This contribution was submitted to the National Science Foundation as part of the NSF CI 2030 planning activity through an NSF Request for Information, https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode).

Author Names & Affiliations

- Anke Kamrath - NCAR
- David Hart - NCAR
- BJ Smith - NCAR
- Rich Loft - NCAR

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Supercomputing for Geosciences

Title of Submission

Cyberinfrastructure for the Atmospheric and Related Sciences

Abstract (maximum ~200 words).

The atmospheric sciences community faces two broad and increasingly interdependent challenges: Improving our understanding and prediction of atmospheric, chemical, and space weather hazards and the impact those hazards have on ecosystems, people, and society; and improving our understanding, prediction, and projection of the consequences of natural and anthropogenic climate variability and change at regional and global scales. These challenges are continually reshaped by the broadening realization of their societal significance and are intertwined in their relationship to our ability to observe and understand the complex Sun-Earth system, including forcing from human systems.

Addressing these challenges requires an evolution in cyberinfrastructure to enable the community to manage increasingly complex, massive data sets with advanced analysis tools and innovative best practices. Software infrastructure and application development also must keep pace, focused on the pursuit of a unified modeling approach and new, multi-scale simulation systems. It will also be important to rethink how cyberinfrastructure is planned, supported, and used to ensure its long-term stability for meeting critical science objectives. Finally, the community needs to encourage and support the training and development of the research computing professionals who are needed to sustain the CI enterprise.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.
The National Center for Atmospheric Research has identified two Grand Challenges, which are summarized here and described in the 2014–2019 NCAR Strategic Plan (https://ncar.ucar.edu/sites/default/files/NCAR_Strat_Plan_Final_102014.pdf). The first challenge is to improve our understanding and prediction of atmospheric, chemical, and space weather hazards and the impact they have on ecosystems, people, and society. This includes determining the relevant processes responsible for hazards; investigating the global context of destructive weather and unhealthy air-quality events; working with operational agencies to enhance observational data and models for space-weather forecasting; and strengthening partnerships with researchers in the applied sciences and health communities to determine the societal and ecosystem impact of hazards.

The second is to improve our understanding, prediction, and projection of the consequences of natural and anthropogenic climate variability and change at regional and global scales. This includes understanding the effects of climate change on the water cycle, water availability, weather extremes, and the health and functioning of marine and terrestrial ecosystems; discerning the potential for abrupt changes in climate; and understanding the limits and options that society has for responding to climate change.

These challenges are not new, but they are continually reshaped by the broadening realization of their societal significance. At their heart is our ability to observe and understand the complex Sun-Earth system, including forcing from human systems. This knowledge must be encapsulated in dynamical models that provide increasing skillful capabilities for predictions of weather, climate, atmospheric chemistry, space weather, and associated effects on socio-ecological systems. A potential tension exists, however, between different directions for progress in modeling. Specifically, should resources be allocated and research efforts be extended to increase model resolution, ensemble size, or model complexity? The answer to this question should be problem-driven. Predicting high-impact weather events, such as severe storms and tropical cyclones, will require significant increases in model resolution to capture important convective-scale forcing, as well as the assimilation of new data sources. Stakeholders are demanding estimates of the uncertainty in predictions, which requires ensemble-forecasting techniques. On the other hand, climate applications with great societal relevance, such as climate change impacts on water resources and sea-level rise, require substantial progress to accurately represent the complex interactions of the earth system (e.g., ice-sheet dynamics and land-surface hydrology). Significant progress will be achieved through the advancement of the Weather Research and Forecasting (WRF) Model and the Community Earth System Model (CESM), the continued development and community support of which remain a research and service priority for NCAR.

Regional weather prediction models and global climate models historically have evolved on separate development paths. However, with increasing computing power and advancing modeling technology, weather models are moving toward global coverage while enhanced-resolution climate models are beginning to resolve mesoscale processes. This trend toward multi-scale simulation capabilities emphasizes that the traditional distinctions between “weather” and “climate” models are becoming increasingly blurred, and there is great synergistic benefit in developing unified modeling approaches for both physical processes and model dynamics. The large-scale climate, for instance, determines the environment for microscale and mesoscale processes that govern weather, and these small-scale processes likely have significant impacts on the evolution of the large-scale circulation. Accurately representing this continuum of variability in numerical models is a challenging but essential goal.

Fundamental barriers to advancing weather and climate prediction on time scales from days to decades, as well as long-standing systematic errors in prediction models, are partly attributable to our limited understanding of and capability for simulating the complex, multi-scale interactions that are intrinsic to atmospheric, oceanic, and cryospheric fluid motions. For weather prediction, moreover, there are opportunities to benefit from advances in climate modeling that reduce systematic errors due to mean-state drift, leading to more accurate medium-range weather and air-quality forecasts. Global climate and chemistry modeling can benefit in turn from the experience of weather modelers in designing parameterizations for model physics that more accurately capture mesoscale weather systems, thus enhancing understanding of multi-scale interactions in the coupled system.

