

## Credibility via Coupling: Institutions and Infrastructures in Climate Model Intercomparisons

MATTHEW S. MAYERNIK  
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH (NCAR)  
UNITED STATES

### Abstract

This study investigates Model Intercomparison Projects (MIPs) as one example of a coordinated approach to establishing scientific credibility. MIPs originated within climate science as a method to evaluate and compare disparate climate models, but MIPs or MIP-like projects are now spreading to many scientific fields. Within climate science, MIPs have advanced knowledge of: a) the climate phenomena being modeled, and b) the building of climate models themselves. MIPs thus build scientific confidence in the climate modeling enterprise writ large, reducing questions of the credibility or reproducibility of any single model. This paper will discuss how MIPs organize people, models, and data through institution and infrastructure coupling (IIC). IIC involves establishing mechanisms and technologies for collecting, distributing, and comparing data and models (infrastructural work), alongside corresponding governance structures, rules of participation, and collaboration mechanisms that enable partners around the world to work together effectively (institutional work). Coupling these efforts involves developing formal and informal ways to standardize data and metadata, create common vocabularies, provide uniform tools and methods for evaluating resulting data, and build community around shared research topics.

### Keywords

institutions; infrastructures; data; reproducibility; credibility; intercomparison project, climate science; climate model

### Introduction

... there should be no such thing as a theory of how credibility is achieved, at least in the sense of one of those grand theories that would offer an adequate formula for how it is done regardless of setting and the nature of the case at hand. In any particular case the resources and tactics relevant to the achievement of credibility are likely to be very diverse, and a different array of resources and tactics is likely to bear on different types of case ([Shapin 1995, 261](#)).

Questions about the credibility of scientific research abound within public discourse about science. Multiple efforts have sprung up to address credibility questions related to the “reproducibility” of scholarly research,

---

Copyright © 2021 (Matthew S. Mayernik). Licensed under the Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International (CC BY-NC-ND 4.0). Available at [estsjournal.org](https://estsjournal.org).

**To cite this article:** Mayernik, Matthew S. 2021. “Credibility via Coupling: Institutions and Infrastructures in Climate Model Intercomparisons.” *Engaging Science, Technology, & Society* 7.2: 10–32.  
<https://doi.org/10.17351/ests2021.769>.

To email contact Matthew S. Mayernik: [mayernik@ucar.edu](mailto:mayernik@ucar.edu).

including high-profile cases of academic fraud, inaccessible data, and the use of questionable statistical methodologies ([Moylan & Kowalczyk 2016](#); [Miyakawa 2020](#)). Digging deeper into commentary on reproducibility within scholarly research, it is clear that reproducibility is a multivalent concept, including both global dimensions, e.g. ensuring scientific integrity through the verification of results, and local dimensions, e.g. facilitating collaboration ([Borgman 2015](#); [Feinberg et al. 2020](#)). Discussions about reproducibility invoke some subset of a group of complex issues, including (but not limited to): accessibility, accuracy, credibility, integrity, quality, reliability, transparency, validity, and verifiability of scientific findings. As such, “reproducibility” tends to serve as a generalization for these other important concerns. In sociological and linguistic terms, the term “reproducibility” serves as a “gloss,” that is, “a formulation which, on its occurrence, is quite adequate, but which turns out to have been incomplete, ambiguous, even misleading” ([Jefferson 1985, 462](#)).

Problems arise when complex issues—like reproducibility that gloss many other concepts—are used to motivate policy, technological, and organizational changes. This is where the epigraph above from Steven Shapin comes into play. Talking about reproducibility as a single thing tends to suggest that there could be a “grand theory” of how reproducibility could be achieved. As Shapin notes, however, achieving credibility, or any concepts glossed by “reproducibility,” is inherently situation specific. Thus, there is a benefit to developing understandings of ways in which different scholarly communities, groups, and individuals achieve credibility, reproducibility, transparency, etc.

Recent works by Sabina Leonelli ([2018](#)) and Bart Penders, Britt Holbrook, and Sarah de Rijcke ([2019](#)) break down the broad concept of “reproducibility” into six or more specific types. If we take Shapin’s point seriously, that there should be no “grand theory” of reproducibility, each type of reproducibility in such typologies should be studied individually, with due examination of the attendant people, organizations, institutions, technologies, data, and work practices involved.

Commentary on the “crisis of reproducibility” within academic research tends to focus on individual research studies or the work of individual scholars. In several research areas, however, groups of researchers come together to define coordinated initiatives based around common methods, data, and/or research questions. In such initiatives, the direct reproducibility of individual contributions are subsumed by the ability to compare and collate results from many contributors. The main research question for this paper is: *how is the credibility of scientific findings and data achieved in organized community research endeavors?*

Model Intercomparison Projects (MIPs) provide an example of a coordinated approach to establishing scientific credibility. MIPs originated within climate science as a method to evaluate and compare disparate climate models. MIPs or MIP-like projects are now spreading to many scientific fields. Within climate science, MIPs have advanced knowledge of: a) the climate phenomena being modeled, and b) the building of climate models themselves ([Wilson 2021](#)). MIPs thus build confidence in the science and policy components of the climate modeling enterprise, downsizing questions of the credibility or reproducibility of any single model.

This paper uses document analysis, ethnographic research, and participant observation to build on Eric Winsberg’s argument that “climate science is, in a thorough-going way, a socially organized kind of science, and that many features of its epistemology need to be surveyed at the social level in order to be properly understood” ([Winsberg 2018, 209–210](#)). It incorporates insights from peer-reviewed literature, gray literature, and websites by and about MIPs over the past forty years. This analysis is also informed by

my position within the National Center for Atmospheric Research (NCAR), through which I have engaged in ethnographic and participant observation intermittently over the past ten years in workshops, seminars, small group meetings, and informal interactions with scientists, technology and data experts. I also incorporate quotes from a series of semi-structured interviews conducted in 2017 (at NCAR).

The following sections discuss how MIPs organize people, models, and data through institution and infrastructure coupling (IIC). IIC involves establishing mechanisms and technologies for collecting, distributing, and comparing data and models (infrastructural work), alongside corresponding governance structures, rules of participation, and collaboration mechanisms that enable partners around the world to work together effectively (institutional work). Coupling these efforts involves developing formal and informal ways to standardize data and metadata, create common vocabularies, provide uniform tools and methods for evaluating resulting data, and build community around shared research topics. I argue that this coupling of the institutional and infrastructural work is instrumental in enabling MIPs to achieve credible research outcomes.

### **Background—Credibility and Reproducibility Overview**

STS literature vividly depicts how the reproducibility, accuracy, and credibility of scientific instruments and findings depend on the movement of people with specific standing and expertise. Replications of scientific findings encounter “regress” challenges, namely, difficulties in knowing whether experimental or theoretical findings are real if they have never been encountered before ([Collins 1985](#); [Kennefick 2007](#)). This can lead to debates that span years or even decades, across multiple generations of people and research studies ([Galison 1987](#)). Work to resolve such debates occurs at both individual and institutional levels ([Braun & Kropp 2010](#)).

Most scholarly communities, and in particular the geosciences ([Yan et al. 2020](#)), do not organize themselves to achieve “reproducibility” per se. As noted in a recent report by National Academies of Sciences, Engineering, and Medicine (NASEM): “A predominant focus on the replicability of individual studies is an inefficient way to assure the reliability of scientific knowledge. Rather, reviews of cumulative evidence on a subject, to assess both the overall effect size and generalizability, is often a more useful way to gain confidence in the state of scientific knowledge” ([NASEM 2019, 2](#)). Cumulative evidence gives credibility to research findings even if strong reproducibility is difficult or impossible to achieve.

Credibility is used here to refer to the quality of engendering trust or belief, and of being convincing or inspiring confidence. This understanding of credibility is aligned with both dictionary definitions and past sociology of science research, such as that of Shapin ([1995](#)). Organized community research endeavors often arise due to challenges in reproducing or replicating certain findings. Climate science provides an excellent case of how groups of researchers and other stakeholders design for comparison and cumulative evidence. As recently stated by a group of climate science experts:

numerical reproducibility is difficult to achieve with the computing arrays required by modern GCMs [General Circulation Models]. . . . Therefore, the focus of the discipline has not been on model run reproducibility, but rather on replication of model phenomena observed and their magnitudes, which is performed mostly in organized multi-model ensembles ([Bush et al. 2020, 10](#)).

MIPs are a high-profile example of a multi-model ensemble. Organized at an international scale, MIPs provide a venue through which dozens of computational modeling teams can compare, evaluate, and diagnose their models. In the literature focused on reproducibility, such as Leonelli (2018), computational simulation-based research projects are often grouped together under the broad notion of “computational reproducibility.” Penders et al. (2019)—in extending Leonelli’s typology—list simulations as having “high” reproducibility due in part to an “absolute” level of control over the research environment.

