

Cyberinfrastructure for the Atmospheric Sciences in the 21st Century

The CyRDAS Committee – August 2003

The steering committee for Cyberinfrastructure for Research and Development in the Atmospheric Sciences (CyRDAS) is gathering input from the community on the questions of how cyberinfrastructure could lead to more rapid and more substantial progress in research and more efficient and effective education. We are also seeking to determine what cyberinfrastructure barriers are impeding progress. The goal is to identify the central issues that atmospheric scientists, educators and technologists consider most important, from their individual perspectives, to help them achieve what they hope to accomplish. The following list of questions, which is neither comprehensive nor exhaustive, is simply intended to stimulate thinking on these issues.

SOCIAL/CULTURAL ISSUES

- **Work force.** What are the CI-related atmospheric science work force needs of the Nation? Will different sectors (academia, government, private sector) have different needs? Is appropriate CI support best facilitated by the staff in large centers, by a departmental pool of computer scientists/programmers, by providing each PI with the funding to support his or her own computer scientist/programmer, or something else? What changes, if any, should be made to the academic reward structure (tenure) to appropriately recognize young faculty members who are actively involved in scientific computing? In collaborative research? In outreach efforts in education?
- **Students.** How can we ensure that students who intend to pursue careers in atmospheric sciences are aware of and know how to use the appropriate hardware and software tools? How can CI be best leveraged to attract undergraduates and graduate students to the atmospheric sciences? How can CI contribute to guiding pre-college students toward careers in the atmospheric sciences?
- **Collaboration.** How best can computer scientists and atmospheric scientists be encouraged to work together? Are collaborative environments for scientists truly beneficial, and if so, how can lessons be learned from previous experiments? How can CI help overcome the barriers to progress experienced in community modeling activities?
- **Integrating research and education.** Is it reasonable to assume that the CI can stimulate closer relationships between the research and education communities? Is the atmospheric science community prepared to produce the needed materials and resources developing digital library? How can the atmospheric science community use the CI to help prepare, or improve, the science training of the K-12 teaching community?
- **Intellectual property.** How should the intellectual property rights culture in the atmospheric research and education community be structured? What are the potential benefits and pitfalls of managing intellectual property rights in the expanded collaboration space that can be supported with the evolving CI?
- **Access and availability.** To what extent should CI resources be available to PIs, national laboratories' staff, graduate faculty, undergraduate faculty, graduate students, and K-12 educators? What is a reasonable level of cost to ensure universal access to CI?
- **Intra-NSF and inter-agency coordination.** What mechanisms and structures should be put in place to enable and ensure coordination of CI activities among the various divisions of the Geosciences directorate? Between the Computer and Information Sciences and Engineering and the Geosciences directorates? How should CI planning and investments at the NSF be coordinated with similar activities at DOE, NASA, NOAA and other federal agencies funding atmospheric sciences research?

HIGH-END COMPUTING ISSUES

- **Hardware**. Are the high-end computing solutions (architectures, capability, and capacity) that are currently available commercially, and those that may reasonably be anticipated to be available in the near future, meet the needs of atmospheric sciences research? Is special-purpose design and development of computational systems required for atmospheric science research projects that may arise over the next ten years?
- **Modeling requirements**. What are the requirements in the next decade for implementing more accurate, and better spatially resolved (“larger”) problems, and problems with more realistic physical parameters? What are the requirements for longer duration models, and ensembles of model calculations needed to carry out key statistical studies? What are the requirements for predictive computing model results in real time or near real time?
- **Centralized to distributed computing**. What demands will the atmospheric sciences community have for different types of computing facilities, including large shared-memory multiprocessor supercomputers, local Beowulf clusters, and distributed grid computing capabilities? How should resources be distributed among individual PIs, small centers and large centers? Which computing resources should be centralized? distributed?

DATA ISSUES

- **Observational requirements**. What are the requirements of the atmospheric sciences research community for data archiving and dataset organization, management and access for the vast datasets that will be accumulated using multi-point measurements and multi-station observational datasets (weather stations, spacecraft, etc)?
- **Centralized to distributed data management**. What is the proper balance between archiving data at a central location, such as NCAR, and at distributed sites such as observatories or PI institutions? Can anticipated developments in networking meet the needs of globally-distributed, high-data-rate atmospheric science instrumentation? Is there a requirement for and what can be done to make old analog data available in digital form, and to ensure that both old and new digital data are preserved on current media?
- **Metadata**. Are the ongoing efforts in metadata generation by digital library communities sufficient to address the needs of the entire atmospheric sciences community? Is a concerted effort needed to produce metadata for various digital collections? Given the complexities of generating metadata, what barriers do individual PIs foresee in generating metadata for their own datasets? Should future CI investments be made to produce ontologies for atmospheric and related sciences?
- **Data mining**. Are scientific data mining tools being used in current research? What scientific and computational advancements need to be made to enhance the current generation of scientific data mining tools?

SOFTWARE ISSUES

- **Software engineering**. What software practices and design methods are most critical for smaller projects and large efforts in the atmospheric sciences, and how can those skills be taught? When is it appropriate to train individuals within the atmospheric sciences community and when is it appropriate to seek expertise from outside our community? How can software processes help to bridge the gap between the computer scientists and software engineers who develop software and atmospheric scientists who use software?
- **Software project management**. What can we learn from the commercial domain and other fields about the management of software projects? How can we identify and train software project managers? How can we best mentor young software engineers?
- **Modeling frameworks**. What role do frameworks play in enabling new atmospheric science? How can frameworks be designed to accommodate scientific flexibility and technical innovation? How do we balance community-wide framework efforts with smaller, more focused frameworks? How should the current generation of atmospheric

science frameworks evolve over the next 5 years to best serve the modeling community?
Over the next 10 years?

- **Integrating GIS and scientific information systems** . How important are geographic information systems (GIS) tools and services in current and future atmospheric sciences education and research activities? Would there be a benefit and what specific steps are needed to bridge the gap between GIS tools and scientific analysis and display applications? How should atmospheric science students and professionals be trained in the use of GIS? What is the appropriate balance between using proprietary and standards-based GIS tools and services? And how do we attain that balance?