NCAR identified additional research challenges related to future cyberinfrastructure needs in planning for the construction of the NCAR-Wyoming Supercomputing Center (NWSC), as described at https://www2.cisl.ucar.edu/user-support/allocations/nwsc-science-impact. Expanding on a series of reports from the Earth Sciences, NCAR justified deployment of Yellowstone, Cheyenne, and future HPC systems, as well as the construction of the NWSC, with an extensive evaluation of the potential science impact of petascale and exascale computing capabilities on the Earth System sciences. The NWSC science justification describes how many research challenges place demands on and push the limits of cyberinfrastructure in six areas or dimensions [p. 14–15]: increased spatial resolution; the inclusion of new processes and phenomena plus the inclusion of more diverse processes; higher-fidelity physics; increased capabilities for data assimilation; increased ensemble sizes for better statistics; and execution over longer timescales. For a particular computational level – and assuming the time to solution is fixed by practical or real-time limitations – advances in any one of these six dimensions must come at the expense of the others. Depending on their specific research interests, different modelers will make different choices in how to expend an increment of available resources along the different axes. The ability to open new scientific frontiers by advancing all areas simultaneously would require both radical advances in current computational capabilities and the development of novel enabling algorithms. The latter include nested and
computing community.

accelerating scientific progress in these areas is encouragement of cross-disciplinary discovery across the research and scientific development of radically new approaches to testing and optimizing software and new model validation strategies. Also important for computers will necessitate new algorithms and revision of existing model designs. Meeting these standards will also require the analytics, and computer science capabilities. Achieving good performance, scalability, and fault tolerance on petascale to exascale Such new multi-scale simulation systems will push modeling scale and complexity beyond the limits of current software engineering, data analytics, and computer science capabilities. Achieving good performance, scalability, and fault tolerance on petascale to exascale computers will necessitate new algorithms and revision of existing model designs. Meeting these standards will also require the development of radically new approaches to testing and optimizing software and new model validation strategies. Also important for accelerating scientific progress in these areas is encouragement of cross-disciplinary discovery across the research and scientific computing community.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

To address the research challenges described in the response to Question 1, cyberinfrastructure must evolve dramatically in the areas of data and software infrastructure while adapting to rapidly evolving computational technology. The 2014–2019 NCAR Strategic Plan describes some of the present cyberinfrastructure limitations and the advancements needed to surmount them, as summarized here. For example, data sets that are impractical to deal with because they are either too large or too complex are a growing problem for the atmospheric and Sun-Earth science community. They are rapidly outstripping the capabilities of current analysis tools and workflow practices and threaten to become a significant barrier to scientific progress. As the scientific community’s capacity to generate data continues to increase, innovation in both data storage and analysis technology will need to be matched by innovation in best practices for data management, workflow, planning, and conducting data-intensive computational campaigns. Developing new capabilities for extracting useful information from large, diverse, distributed, and heterogeneous data sets will be integral to meeting these challenges. It will be critical to conducting novel scientific research, including the development, validation, and application of multi-scale simulation systems. Addressing this challenge will require community collaboration to standardize metadata and promote transparent data discovery, access, analysis, legacy, preservation, and stewardship. Similarly, meeting the OSTP memorandum on public access of research data requires promoting data publication as a first-class scientific activity to support these efforts. Finally, while new data and storage technologies are essential to coping with the big data challenges in the Earth Sciences, the availability of such technologies must be matched by a corresponding investment to procure, deploy, and operate these solutions as key infrastructure components.

Similarly, software infrastructure and application development must keep pace with other advances to take the fullest advantage of computational infrastructure, particularly in the realm of multicore technology. This technology can support the development and use of new multi-scale simulation capabilities as well as the enhancement of existing community models. Existing models should be evaluated for their adaptability to multicore petascale and exascale supercomputing architectures that will enable them to more efficiently address the pressing problems in Sun-Earth system science. These problems include many unanswered scientific questions about the processes and interactions that determine weather, climate, atmospheric chemistry, space weather and their predictability. Projecting changes in weather and climate over the coming decades, for example, depends on understanding the components of the land-atmosphere-ocean-cryosphere-human system and their multi-scale interactions, particularly those that regulate the cycling of water and carbon. New global multi-scale simulation and prediction applications will provide the opportunity to better address important weather-climate interactions that occur across a wide range of scales. Using seamless variations in resolution, for instance, researchers will be able to assess the local impact of global climate trends with less ambiguity, as well as the upscale influences of smaller scale processes. Developing multi-scale simulation capabilities will also enhance the prediction and warning of hazardous weather systems by allowing observational information to be synthesized effectively at all scales by advanced data-assimilation systems.

Pursuing the research priorities identified by NCAR requires the continued pursuit of a more unified modeling approach to support the development of global models that run efficiently and accurately at a variety of horizontal resolutions, and models with limited-area, higher-resolution grids for both regional downscaling and upscaling of information from local events. These next-generation, community-supported modeling systems will also necessitate the development of scalable physical parameterizations and numerical algorithms, along with new techniques for modeling and coupling socio-ecological systems, human activities, and decision-making with the biophysical components of the Earth system.