The challenge with climate models, as noted by Bush et al., is that climate models simulate chaotic phenomena. Thus, re-running a model with the same configuration and input data may not provide the same bit-level output. Also, climate models are optimized to run on particular hardware and software systems (supercomputers) and can produce different output when run on other computational platforms (Easterbrook 2014). For example, in a workshop held in May 2020 focused on geoscience model output archiving and reproducibility (of which I was a co-convenor), questions about bit-wise reproducibility were quickly set aside by participants as being not useful due to the difficulty of achieving bit-wise equivalent simulations and the low value for modelers in doing so. Instead, discussions focused on “feature reproducibility,” namely, whether the same physical phenomena (or statistics about those phenomena) could be seen on subsequent simulations. Examples include whether tornadoes emerged in simulations consistently when the same atmospheric conditions were present, or whether geographic temperature trends were consistent across long-term climate simulations. Within climate model research papers and policy documents, the term “reproducibility” is most commonly used in this way, e.g. the extent to which models can reproduce observed temperature trends or particular well-known climate phenomena such as El Niño.

The MIP approach to organizing research has more in common with Leonelli’s (2018) third type of reproducibility, “Semi-Standardized Experiments.” As described by Leonelli, in this category, “[research] methods, set-up, and materials used have been construed with ingenuity in order to yield very specific outcomes, and yet some significant parts of the set-up necessarily elude the controls set up by experimenters.” (ibid., 136). In this category, research is not aimed for direct reproducibility, but instead emphasize other things, including comparability, validity, and predictability.

MIPs complement other considerations involved assessing the reproducibility and credibility of climate models. Climate models contain multiple sources of uncertainty, and validating their outputs necessarily require multiple approaches (Randall et al. 2007). Decisions within climate modeling centers about model development and assessment are influenced by the centers’ objectives, conceptual assumptions, community norms, and the availability of funding and computational resources (Morrison 2021). Climate model results are evaluated quantitatively and qualitatively via comparisons to observations, to known physical laws (such as conservation of energy), and to prior generations of climate models (Rood 2019). MIPs do not encompass all sources of model variation and uncertainty. In general, any model that can meet the requirements of a given MIP can participate, resulting in “ensembles of opportunity,” rather than a random or systematic sample of all possible Earth system models (Winsberg 2018). Other kinds of simulation ensembles, such as large ensembles of simulations from a single model, are better suited for studying the internal variability of the climate system or of an individual model’s components (Deser et al. 2020). MIPs do, however, provide robust social and technical scaffolding to reduce certain kinds of variation and uncertainty, and to buttress the establishment of credibility in climate models more broadly (Leonelli

2019). Such social and technical scaffolding is critical to enabling climate science to meet external demands for transparency, accountability, and credibility (Edwards 2019b; Mayernik 2019).

### Institutional And Infrastructural Coupling (IIC)

Before going further into the MIP case, this section provides an outline of the key concepts of this paper, namely, infrastructures and institutions. A full review of the extensive literature that exists around each concept is beyond the scope of this article. Here, however, I discuss a few characteristics related to each concept to frame the rest of the study.

As developed by Susan Leigh Star and Karen Ruhleder (1996), groupings of technical systems, human practices, and organizations can be studied as infrastructure if they present certain characteristics, including—that they are embedded within other social arrangements and technologies, are built upon an installed base of prior systems, and are typically invisible to the user until they break down. Infrastructures are also deeply connected with routines and habits involved in their use. Their “invisibility” often comes from the ways that habits and norms fade into the background of routine interactions with built systems (Edwards, 2019a). “Infrastructure” is thus a concept that denotes human-built networks of technical systems that underpin distributed sets of human practices and movements of material entities (Edwards et al. 2007). This paper follows the recommendation of Charlotte Lee and Kjeld Schmidt—to delineate the scope of the concept of “infrastructure” more precisely (2018). Lee and Schmidt’s analysis details how the gradual expansion of the concept has led to terminological vagueness and conceptual imprecision. They depict how these problems emerge from Star and Ruhleder’s initial studies, and carry forward through subsequent studies of “information infrastructures” and “cyberinfrastructure.” The result of this imprecision is that “the term ‘infrastructure’ can be used to mean just about anything. This licenses not just semantic drift but a conceptual landslide” (Lee & Schmidt 2018, 191-2). The range of entities that have been characterized as “infrastructure” within various literatures is indeed remarkable, encompassing even the sky and non-human animals (Hoeppe 2018; Barua 2021). This begs the question posed by Lee and Schmidt: “Does ‘infrastructure’ then simply mean the infinite assortment of stuff upon which a practice, any practice, relies?” (Lee & Schmidt 2018, 192).

To avoid these conceptual challenges, I use a narrower view of infrastructure, one of four from Lee and Schmidt: “a technical structure or installation or material substrate ([e.g.] ‘networked computing’) conceived of in terms of its structure and the services it provides to some social system” (ibid., 207). This definition maps well to types of infrastructures that are commonly used as examples of the concept, such as the electricity grid, the US interstate highway system, the telephone networks (wired and cellular), and the internet. I thus use the term “infrastructural work” to refer to the work required by people and organizations to establish built systems as infrastructure.

In this study, I also bring in the lens of institutional theory. This body of literature provides a complementary set of terminologies and concepts that can shed light on the connections between technological systems and human processes of organization and coordination.

Institutions are generally understood to be “complex social forms that reproduce themselves, such as governments, family, human languages, universities, hospitals, business corporations, and legal systems” (Miller 2019, n. p.). They manifest as stable patterns of individual and organizational behavior that structure and legitimize actions, relationships, and understandings within specific situations. In Douglass

North's metaphor—institutions are “rules of the game” within social interactions, while individuals and organizations are the “players in the game” (1990). Institutions can be understood to be social structures, orders, or patterns that enable cooperation across formal organizations or where formal organizations are absent. I use the phrase “institutional work” to refer to the work involved in establishing stable processes and practices for coordination of heterogeneous stakeholders (Mayernik 2016).

Bruno Latour, in his book “An Inquiry into Modes of Existence” (2013), notes the connection between the trustworthiness of science and the robustness of its institutions. In particular, he draws attention “to the institutions that would allow [truths] to maintain themselves in existence a little longer (and it is here, as we have already seen, that the notion of *trust* in institutions comes to the fore)” (ibid., 18-19, italics in original). Latour discusses how the validity of “truths” is buttressed by particular configurations of practices, values, and institutions. In another example, Harry Collins depicts the organizational work of gravity wave physicists as being central to the operation of their research agendas:

One thing I have discovered about physicists, or at least this group of physicists, is that they love to try to solve problems by inventing organizational structures. I have often been surprised that, when I have asked a question of a senior member of the collaboration about what the members are thinking about this or that conundrum of analysis or judgment, the reply refers to the committees or bureaucratic units they are putting together to deal with it. It is as though a properly designed organization can serve the same purpose as a properly designed experiment—to produce a correct answer (Collins 2013, 81).

My analysis is rooted in the coupling of these two conceptual frameworks. As Paul Edwards notes—infrastructures enable people to “generate, share, and maintain specific knowledge about the human and natural worlds” (2010, 17). Infrastructures shape the kinds of entities that can exist to play roles in knowledge generation, sharing, and maintenance (Edwards et al. 2007). Institutions, on the other hand, provide “vehicles through which the validity of new knowledge can be accredited” (Jasanoff 2004, 39-40). Institutional work is also central to mediate information and knowledge exchanges on the science-policy interface (Miller 2001). As described by Oran Young, Paul Berkman, and Alexander Vylegzhanin in a discussion of governance of environmental systems, considering infrastructures and institutions as coupled phenomena exposes “the relationship between the design and establishment of institutions that form the core of governance systems on the one hand and the administration of these arrangements on a day-to-day basis on the other” (2020, 348). The following discussion of MIPs investigates how infrastructures and institutions support distributed scientific collaboration and data sharing via a coupling of technical systems, research coordination mechanisms, governance structures, and rules of participation.

