with G6solar to determine differences between sulfate aerosol effects and solar irradiance effects.	
Requirements	
Internal Stratospheric Aerosol Precursors RCP85 to RCP45: Injection of stratospheric sulfate aerosol precursors to reduce the radiative forcing of ScenarioMIP high forcing scenario (SSP5-85) to match that of the ScenarioMIP medium forcing scenario (SSP2-45). Modelling groups that have an internal sulfate aerosol treatment should calibrate the radiative response to sulfate aerosols individually so that the results will be internally consistent (a procedure that will be more difficult for models that have a complex microphysical treatment of aerosols). Potential methods include a double radiation call, once with and once without the stratospheric aerosols, and also the use feedback methods. Simulations to be conducted as if the aerosols or aerosol precursors are emitted in a line from 10°S to 10°N along a single longitude band (0°). The injected aerosols or aerosol precursors should be evenly spread across model layers between 18 and 20 km. Note that sedimentation processes and self-lofting due to heating are likely to result in the aerosols being distributed between 16-25 km in altitude.	**External Stratospheric Aerosol Precursors RCP85 to RCP45**: Injection of stratospheric sulfate aerosol precursors to reduce the radiative forcing of ScenarioMIP high forcing scenario (SSP5-85) to match that of the ScenarioMIP medium forcing scenario (SSP2-45). For modelling groups that have no dynamical treatment of sulfate aerosols, GeoMIP will provide a data set of aerosol optical depth, as well as ozone fields that are consistent with this aerosol distribution. Note that the amount of sulfate injection needed for a given model to achieve the goals of this experiment may vary, so modelling groups should scale the aerosol and ozone perturbation fields as necessary. Simulations to be conducted as if the aerosols or aerosol precursors are emitted in a line from 10°S to 10°N along a single longitude band (0°). The injected aerosols or aerosol precursors should be evenly spread across model layers between 18 and 20 km. Note that sedimentation processes and self-lofting due to heating are likely to result in the aerosols being distributed between 16-25 km in altitude.
RCP85 Forcing: Impose RCP8.5 forcing. *Additional Requirements:* • Representative Concentration Pathway 8.5 Well Mixed Greenhouse Gases • Representative Concentration Pathway 8.5 Short Lived Gas Species • Representative Concentration Pathway 8.5 Aerosols • Representative Concentration Pathway 8.5 Aerosol Precursors • Representative Concentration Pathway 8.5 Land Use for Shared Socioeconomic Pathway 5	**SSP5-85Initialisation2020**: Initialisation is from the beginning of year 2020 of the SSP5-8.5 experiment.
AOGCM Configuration: Use a coupled Atmosphere-Ocean general circulation model	**2020-2100 81yrs**: Scenario, from 2020 to the end of the 21st century
SingleMember: One ensemble member	

Table 8. GeoMIP is clear about what the forcing should achieve (reduction in radiative forcing from rcp8.5 to rcp4.5) but leave it open to the modelling groups to choose a method that best suits their aerosol scheme. See https://documentation.es-doc.org/cmip6/experiments/g6sulfur for more information.