Such new multi-scale simulation systems will push modeling scale and complexity beyond the limits of current software engineering, data analytics, and computer science capabilities. Achieving good performance, scalability, and fault tolerance on petascale to exascale computers will necessitate new algorithms and revision of existing model designs. Meeting these standards will also require the development of radically new approaches to testing and optimizing software and new model validation strategies. Also important for accelerating scientific progress in these areas is encouragement of cross-disciplinary discovery across the research and scientific computing community.
Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

Today it is widely accepted that computational methods have become the third pillar of the scientific enterprise, alongside experimental and observational methods. However, NSF’s investment in cyberinfrastructure has been, at best, inconsistently treated as an infrastructural pillar. Cyberinfrastructure at the NSF needs to be treated more akin to how NSF handles infrastructure for Antarctic research stations or the FFRDRC model used with NCAR to support research in the atmospheric sciences. Infrastructure implies stability and predictability over the long term, with measured and planned response to change. By contrast, the XD program, a cornerstone of NSF’s cyberinfrastructure investment strategy, supports, awards, and reviews the XSEDE project and various computational investments as research activities within the CISE directorate, rather than as infrastructure activities that in practice serve all NSF directorates. In other words, science objectives and priorities must drive CI strategy and investment, yet most NSF cyberinfrastructure investments are made by peer-review panels managed by CISE, even though the vast majority of scientific use of those investments comes from outside of CISE.

In considering cyberinfrastructure as research infrastructure, NSF must also balance between the competing desires for tomorrow’s better and forward-looking technologies, on the one hand, and the baseline needs for today’s best, capable, and highly demanded technologies on the other. That is, cyberinfrastructure investments should address computing, storage, and networking infrastructure for today’s immediate science needs with today’s best available technologies and also support tomorrow’s next-generation technologies for future science challenges.

Similarly, scope and mission for NSF’s collective cyberinfrastructure investments need to be clarified and subsequent investment strategies should support that scope and mission. In particular, NSF’s cyberinfrastructure investment in the XD program not only supports NSF-funded research, but also computational research by any U.S. researcher, regardless of funding source. NSF should consider whether such an approach will prove sustainable for the long-term. (By contrast, cyberinfrastructure support from DOE and NASA facilities requires alignment with those agencies’ missions, if not an awarded grant from those agencies.) Internally, NSF shows a similar lack of clarity on scope and mission; NSF presentations have touted that NSF’s OCI/ACI/OAC budgets in recent years have “tracked the CISE budget,” even though the vast majority of the use of that infrastructure is by non-CISE NSF directorates.

Related to both the issue of stability and science objectives, CI strategy and investments need to encompass the growing significance, both scientifically and financially, of data access and preservation. Expectations by end users and by cyberinfrastructure providers related to requirements for, availability of, and responsibility for long-term storage and preservation lack clarity. NSF’s long-term CI strategy needs to address how and by whom the cyberinfrastructure for data management, access, and preservation will be delivered and implemented. In this realm, cyberinfrastructure is required to support not only the data generated by computational methods, but also the data produced through experimental and observational methods, further compounding the challenge.

It is unlikely that CISE alone can sustain a national cyberinfrastructure sufficient to support the computational needs of the entire NSF, and similarly for NSF alone to sustain a national cyberinfrastructure to support the needs of all university research supported by federal agencies. Furthermore, many campuses are now making more substantial investments in cyberinfrastructure or pursuing CI resources through NSF Major Research Instrumentation programs (which also tends to treat such infrastructure investments as research activities). However, campuses decide such investments in light of plans from agencies such as NSF, so a better defined, longer-term CI strategy will likely affect how campuses pursue and direct local investments. A confounding factor is that many cyberinfrastructure investments tend to have a lifespan of 4-6 years, compared with decades in the case of other infrastructure. Thus, NSF, with community input, should develop and maintain a multi-decadal roadmap for cyberinfrastructure as research infrastructure against which the agency itself and campuses across the country can plan their investments in technologies not only on a 4-6 year timeframe, but also for the longer term.

Investments in the technological components of cyberinfrastructure must be aligned with appropriate investments in the workforce. A side effect of treating infrastructure investments as research awards is the corresponding broader impacts focus of the resulting awards. The academic and industry providers face a current and future shortage of CI professionals to develop, operate, and maintain local and national research cyberinfrastructure environments. Training and outreach efforts should be encouraged and rewarded for increasing the pipeline of students into STEM, and especially in computer science and IT careers. These activities also should be encouraged with the goal of increasing the pool of professionals needed to sustain the CI enterprise, including retraining of adults for second careers or internship programs to fill skilled technical (but non-research) positions. Efforts to increase gender and minority representation in CI also should be encouraged.

Consent Statement

- “I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode).”