### **Model Intercomparison Projects (MIPs)**

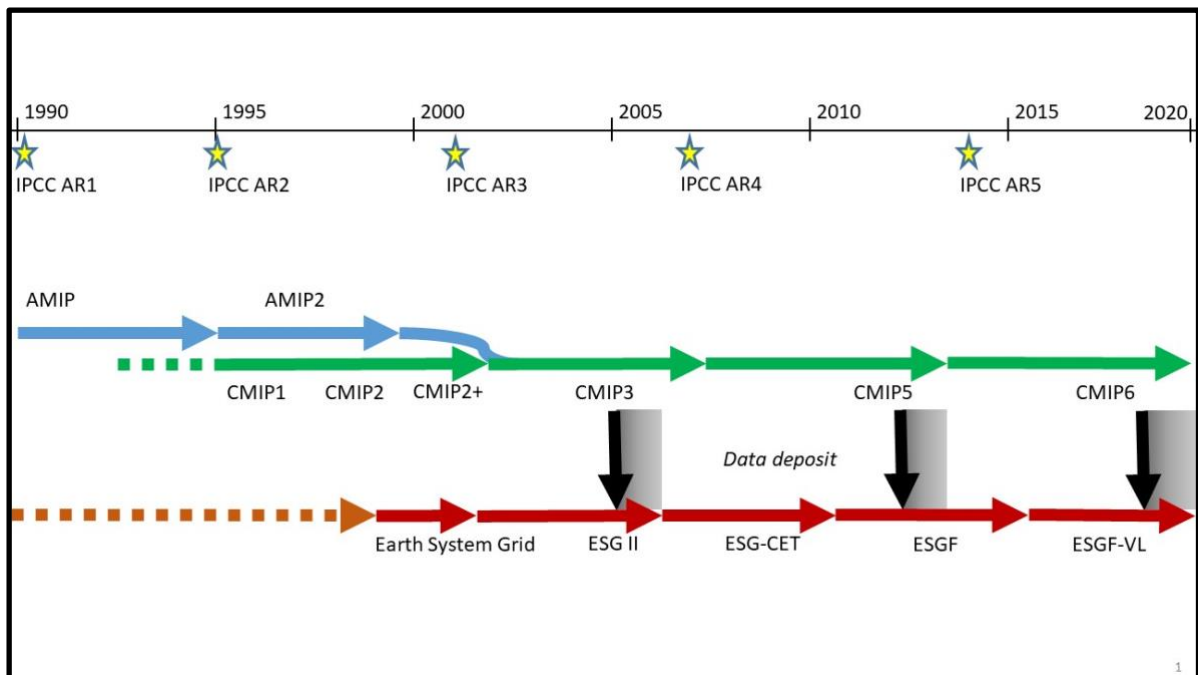
MIP data and results have been used in thousands of climate research papers, largely focusing on exploring future climate change and associated uncertainties, the comparing of MIP simulation results with observations, and informing the interpretation of model results (Touzé-Peiffer, Barberousse, & Le Treut 2020). The first named MIP was the Atmospheric Model Intercomparison Project (AMIP), which began in 1990 under the auspices of the Joint Scientific Committee (JSC) of the World Climate Research Programme (WCRP). WCRP was itself formed in 1980 via joint sponsorship of the International Council for Science (ICSU)



and the World Meteorological Organization (WMO), with the goal of fostering research in support of climate prediction and the influence of human activities on climate.

AMIP was motivated by several smaller scale climate model comparison studies that took place in the 1970s and 1980s ([Gates 1979](#); [Cess et al. 1989](#)). These early projects were focused on evaluating climate models' behavior with respect to specific phenomena, such as clouds, precipitation, or air temperature. They demonstrated that comparing simulations from different models helped to identify where they agree and diverge. AMIP formalized a process for doing intercomparisons. This included defining standard model experiments, often called "model scenarios," such as to simulate climate impacts of a 1% annual increase in the atmospheric CO<sub>2</sub> concentration. Also standardized within AMIP were the requested model output data, the data used to initialize, compare, and validate model outputs, and the model validation procedures themselves ([Gates 1992](#)). The first AMIP set the stage for subsequent AMIPs, as well as the Coupled Model Intercomparison Project (CMIP), which was the next major international MIP. The first iteration of CMIP was initiated in 1996, and focused on characterizing systematic simulation errors of global coupled climate models ([Meehl et al. 1997](#)). Since the beginning the CMIPs have been organized by the Working Group on Coupled Modeling (WGCM), which has operated under the joint auspices of the WCRP and the Climate Variability and Predictability (CLIVAR) organization. The CMIP operations have been managed by a "CMIP Panel," which is constituted within the WGCM. The CMIP Panel is responsible for overseeing the design of the CMIP experiments, the input and output datasets, and for resolving problems that arise.

Many iterations of CMIP and other MIPs have taken place since the turn of the century. [Figure 1](#) shows a timeline of the iterations of AMIP and CMIP since 1990. All of these MIPs themselves consisted of numerous sub-projects and/or sub-MIPs. As of 2021, CMIP6 is currently in process. The results from model intercomparisons have become central to policy discussions surrounding global environmental change, especially the scientific "Assessment Reports" of the Intergovernmental Panel on Climate Change ([IPCC 2021](#)), which are shown in [figure 1](#) as AR1–AR5. The IPCC is itself a complex and fascinating endeavor ([Hulme & Mahony 2010](#)), but for the sake of space is only discussed in relation to MIPs. IPCC AR1, published in 1990, makes no mention of any "MIPs," but includes discussion of precursor intercomparisons. By IPCC AR2, published in 1995, AMIP is discussed extensively. For CMIP3 through CMIP6, feeding into the IPCC assessment reports has been an explicit goal. All modeling groups that contributed to the fifth (published in 2013) and sixth IPCC reports (in process) have been required to perform at least the base set of CMIP experiments. One type of modeling experiment conducted in multiple CMIPs, and featured in the IPCC Assessment Reports, is to compare model simulations of the twentieth century climate both with and without anthropogenic greenhouse gas and aerosol (particulate matter in the atmosphere) increases. These experiments demonstrate that no models can reproduce the warming trend of the twentieth century without anthropogenic influence included. As the iterations of CMIP have progressed, they have included more model experiments, oriented to both past and future climates. For example, 23 different MIPs have been organized for the current CMIP6 ([WCRP 2014](#)).



**Figure 1.** Timeline of activities related to (top to bottom) the IPCC Assessment Reports, AMIP, AMIP2, and CMIP1–CMIP6, and the Earth System Grid (ESG) and Earth System Grid Federation (ESGF) data infrastructures (Source Author’s own).

The data components of AMIP and CMIP have been supported since the beginning by the Program for Climate Model Diagnosis and Intercomparison (PCMDI), which is managed by the Lawrence Livermore National Laboratory, a U.S. Department of Energy (DoE) facility. During AMIP1–2 and CMIP1–2, PCMDI staff performed the bulk of the work to gather, compile, and distribute the input and output datasets. As described in a technical report, the PCMDI role included “quality control and archiving of model results, the development of diagnostic, statistical and visualization software, the assembly of observational data, the maintenance of model documentation, and . . . participation in and overall coordination of model diagnosis, validation and intercomparison” (Gates 1995, 8). In later MIPs, PCMDI has been a leading player in the creation of large-scale infrastructures, specifically the Earth System Grid (ESG) and Earth System Grid Federation (ESGF), which have evolved over multiple iterations (shown in figure 1) to involve dozens of international partners and global-scale computational and data systems.

Thus, MIPs have been targeted at establishing the credibility for climate models within climate science research and policy-relevant science products, such as the IPCC reports (Rood 2019). The success of AMIP and CMIP helped to spawn other MIPs (Gates et al. 1999), thus providing good case studies in how “reproducibility” and its associated concepts are achieved in such coordinated projects.

### Why Participate in MIPs?

Participating in MIPs, particularly the more recent generations of CMIP, puts significant amounts of work on the modeling centers. The required model runs for each generation of CMIP typically take months or years to complete on supercomputers, and the data requirements are significant, both in terms of data volume and standardization, as described further in the next section. Here, however, I address the basic question: *why*



*do modeling centers participate?* The motivations for participating range from scientific to social in nature, as encapsulated by the following interview quote by an individual who was involved in organizing the CMIP1–2:

The benefit. . . , and I think it became quickly apparent, is “how well is our model doing comparing with the others?” And if only because the financial sponsors, say in the US, the NSF, and whoever else is sponsoring these models, the managers in Washington would say, “Well, okay. How good is that model compared to others? Is it really state of the art?” And so to be able to give an objective answer to that question, and also in a more practical way to see where model weaknesses lie (Interview A, 2017).

The analysis of MIP model runs thus gives modelers a window into how their model compares to others. This understanding can be instrumental in improving the models themselves, as well as help with the practical need to report to funders. This social aspect of participating in MIPs also manifests via peer pressure, as noted in the following quote from the same individual:

I think it was as simple as once we got one or two [to participate], then the others didn’t want to be left out. “Well, gee whiz. [Organization A & B] are doing this. Shouldn’t we be doing this? We don’t even look like we’re a legitimate model if we’re not up there with the big boys.”

This quote was made about the early CMIPs, but it holds true in later iterations. For example, a collaboration in Brazil has been working since 2008 to develop the Brazilian Earth System Model (BESM). The BESM was not a participant in CMIP5 (completed in 2013) or any earlier CMIPs, but the BESM team used the CMIP modeling scenarios and data requirements as benchmarks and standards when developing and evaluating the BESM output: “. . . we have followed the criteria for participation in phase 5 of the Coupled Model Intercomparison Project (CMIP5) protocol. . . . The atmospheric data were output at a 3-hourly frequency and later processed using the Climate Model Output Re-writer version 2 (CMOR2) software . . . to satisfy all CMIP5 output requirements” (Nobre et al. 2013, 6717, and 6719). A more recent paper explicitly states that “One of the fundamental aims of the BESM project is to participate in the Coupled Model Intercomparison Project’s sixth phase” (Veiga et al. 2019, 1613).

Likewise, a 2015 paper describing the development of a climate model in India, the Indian Institute of Tropical Meteorology Earth System Model (IITM-ESM), explicitly points to the goal of contributing to the IPCC assessment reports: “The model, a successful result of Indo–U.S. collaboration, will contribute to the IPCC’s Sixth Assessment Report (AR6) simulations, a first for India” (Swapna et al. 2015, n.p. from the abstract). The results presented in the paper also feature multiple comparisons between the outcomes of the Indian model to those demonstrated by the CMIP5 models, even though the IITM-ESM was not a participant in CMIP5.

Newly developed climate models are not only targeted toward participating in international projects like MIPs, however. The goals of the IITM-ESM model are also explicitly targeted toward improving the representation of climate phenomena that impact India, such as monsoons. But the CMIP activities have clearly provided important benchmarks for the Brazilian and Indian modeling efforts. Notably, as of April 2021, the IITM-ESM is listed on the CMIP6 directory of data contributions, but the BESM is not (PCMDI, 2021).

The MIP modeling scenarios also provide important benchmarks for long-established models. One participant in many MIPs described how one modeling center used a standard AMIP model run as a way to evaluate new versions of their model.

It turns out the [modeling center] people, they still do an AMIP run every time they change the model, they run an AMIP run, a 10 or 20-year AMIP run, just to see where they stand in the fixed sea surface temperature thing. Make sure that the ocean's not running it all over. So it's a standard, you know (Interview B, 2017).

The AMIP protocol thus provides a well-understood and relatively simple simulation scenario that can be used to evaluate whether newly added changes to the model would result in unexpected simulation outcomes. Thus, the MIPs serve both scientific and social purposes for those modeling centers that participate.

### **Institution And Infrastructure Coupling (IIC) Within MIPs**

To illustrate IIC in relation to the use of MIPs within climate sciences, I start with the following quote from Meehl et al. (2007) about the compiled CMIP3 dataset: "This unique and valuable multimodel dataset will be maintained at PCMDI and overseen by the WGCM Climate Simulation Panel for at least the next several years." (ibid., 1393). This quote notes two key components of the MIP projects that will be featured in this section, namely, the data infrastructures and the institutionalized governance structures. These components and their coupling have evolved iteratively through the past few decades.

The WGCM panels have been responsible for the MIP experimental designs and overall project organization. A range of institutional structures and processes have been created over time related to governance and collaboration. This institutional work has involved establishing sets of rules related to participation and roles necessary to support the projects' goals. As an example, since the initial AMIP there has been the idea of having a core set of experiments that all participants contribute to, along with a set of focused investigations into specific phenomena. In the first phases of AMIP and CMIP, the specific focused investigations were called "diagnostic subprojects" (Gates 1992). In CMIP6, the most recent MIP, these have been called "endorsed MIPs" that complement the core set of experiments. Each "endorsed MIP" was required to submit a formal proposal that detailed the scientific goals of the effort, the model output to be generated, and any unique data/metadata characteristics, and the designated co-chairs and Scientific Steering Committee (WCRP 2014).

The institutional work done by the CMIP Panel and associated committees also encompasses setting requirements for data and metadata. This has involved forming recommendations and requirements for the datasets used as input for the modeling experiments, and for the output data to be generated by the modeling groups for submission to the central CMIP data collection. These requirements also specify data and metadata file formats, data gridding and coordinate systems, data file organizing scheme, variable names, units, and sign conventions, and more recently, the structure and content of the citation for the output data. For example, the CMIP5 data request included over 400 variables, spanning the atmosphere, oceans, land surface and many other phenomena. The full specification of the CMIP5 model output request runs 133 pages of tables (Taylor 2013).

The infrastructural work involved in the operation of these MIPs has increased in scope over time, being both enabling and constraining on the ambitions of the projects. In the initial AMIPs and CMIP1-2, very little “infrastructure” existed per se, beyond the internet as it existed in the 1990s. Participating modeling teams were asked to send their model output data directly to PCMDI via email or file transfer protocol (FTP). In CMIP3, as the scientific ambitions grew, so did the size of the requested data, along with the concomitant requirements for data infrastructure. As depicted in [figure 1](#), the ESG system was first used to support data collection and distribution for CMIP3. The large data volumes for CMIP3, however, precluded the data from being sent over the internet. Thus, the CMIP3 modeling groups “were sent hard disks and asked to copy their model data onto the disks in netCDF format and then mail the disks to PCMDI where the model data were downloaded and cataloged” ([Meehl et al. 2007, 1385](#)). PCMDI staff then performed the work to load the data onto the ESG for wider distribution. Starting with CMIP5, participating groups were asked to stage their data directly to the ESGF, through a distributed system of fifteen ESGF nodes operated by organizations on four continents ([Cinquini et al. 2014](#)). This was a major transition that was characterized as a “nightmare” by one interviewee (Interview C, 2017), due to the heavy demands put on the modeling centers, who were required to install and operate new large-scale software infrastructures for data publication, transfer, replication, authorization, and quality control.

The infrastructural work in recent MIPs has been aimed toward creating and operating global data collection, management, and distribution systems. In the earlier MIPs, where the technical ambitions for the data infrastructure were lower, significant amounts of work went into producing and distributing the input datasets, and compiling the output from the modeling teams in a common format. Over time, additional infrastructural work has taken place to create a framework for documenting models, data, and their provenance, assign persistent identifiers to data, secure data quality and integrity, enable data discovery and access, data replication, and provide software to visualize and analyze data. For CMIP3, for example, PCMDI staff wrote a software program called Climate Model Output Rewriter (CMOR) that transformed native model output into data files that met the CMIP data request. The stated goal of CMOR was “simply to reduce the effort required to prepare and manage MIP data” ([Taylor, Doutriaux, & Peterschmitt 2006, 1](#)). Later, a new data quality control approach was developed for CMIP5 that formalized three “levels” of data. Within each level, a series of quality control processes and tools were invoked to check for data completeness and conformance to the stipulated standards, to establish version tracking and integrity checking across data replications, and to create persistent identifiers for the data ([Stockhause et al. 2012](#)).

The coupling of the institutional and infrastructural work described here is necessary for projects like MIPs to move forward amid the complex configurations of human, organizational, and technical factors at play. As one example, the metadata requirements evolved significantly from CMIP3 to CMIP5, becoming much more labor-intensive in response to concerns coming out of CMIP3 about the version control and provenance tracking for the model components ([Guilyardi et al. 2011](#)). One individual who was involved in the technical development of the ESG during CMIP3 indicated that these problems were anticipated, but ultimately not addressed by earlier versions of the ESG:

Within the [ESG-CET] project, when it was conceived and proposed and even reviewed, there were things that we knew were gonna be absolutely essential and that if we didn't build them in, they were going to be really hard to tack on. So one was provenance, the other was a robust handling of semantics, and both

those basically got taken off the table. I believe, I don't have this directly from a DOE [Department of Energy] program manager, but we were told, DOE told us not to focus on that (Interview D, 2017).

Retrofitting the collection of metadata and provenance information into the CMIP and ESG workflows was indeed a challenging process. In my interactions with MIP participants, I have heard multiple accounts about how documenting the models and resulting output data via the process used for CMIP5 typically took multiple weeks. CMIP6 metadata requirements and infrastructural components were largely the same as for CMIP5, enabling the modeling centers to carry over lessons learned, and tools built from CMIP5.

This metadata example is one illustration of how the coupling between infrastructural and institutional components of scientific work can reveal frictions. As Ville Aula (2019) notes, infrastructural and data frictions are often addressed via institutional means, whether through formal regulation or informal negotiations among stakeholders. The creation of the ESGF, around the time of CMIP5, was a lengthy and occasionally contentious process that involved negotiations among the prior ESG collaborators, with the WCRP as the orchestrating body. The initial ESG collaboration involved partners funded to investigate a specific problem—how to do high-speed transfers of large-scale data across the internet—while operating as a somewhat informal collaboration. In contrast, a new formal governance structure was set up for ESGF to establish it as an open consortium of organizations, with official documents that stipulate the roles and responsibilities of an ESGF steering committee, its executive committee, and working teams.

In another illustration of IIC, the approach to solving data infrastructure problems that cropped up in CMIP5 was the creation of the WGCM Infrastructure Panel (WIP) for CMIP6 (Balaji et al. 2018). The ESGF infrastructures were noted in a 2019 WGCM document as being critical to the operation of CMIP, while also being “fragile” and a “single point of failure” for the enterprise (WCRP 2019). A prominent illustration of this was when the ESGF was brought down by a security breach in 2015. No data were corrupted, but the ESGF was completely offline for about six months as reengineering took place. According to multiple participants, during that time CMIP5 data users either hit a roadblock or had to find back channel sources. The terms of reference for the new WIP committee included eight clauses, touching on both infrastructural and institutional responsibilities, as demonstrated by the below excerpt:

1. Serve the interests of the WGCM in establishing and maintaining standards and policies for sharing climate model output and derived products . . .
4. Review and provide guidance on requirements of the infrastructure (e.g. level of service, accessibility, level of security) . . .
6. Collaborate with and rely on the ideas and leadership of other groups with interests in standards and infrastructure for climate data (e.g., CMIP, obs4MIPs, CORDEX, ESGF, ES-DOC, CF conventions), with the understanding that the WGCM expects the WIP to provide oversight (WIP 2014).

The institutional work of the WIP was thus focused on mitigating the sources of risk that had cropped up over time as the ESG and ESGF infrastructures became indispensable to the CMIPs.

IIC is thus critical to understanding how MIP-based research achieves the goals of the recent “reproducibility” movements, even if it is not possible and/or practical to bit-for-bit reproduce the outputs of the climate models involved. The rules of participation, experimental specifications, metadata and data standards, and data delivery systems are intertwined in ensuring the comparability of the MIP-generated data and results. The credibility of the data and findings likewise emerge from the articulations between the

experimental design, metadata and provenance tracing, the assignment of persistent identifiers to data, and the security features implemented within the data management and preservation systems. The credibility of organizations and processes involved are both critical to ensuring the trustworthiness of the data and the scientific findings from MIPs. This is nicely encapsulated by this quote from a contributing author to the IPCC Report AR4, which was in fact included in the AR4:

It proved important in carrying out the various MIPs to standardize the model forcing parameters and the model output so that file formats, variable names, units, etc., are easily recognized by data users. The fact that the model results were stored separately and independently of the modeling centers, and that the analysis of the model output was performed mainly by research groups independent of the modelers, has added to confidence in the results. AMIP and CMIP opened a new era for climate modeling, setting standards of quality control, providing organizational continuity, and ensuring that results are generally reproducible ([Somerville 2011, 244](#)).

Note how many different details are mentioned with regard to “ensuring that results are generally reproducible”: standardizing parameters and model output files, the separation between modeling centers and the model output, standards for quality control, and organizational continuity. Coordinating and achieving these many details at international scales requires both institutional and infrastructural work.

### Science Case Study—Clouds And Aerosols

In the past few years, participants in CMIP6 have found that their new generation of climate models are producing future climates with higher “equilibrium climate sensitivity” (ECS) than prior models. The ECS is an important statistic for climate models because it indicates how much the planet may warm due to a doubling of atmospheric CO<sub>2</sub> compared to pre-industrial levels. This section provides insight into how the credibility of a finding like this potentially higher ECS is established.

For the climate modeling group based at the National Center for Atmospheric Research (NCAR), this higher ECS was produced by a new version of a model. Developing this new model involved running nearly 300 different model runs, with varying configurations and inputs. Many of the developmental runs of this model produced surface temperature trends that did not track with the observed twentieth century temperature increase. By iteratively running the model, the team diagnosed that the model produced different climate trends when using aerosol emissions input datasets that were produced for the CMIP6 project, versus emissions datasets produced for CMIP5. (Both sets of emissions data were distributed through the ESGF.) Further diagnosis indicated that cloud production components of the model were the primary cause of the output changes, as cloud generation is tied to the presence of aerosols within the atmosphere.

Accurately simulating feedbacks related to clouds and aerosols have been long-term challenges for climate modelers ([Cess et al. 1989](#)), so this finding was not itself surprising, but the modeling team then faced questions about whether the preliminary simulations of the twentieth century were inaccurate due to problems with the model, the aerosol emissions input datasets, or both. After several iterations, the team found a model configuration that simulated a temperature trend close to the observations when using CMIP5 emissions input data, but not with the CMIP6 emissions. During a 2017 presentation about this investigation, one scientist observed that questions about the input data required a broader discussion within the CMIP6

project, and that data questions needed to be sorted out with the CMIP governance teams before the group could start doing their CMIP6 model runs.

The emissions data had already been re-released a few times to correct problems, once to add back in data that had been dropped due to a “limitation in the ESGF” (Hoesly et al. 2018), and once to correct gridding errors discovered by users (including by the NCAR modeling group). But no further problems were found in the data that resulted in any new releases of the CMIP6 emissions input datasets. Thus, the NCAR modelers continued to evolve their model using these particular datasets as input. Iterative model development continued until their new model was simulating the twentieth century climate trends acceptably while using the later releases of the CMIP6 input emissions forcing data. At this point, the team began analyzing the simulated climates in more detail, including analyzing the new model’s ECS. Subsequent analyses showed that the higher ECS was due to cloud processes. The group did not, however, adjust the model directly to affect the ECS.

As noted in the following quote from a publication in which the higher ECS is presented and analyzed in detail, the importance of the overall finding of a potentially increased ECS required coordinated study:

An ECS [equilibrium climate sensitivity] of 5.3 K would lead to a high level of climate change and large impacts. It is imperative that the community work in a multimodel context to understand how plausible such a high ECS is. What scares us is not that [our] ECS is wrong (all models are wrong, [Box 1976]) but that it might be right (Gettelman et al. 2019, 8336).

This quote closes the cited paper. Note the reference to statistician George Box and his famous statement that “all models are wrong, but some are useful” (Box 1976). Also note the explicit call to study this finding in a “multimodel context.” Subsequent studies of CMIP6 models show that many of them are producing higher ECS numbers than were produced for any previous iteration of CMIP (Zelinka et al. 2020). Thus, the credibility of any one model is tied to the coordinated effort to standardize the inputs, outputs, and evaluation across many models.

### Generalization—Semi-Controlled Experiments

I have so far argued that MIP-based research could be characterized as a “semi-controlled experiment,” following Leonelli’s (2018) typology of reproducibility. This aligns with previous work that has noted that MIP-based climate science is more akin to laboratory experimentation (Schmidt & Sherwood 2014) or a “panel of experts” (Jebeile & Crucifix 2020) than to a repeatable theoretical calculation. I suggest that IIC can be a framework for analyzing research efforts that fit into the “semi-controlled experiment” category. I use three non-climate science examples to illustrate how the IIC lens is useful in cases where the data collection/generation methods are themselves sources of uncertainty and variability. These cases may have less political and social scrutiny as climate research. But they are all subject to external scrutiny of some kind, and demonstrate interconnections between the development of infrastructure and associated institutions.

*Model Organisms:* Leonelli (2018) points to model organisms-based biological research as one example of “semi-controlled experiments.” For model organism-based research, particular strains of organisms, such as fruit flies, mice, or weeds, are used to develop knowledge that is generalizable to other



organisms ([Ankeny & Leonelli 2020](#)). These organisms are “semi-controlled” because, while engineered to conform to behavior and genetic standards, they exist as living entities, and thus inevitably present variations and uncontrollable factors. Model organism-based research is scaffolded by data compilation and distribution infrastructures, as well as committees, consortia, and other organizations that exist to manage and steer the communities of researchers involved ([Leonelli 2019](#)). The credibility of research using model organisms emerges from the comparison and compilation of data derived from the same strain of organism, not the reproducibility of studies that use individual animals of the model species. Model organism-based data differs in important ways from climate data, in particular in how data represent particular entities, such as genes, chromosomes, and species. Climate science data infrastructures, on the other hand, are structured around geo-located time-series of variables such as temperature, wind speed, and precipitation rates ([Lloyd et al. forthcoming](#)). But as Ankeny and Leonelli ([2020, 63](#)) noted, external scrutiny from funding agencies and concerns from within the biomedical communities required concerted political and social effort to overcome: model organisms “would not have been widely adopted without various forms of institutional legitimization, explication of shared conceptual commitments, and technological developments” ([ibid., 63](#)).

*Text Retrieval:* The Text Retrieval Conferences (TREC) have been held every year since 1992 as a venue to advance the field of information and text retrieval. The focus of TREC has been to establish a coordinated mechanism for conducting, comparing, and evaluating text retrieval experiments, with the goal of improving document search algorithms ([Harman & Voorhees 2007](#)). Similar to MIPs, the TREC series is organized around common input datasets, experiment specifications, and evaluation methods, and the explicit goal is to conduct comparative assessments of particular text retrieval algorithms. Also similar to MIPs, the TREC activities have been sub-divided according to particular research questions, such as a “routing” task, in which participating research teams were asked to search specified collections of documents, such as news clipping services, to find relevant documents for a particular subject query. The “semi-controlled” aspect of the TREC projects is the evaluation process, which is based on human judgements of the relevance of documents for particular information retrieval tasks. The inconsistency of human relevance assessments is a systematic challenge in evaluating the effectiveness of information retrieval algorithms ([Harter 1996](#)). TREC addressed this factor by standardizing the relevance judgements used to evaluate the outcomes of the various TREC experimental tracks. The TREC series has been supported since the beginning by the US National Institute of Standards and Technology (NIST) and the US Defense Advanced Research Projects Agency (DARPA). NIST has been the primary site for the “infrastructural work” within TREC, in this case, compiling, standardizing, and distributing the document collections used for each TREC. A TREC Program Committee oversees the meeting programs, including the definition of participation roles, the organization of the experimental tracks, and the specification of their test corpuses.

*Ecological Observatories:* The last example of a “semi-controlled” scientific endeavor discussed here is ecological observatories like the Long Term Ecological Network (LTER) and the National Ecological Observatory Network (NEON), where multiple ecological sites are operated in coordination to support data collection and scientific projects that span time and space. Significant effort goes into achieving consistent, comparable, and traceable datasets from field sites that are “semi controllable” due to unpredictable flora, fauna, and weather. Staff of such observatories use processes that are designed to enable the comparability of data over time ([Ribes & Jackson 2013](#)). These observatories are complex to build, operate, and sustain.

Henry Loescher, Eugene Kelly, and Russ Lea (2017), writing as then-members of NEON leadership about NEON's construction, discussed the coupling of institutional and infrastructural issues explicitly:

[S]taff scientists are faced with an unfamiliar organizational structure of NEON that acts as a construction company, scientific institution, and a start-up company combined, each with its own culture, that often manifest in needs to build internal organizational function/structures for one culture. Compounding this dynamic, are the rapidly changing institutional needs, and the changing and unforeseen reporting and oversight of the sponsors themselves. The need to engage with the user community has never been greater and, at the same time, always outweighs the institutional capability to do so. This is not meant as an excuse, but rather a common, reoccurring reality seen by all research infrastructure during their construction ([ibid., 44](#)).

These challenges resulted in a shake-up in 2016, when the NSF abruptly changed NEON's management, with the goal of speeding up the construction. This changeover did result in the successful completion of the NEON data collection infrastructure, but also involved removal of key scientists and dissolution of scientific advisory committees. This led to questions about the observatory within the ecological science community. One prominent ecologist was quoted in Science magazine as posting on Twitter about NEON as follows: "Great data, no users, no trust = failure" ([Mervis 2019, 212](#)). A NEON Science, Technology & Education Advisory Committee has since been reconstituted, along with 25 other "technical committees" that provide NEON guidance across a range of topics, including community engagement, data standards, and soil sensors.

## Conclusion

This paper argues that the credibility of scientific findings and data for projects that are focused on enabling comparison, such as MIPs in climate science, is strongly related to the coupling of robust infrastructures for data collection, curation, distribution, and preservation, with robust institutions that facilitate the delineation of roles, responsibilities, rules of participation, and coordination processes. The cases depicted in this paper suggest that trust and credibility of the data and findings from such projects comes not from either infrastructures or institutions in isolation, but rather the articulations between them. The concept of "reproducibility" is nuanced for these kinds of projects, due to the semi-controlled nature of the research. In the case of MIP-based climate science, critical aspects of the investigation are beyond the full control of the research community: the complexity of the climate system, the non-deterministic mathematics used to simulate that system, and the portability of the computational manifestations of those mathematics.

Institutional and infrastructural coupling (IIC) is introduced as a framework for understanding how scientific endeavors achieve other things that are more applicable than "reproducibility," namely trustworthiness, credibility, and comparability of the data and findings from such projects. The cases depicted in this paper demonstrate how MIPs and other kinds of scientific endeavors can enable the generation of reliable scientific findings and data even when strict notions of "reproducibility" are not useful. MIPs enable people to perform specific kinds of scientific investigations through collaboration around common research methods, data, and evaluation procedures. The structure of the MIPs in climate science has been relatively robust over time and has transferred from one project to another, with some evolution to both the institutional and infrastructural components, as well as the couplings between them. For example, problems encountered in prior MIPs, including governance challenges, technical problems, and unresolved or emergent scientific questions, tend to re-appear as the remit of a new committee, working

group, or sub-project, and often as the subject of a new rule or recommendation related to participation and data contributions.

The term “coupling” within the IIC concept perhaps implies a one-to-one connection, like between rail cars. I suggest, however, that it is more useful to think of these couplings like those between the bones of the skull, where there are seams of unique and multifaceted connections. One skull bone is only as effective as its coupling to the other bones adjacent to it. Likewise, the infrastructures that support MIPs (and other kinds of research that can be characterized as “semi-controlled experiments” (Leonelli 2018)) are only effective to the extent that they are coupled with institutions that facilitate the credibility of the scientific effort, and vice versa.

### Acknowledgements

I thank the study participants for their time and insight. Thanks to Christine Borgman, Justin Donhauser, Andrew Gettelman, Elisabeth Lloyd, Seth McGinnis, Ryan O’Loughlin, Gary Strand, and the two anonymous reviewers for comments on earlier versions of this work. This material is based upon work supported by the National Center for Atmospheric Research, which is sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. This paper was also supported by the NSF EarthCube program awards 1929757 and 1929773. I also appreciate support from the Association for Information Science and Technology (ASIS&T) Bob Williams Research Grant. Any opinions, findings, and conclusions expressed in this publication are those of the author and do not necessarily reflect the views of NCAR or the NSF.

### Author Biography

Matthew S. Mayernik is a project scientist and research data services specialist in the library of the National Center for Atmospheric Research / University Corporation for Atmospheric Research. He conducts research and leads service development related to scientific data and metadata curation.

### References

- Ankeny, Rachel A., and Sabina Leonelli. 2020. *Model Organisms* Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781108593014>.
- Aula, Ville. 2019. “Institutions, Infrastructures, and Data Friction—Reforming Secondary Use of Health Data in Finland.” *Big Data & Society* 6(2): 205395171987598. <https://doi.org/10.1177/2053951719875980>.
- Balaji, Venkatramani, Karl E. Taylor, Martin Juckes, Bryan N. Lawrence, et al. 2018. “Requirements for a Global Data Infrastructure in Support of CMIP6.” *Geoscientific Model Development* 11(9): 3659–80. <https://doi.org/10.5194/gmd-11-3659-2018>.
- Barua, Maan. 2021. “Infrastructure and Non-Human Life: A Wider Ontology.” *Progress in Human Geography* 45(6): 1467–89. <https://doi.org/10.1177/0309132521991220>.
- Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.
- Box, George E. P. 1976. “Science and Statistics.” *Journal of the American Statistical Association* 71(356): 791–99. <https://doi.org/10.1080/01621459.1976.10480949>.

- Braun, Kathrin, and Cordula Kropp. 2010. "Beyond Speaking Truth? Institutional Responses to Uncertainty in Scientific Governance." *Science, Technology, & Human Values* 35(6): 771–82. <https://doi.org/10.1177/0162243909357916>.
- Bush, Rosemary, Andrea Dutton, Michael Evans, Rich Loft, et al. 2020. "Perspectives on Data Reproducibility and Replicability in Paleoclimate and Climate Science." *Harvard Data Science Review* 2(4). <https://doi.org/10.1162/99608f92.00cd8f85>.
- Cess, Robert D., Potter, Gerald L., Jean-Pierre Blanchet, George J. Boer, et al. 1989. "Interpretation of Cloud-Climate Feedback as Produced by 14 Atmospheric General Circulation Models." *Science* 245(4917): 513–16. <https://doi.org/10.1126/science.245.4917.513>.
- Cinquini, Luca, Daniel Crichton, Chris Mattmann, John Harney, et al. 2014. "The Earth System Grid Federation: An Open Infrastructure for Access to Distributed Geospatial Data." *Future Generation Computer Systems* 36 (July): 400–417. <https://doi.org/10.1016/j.future.2013.07.002>.
- Collins, Harry. 1985. *Changing Order: Replication and Induction in Scientific Practice*. Chicago, IL: University of Chicago Press.
- . 2013. *Gravity's Ghost and Big Dog: Scientific Discovery and Social Analysis in the Twenty-First Century*. Chicago, IL: University of Chicago Press.
- Deser, Clara, Flavio Lehner, Keith B. Rodgers, Toby Ault, et al. 2020. "Insights from Earth System Model Initial-Condition Large Ensembles and Future Prospects." *Nature Climate Change* 10(2020): 277–86. <https://doi.org/10.1038/s41558-020-0731-2>.
- Easterbrook, Steve M. 2014. "Open Code for Open Science?" *Nature Geoscience* 7(2014): 779–81. <https://doi.org/10.1038/ngeo2283>.
- Edwards, Paul N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.
- . 2019a. "Infrastructuration: On Habits, Norms and Routines as Elements of Infrastructure\*." In *Thinking Infrastructures (Research in the Sociology of Organizations)*, edited by Martin Kornberger, Geoffrey C. Bowker, Julia Elyachar, Andrea Mennicken, et al. 62: 355–66. Bingley, Emerald Publishing Limited. <https://doi.org/10.1108/s0733-558x20190000062022>.
- . 2019b. "Knowledge Infrastructures Under Siege: Climate Data as Memory, Truce, and Target." In *Data Politics: Worlds, Subjects, Rights*, edited by Didier Bigo, Engin Isin, Evelyn Ruppert, 21–42. New York: Routledge. <https://doi.org/10.4324/9781315167305-2>.
- , Steven J. Jackson, Geoffrey C. Bowker, & Cory P. Knobel. 2007. *Understanding Infrastructure: Dynamics, Tensions, and Design*. Ann Arbor, MI: University of Michigan. <http://deepblue.lib.umich.edu/handle/2027.42/49353>.
- Feinberg, Melanie, Will Sutherland, Sarah Beth Nelson, Mohammad Hossein Jarrahi, et al. 2020. "The New Reality of Reproducibility: The Role of Data Work in Scientific Research." *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW1): 1–22. <https://doi.org/10.1145/3392840>.
- Galison, Peter. 1987. *How Experiments End*. Chicago, IL: University of Chicago Press.
- Gates, W. Lawrence, ed. 1979. "Report of the JOC Study Conference on Climate Models: Performance, Intercomparison and Sensitivity Studies, Volume I." *Global Atmospheric Research Programme (GARP) Publications Series No. 22*. Geneva: World Meteorological Organization.

- . 1992. “AMIP: The Atmospheric Model Intercomparison Project.” *Bulletin of the American Meteorological Society* 73(12): 1962–70.  
[https://doi.org/10.1175/1520-0477\(1992\)073<1962:atamip>2.0.co;2](https://doi.org/10.1175/1520-0477(1992)073<1962:atamip>2.0.co;2).
- . 1995. “An Overview of AMIP and Preliminary Results.” In *Proceedings of the First International AMIP Scientific Conference. World Climate Research Programme, WCRP-92, WMO/TD-No. 732*, 1–8. Geneva: World Meteorological Organization.  
[https://library.wmo.int/index.php?lvl=notice\\_display&id=11852](https://library.wmo.int/index.php?lvl=notice_display&id=11852).
- , James S. Boyle, Curt Covey, Clyde G. Dease, Charles M. Doutriaux, et al. 1999. “An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I).” *Bulletin of the American Meteorological Society* 80(1): 29–56.  
[https://doi.org/10.1175/1520-0477\(1999\)080<0029:aootro>2.0.co;2](https://doi.org/10.1175/1520-0477(1999)080<0029:aootro>2.0.co;2).
- Gottelman, Andrew, Cecile Hannay, Julio T. Bacmeister, Richard B. Neale, et al. 2019. “High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2).” *Geophysical Research Letters* 46(14): 8329–37. <https://doi.org/10.1029/2019gl083978>.
- Guilyardi, Eric, Venkatramani Balaji, Sarah Callaghan, Cecelia DeLuca, et al. 2011. “The CMIP5 Model and Simulation Documentation: A New Standard for Climate Modelling Metadata.” *CLIVAR Exchanges* 56: 16(2); 42–46. <http://centaur.reading.ac.uk/25733/>.
- Harman, Donna K., and Ellen M. Voorhees. 2007. “TREC: An Overview.” *Annual Review of Information Science and Technology* 40(1): 113–55. <https://doi.org/10.1002/aris.1440400111>.
- Harter, Stephen P. 1996. “Variations in Relevance Assessments and the Measurement of Retrieval Effectiveness.” *Journal of the American Society for Information Science* 47(1): 37–49. [https://doi.org/10.1002/\(sici\)1097-4571\(199601\)47:1<37::aid-asi4>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-4571(199601)47:1<37::aid-asi4>3.0.co;2-3).
- Hoepe, Götz. 2018. “Mediating Environments and Objects as Knowledge Infrastructure.” *Computer Supported Cooperative Work (CSCW)* 28(1–2): 25–59. <https://doi.org/10.1007/s10606-018-9342-0>.
- Hoesly, Rachel M., Steven J. Smith, Leyang Feng, Zbigniew Klimont, et al. 2018. “Historical (1750–2014) Anthropogenic Emissions of Reactive Gases and Aerosols from the Community Emissions Data System (CEDS).” *Geoscientific Model Development* 11(1): 369–408.  
<https://doi.org/10.5194/gmd-11-369-2018>.
- Hulme, Mike, and Martin Mahony. 2010. “Climate Change: What Do We Know about the IPCC?” *Progress in Physical Geography: Earth and Environment* 34(5): 705–18.  
<https://doi.org/10.1177/0309133310373719>.
- Intergovernmental Panel on Climate Change (IPCC). 2021. Reports. <https://www.ipcc.ch/reports/>.
- Jasanoff, Sheila. 2004. “Ordering Knowledge, Ordering Society.” In *States of Knowledge: The Co-Production of Science and Social Order*, edited by Sheila Jasanoff, 13–45. New York: Routledge.
- Jebeile, Julie, and Michel Crucifix. 2020. “Multi-Model Ensembles in Climate Science: Mathematical Structures and Expert Judgements.” *Studies in History and Philosophy of Science Part A* 83(October): 44–52. <https://doi.org/10.1016/j.shpsa.2020.03.001>.
- Jefferson, Gail. 1985. “On the Interactional Unpackaging of a ‘Gloss.’” *Language in Society* 14(4): 435–66.  
<https://doi.org/10.1017/s0047404500011465>.

- Kennefick, Daniel. [2007](#). *Traveling at the Speed of Thought: Einstein and the Quest for Gravitational Waves*. Princeton, NJ: Princeton University Press.
- Latour, Bruno. [2013](#). *An Inquiry into Modes of Existence: An Anthropology of the Moderns*. Trans. Catherine Porter. Cambridge, MA: Harvard University Press.
- Lee, Charlotte P., and Kjeld Schmidt. [2018](#). "A Bridge Too Far? Critical Remarks on the Concept of 'Infrastructure' in Computer-Supported Cooperative Work and Information Systems." In *Socio-Informatics: A Practice-based Perspective on the Design and Use of IT Artifacts*, edited by Volker Wulf, Volkmar Pipek, David Randall, Markus Rohde, et al., 177–217. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198733249.003.0006>.
- Leonelli, Sabina. [2018](#). "Rethinking Reproducibility as a Criterion for Research Quality." In *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise*, Volume 36B, edited by Luca Fiorito, Scott Scheall, & Carlos Eduardo Suprinyak, 129–46. Emerald Publishing Limited. <https://doi.org/10.1108/s0743-41542018000036b009>.
- . [2019](#). "Scientific Agency and Social Scaffolding in Contemporary Data-Intensive Biology." In *Beyond the Meme: Development and Structure in Cultural Evolution*, edited by Alan C. Love and William Wimsatt, 42–63. Minneapolis, MN: University of Minnesota Press.
- Lloyd, Elisabeth A., Greg Lusk, Stuart M. Gluck, and Seth McGinnis. ([forthcoming](#)) "Varieties of Data-Centric Science: Regional Climate Modeling and Model Organism Research." *Philosophy of Science*.
- Loescher, Henry W., Eugene F. Kelly, and Russ Lea. [2017](#). "National Ecological Observatory Network: Beginnings, programmatic and scientific challenges, and ecological forecasting." In *Terrestrial Ecosystem Research Infrastructures*, edited by Abad Chabbi and Henry W. Loescher, 27–51. Boca Raton, FL: CRC Press.
- Mayernik, Matthew S. [2016](#). "Research Data and Metadata Curation as Institutional Issues." *Journal of the Association for Information Science and Technology* 67(4): 973–93. <https://doi.org/10.1002/asi.23425>.
- . [2019](#). "Metadata Accounts: Achieving Data and Evidence in Scientific Research." *Social Studies of Science* 49(5): 732–57. <https://doi.org/10.1177/0306312719863494>.
- Meehl, Gerald A., George J. Boer, Curt Covey, Mojib Latif, et al. [1997](#). "Intercomparison Makes for a Better Climate Model." *Eos, Transactions American Geophysical Union* 78(41): 445–51. <https://doi.org/10.1029/97eo00276>.
- , Curt Covey, Thomas Delworth, Mojib Latif, et al. [2007](#). "THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate Change Research." *Bulletin of the American Meteorological Society* 88(9): 1383–94. <https://doi.org/10.1175/bams-88-9-1383>.
- Mervis, Jeffrey. [2019](#). "Shake-Up Threatens Novel U.S. Ecology Facility." *Science* 363(6424): 211–12. <https://doi.org/10.1126/science.363.6424.211>.
- Miller, Clark. [2001](#). "Hybrid Management: Boundary Organizations, Science Policy, and Environmental Governance in the Climate Regime." *Science, Technology, & Human Values* 26(4): 478–500. <https://doi.org/10.1177/016224390102600405>.
- Miller, Seumas. [2019](#). "Social Institutions." In *Stanford Encyclopedia of Philosophy*. Stanford, CA: Stanford University. <http://plato.stanford.edu/entries/social-institutions/>.



- Miyakawa, Tsuyoshi. 2020. "No Raw Data, No Science: Another Possible Source of the Reproducibility Crisis." *Molecular Brain* 13, 24 (2020). <https://doi.org/10.1186/s13041-020-0552-2>.
- Morrison, Monica A. 2021. "The Models Are Alright: A Theory of the Socio-Epistemic Landscape of Climate Model Development." PhD Diss. Indiana University.
- Moylan, Elizabeth C., and Maria K. Kowalczyk. 2016. "Why Articles Are Retracted: A Retrospective Cross-Sectional Study of Retraction Notices at BioMed Central." *BMJ Open* 6(11): e012047. <https://doi.org/10.1136/bmjopen-2016-012047>.
- [NASEM] National Academies of Sciences, Engineering, and Medicine. 2019. "Reproducibility and Replicability in Science." *Consensus Study Report* Washington, DC: National Academies Press. <https://doi.org/10.17226/25303>.
- Nobre, Paulo, Leo S. P. Siqueira, Roberto A. F. de Almeida, Marta Malagutti, et al. 2013. "Climate Simulation and Change in the Brazilian Climate Model." *Journal of Climate* 26(17): 6716–32. <https://doi.org/10.1175/jcli-d-12-00580.1>.
- North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.
- [PCMDI] Program for Climate Model Diagnosis & Intercomparison. 2021. *ESGF CMIP6 Data Holdings*. Livermore, CA: Lawrence Livermore National Laboratory. [https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf\\_data\\_holdings/](https://pcmdi.llnl.gov/CMIP6/ArchiveStatistics/esgf_data_holdings/).
- Penders, Bart, Holbrook, J. Britt, & Sarah de Rijcke. 2019. "Rinse and Repeat: Understanding the Value of Replication across Different Ways of Knowing." *Publications* 7(3): 52. <https://doi.org/10.3390/publications7030052>.
- Randall, David A., Richard A. Wood, Sandrine Bony, Robert Colman, et al. 2007. "Climate Models and Their Evaluation." In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. New York: Cambridge University Press. <https://www.ipcc.ch/report/ar4/wg1/climate-models-and-their-evaluation/>.
- Ribes, David and Steven J. Jackson. 2013. "Data Bite Man: The Work of Sustaining a Long-Term Study." In "Raw Data" is an Oxymoron, edited by Lisa Gitelman. 147–166. Cambridge, MA: MIT Press.
- Rood, Richard B. 2019. "Validation of Climate Models: An Essential Practice." In *Simulation Foundations, Methods and Applications*, edited by Claus Beisbart and Nicole J. Saam, 737–62. Springer International Publishing. [https://doi.org/10.1007/978-3-319-70766-2\\_30](https://doi.org/10.1007/978-3-319-70766-2_30).
- Schmidt, Gavin A., and Steven Sherwood. 2014. "A Practical Philosophy of Complex Climate Modelling." *European Journal for Philosophy of Science* 5(2): 149–69. <https://doi.org/10.1007/s13194-014-0102-9>.
- Shapin, Steven. 1995. "Cordelia's Love: Credibility and the Social Studies of Science." *Perspectives on Science* 3(3): 255–275. <http://nrs.harvard.edu/urn-3:HUL.InstRepos:3293019>.
- Somerville, Richard C. 2011. "The Co-evolution of Climate Models and the Intergovernmental Panel on Climate Change." In *The Development of Atmospheric General Circulation Models: Complexity, Synthesis, and Computation*, edited by Leo Donner, Wayne Schubert, and Richard C. Somerville, 225–252. New York: Cambridge University Press.

- Star, Susan Leigh, and Karen Ruhleder. 1996. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7(1): 111–34. <https://doi.org/10.1287/isre.7.1.111>.
- Stockhause, Martina, Heinke Höck, Frank Toussaint, and Martin Lautenschlager. 2012. "Quality Assessment Concept of the World Data Center for Climate and its Application to CMIP5 Data." *Geoscientific Model Development* 5(4): 1023–32. <https://doi.org/10.5194/gmd-5-1023-2012>.
- Swapna, P., M. K. Roxy, K. Aparna, K. Kulkarni, et al. 2015. "The IITM Earth System Model: Transformation of a Seasonal Prediction Model to a Long-Term Climate Model." *Bulletin of the American Meteorological Society* 96(8): 1351–67. <https://doi.org/10.1175/bams-d-13-00276.1>.
- Taylor, Karl. 2013. CMIP5 Standard Output. Livermore, CA: Program for Climate Model Diagnosis and Intercomparison (PCMDI). [https://pcmdi.llnl.gov/mips/cmip5/docs/standard\\_output.pdf?id=4](https://pcmdi.llnl.gov/mips/cmip5/docs/standard_output.pdf?id=4).
- , Charles Doutriaux, and Jean-Yves Peterschmitt. 2006. *Climate Model Output Rewriter (CMOR)*. Livermore, CA: Program for Climate Model Diagnosis & Intercomparison (PCMDI). [https://pcmdi.github.io/cmor-site/media/pdf/cmor\\_users\\_guide.pdf](https://pcmdi.github.io/cmor-site/media/pdf/cmor_users_guide.pdf).
- Touzé-Peiffer, Ludovic, Anouk Barberousse, and Hervé Le Treut. 2020. "The Coupled Model Intercomparison Project: History, Uses, and Structural Effects on Climate Research." *WIREs Climate Change* 11(4). <https://doi.org/10.1002/wcc.648>.
- Veiga, Sandro F., Paulo Nobre, Emanuel Giarolla, Vinicius Capistrano, et al. 2019. "The Brazilian Earth System Model Ocean—Atmosphere (BESM-OA) Version 2.5: Evaluation of its CMIP5 Historical Simulation." *Geoscientific Model Development* 12(4): 1613–42. <https://doi.org/10.5194/gmd-12-1613-2019>.
- [WCRP] World Climate Research Program. 2014. *Application for CMIP6—Endorsed MIPs*. World Climate Research Programme. <https://www.wcrp-climate.org/modelling-wgcm-mip-catalogue/modelling-wgcm-cmip6-endorsed-mips>.
- . 2019. *Report of the 22nd Session of the Working Group on Coupled Modeling: 25th and 29th March 2019*, Barcelona, Spain. WCRP Publication 14(2019). World Climate Research Program. <https://www.wcrp-climate.org/WCRP-publications/2019/WCRP-Report-No14-2019-WGCM22.pdf>.
- [WIP] WGCM Infrastructure Panel. 2014. Terms of Reference for the WGCM Infrastructure Panel (WIP). World Climate Research Programme. [https://wcrp-cmip.github.io/WGCM\\_Infrastructure\\_Panel/Papers/WIP\\_Terms\\_of\\_Reference.pdf](https://wcrp-cmip.github.io/WGCM_Infrastructure_Panel/Papers/WIP_Terms_of_Reference.pdf).
- Wilson, Joseph. 2021. "Two Exploratory Uses for General Circulation Models in Climate Science." *Perspectives on Science* 29(4): 493–509. [https://doi.org/10.1162/posc\\_a\\_00380](https://doi.org/10.1162/posc_a_00380).
- Winsberg, Eric. 2018. *Philosophy and Climate Science*. New York: Cambridge University Press.
- Yan, An, Caihong Huang, Jian-Sin Lee, and Carole L. Palmer. 2020. "Cross-Disciplinary Data Practices in Earth System Science: Aligning Services with Reuse and Reproducibility Priorities." *Proceedings of the Association for Information Science and Technology* 57(1): e218. <https://doi.org/10.1002/pr2.218>.
- Young, Oran R., Paul Arthur Berkman, and Alexander N. Vylegzhanin. 2020. "Informed Decisionmaking for the Sustainability of Ecopolitical Regions." In *Informed Decisionmaking for Sustainability*, edited by Oran R. Young, Paul Arthur Berkman, and Alexander N. Vylegzhanin, 341–53. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-25674-6\\_15](https://doi.org/10.1007/978-3-030-25674-6_15).

Zelinka, Mark D., Timothy A. Myers, Daniel T. McCoy, Stephen Po-Chedley, et al. 2020. "Causes of Higher Climate Sensitivity in CMIP6 Models." *Geophysical Research Letters* 47(1). <https://doi.org/10.1029/2019gl085782>.