J2.2

THE NATIONAL FORUM FOR GEOSCIENCES INFORMATION TECHNOLOGY ("figit")

[†]Kelvin K. Droegemeier

Center for Analysis and Prediction of Storms and School of Meteorology
University of Oklahoma
Norman, Oklahoma

and

Robert B. Wilhelmson
National Computational Science Alliance and Department of Atmospheric Sciences
Urbana, Illinois

1. MOTIVATION AND PURPOSE

During the past several years, considerable attention has been directed by researchers, funding agencies, lawmakers, policymakers and private companies to the role, provision, and innovation of persistent [1]cyberinfrastructure in the U.S. science and engineering enterprise (e.g., NSF 2003; OSTP 2004). Numerous community workshops have been held to organize views on this topic within and across traditional academic disciplines, and visionary plans abound (e.g., ONR/NSF 2002; Estrin et al. 2003; Oldo et al. 2003; NSF 2003a,b; Walker et al. 2003; NSF 2004). Not surprisingly, most disciplines have espoused common needs (e.g., powerful computers, large storage devices, fast networks, federated data repositories, a range of visualization capabilities, and tight coupling among each).

In addition to funding these resources and establishing administrative frameworks for their governance, perhaps one of the greatest challenges involves applying them in novel ways to attack new problems that require on-demand accessibility, real time response, the steering of remote instruments, and an ability to dynamically adapt to changing circumstances in any part of the problem, including within the cyberinfrastructure itself.

A broad portfolio of information technology research and development now is underway on topics ranging from medicine to weather, particle physics to ecology (e.g., LEAD, ESG, NOMADS, GEON, ESMF, NEESGrid, GriPhyN, VGEE, EarthScope, NVODS, CONDUIT, CRAFT, THREDDS, ESDIS, MADIS, DLESE, GCMD). These activities not only are establishing the clear need for persistent cyberinfrastructure, but also providing a glimpse, through actual implementation, of the many benefits to be wrought by the research and education enterprise as well as society in general.

In spite of these efforts – which are emplacing practical tools for broad community use, engaging numerous agencies and creating entirely new communities – there exists considerable duplication of effort, a continuing lack of standards, notable disconnects between the research and operational communities, and a lack of support for technology deployment, adoption and maintenance. Driven in large part by continuous rapid changes in technology and thus the absence of stable environments in which to design and deploy, the overall effectiveness and impact of cyberinfrastructure activities easily is diminished.

To begin a community dialog about these and other issues in a manner that builds upon the many excellent planning and research activities now underway, the National Science Foundation's Large Information Technology Research (ITR) grant known as "Linked Environments for Atmospheric Discovery (LEAD)" (Droegemeier et al. 2004) is organizing the *National Forum for*

[†] Corresponding Author Address: Prof. Kelvin K. Droegemeier, University of Oklahoma, 100 East Boyd Street, Suite 1110, Norman, OK 73019. Email: kkd@ou.edu

[1]Fran Berman, Director of the San Diego Supercomputing Center, defines cyberinfrastructure as "the integration, coordination, and deployment of information technology and human resources to support modern science and engineering problems." Atkins (2001) notes that "cyberinfrastructure" is to the information age what physical infrastructure was to the industrial age.

Geosciences Information Technology (pronounced "figit"). Working with Federal agencies (principally though not limited to the NSF, NASA, NOAA, USGS, EPA, DOE, DOD, DHS, USDA), professional societies and other agencies, *FGIT will serve as a focal point for national dialog on cyberinfrastructure in the geosciences with particular emphasis on the linking of IT systems and the sharing of knowledge and resources.*

2. CYBERINFRASTRUCTURE CHALLENGES IN THE GEOSCIENCES

The geosciences (geology, and geophysics; atmospheric, environmental, oceanographic, and hydrologic sciences; and space weather) encompass an enormously complex human-natural system that operates over vast temporal (milliseconds to billions of years) and spatial (nanometers to thousands of kilometers) scales and exhibits highly nonlinear behavior. Massive data sets are common, involving both physical (e.g., core samples) and digital information that is geo-spatially referenced, much of it collected by remote sensing systems or in situ field measuring devices. Streaming data play a vital role as well, particularly from atmospheric and hydrologic sensors that sample rapidly changing phenomena. Geosciences data are highly heterogeneous, impact numerous other disciplines (e.g., ecology, medicine, biology), and are notable for their open access, particularly in weather and climate forecasting. All of these data can be characterized by rich meta data and must cataloged, often for decades or centuries owing to the nature of the associated research challenges.

The geosciences also are among the most computationally intensive disciplines, with numerical weather prediction a longstanding driver of digital computing since the days of the ENIAC at Princeton. Today, sophisticated models are being coupled, giving rise to entirely new sets of challenges – both in hardware and software – and requiring computational capabilities orders of magnitude beyond what are available today.

Although numerous other comments could be made about cyberinfrastructure needs in the geosciences, they have been articulated clearly by the community in a variety of documents (see aforementioned references). However, one need that remains unmet is a mechanism for communicating effectively the needs and developments in cyberinfrastructure among all stakeholders in and outside of the geosciences, including those who create and deploy cyberinfrastructure as well as use it in their classrooms. It is for this reason that FGIT is being established.

3. AUDIENCE, GOALS, AND EXPECTED OUTCOME

FGIT is an informal grass roots activity having a national audience with representation from relevant agencies, societies, laboratories, and educational institutions, with a view toward achieving the following:

- Engendering broad dialog among the domain geosciences, information technology and cyberinfrastructure research communities to promote the sharing of ideas on issues of mutual relevance (research, prototyping, hardening, technology transition, maintenance, support,) via the use of appropriate frameworks (e.g., working groups). In the true sense of a forum, FGIT will provide all participants an equal voice;

- Serving as a vehicle for disseminating information on information technology and cyberinfrastructure activities and issues to the broad geosciences community;

- Providing an open forum for the geosciences community to provide input and feedback to the information technology and cyberinfrastructure communities on needed technologies;

- Providing a framework where researchers and operational practitioners, including those in industry, can better understand each other's needs, strategies and constraints;

- Serving as a mechanism for sharing strategies and best practices in information technology and cyberinfrastructure components of geosciences education and outreach; and

- Serving as a mechanism for advocating information technology, particularly as related to the geosciences, to agencies and policymakers.

FGIT will not duplicate or supplant other planning activities or projects but rather provide a communication fabric to enhance them. *The principal outcome of FGIT is expected to be a better educated and more effective geosciences*

and associated cyberinfrastructure/information technology enterprise that is involved with the development of tools, and that takes responsibility for their practical implementation and sustenance.

4. ORGANIZATIONAL STRUCTURE

FIGIT is coordinated by a Planning Committee (Table 1) comprising representatives from academia, mission agencies, and professional societies. (Note that private industry will participate in FGIT but is not presently represented on the Planning Committee. This issue will be re-visited in the near future). Careful attention has been given to diversity and to obtaining individuals from a broad cross section of the geosciences.

The governance of FGIT likely will be similar to, though at least initially smaller in scope and with less complexity than the Global Grid Forum (http://www.gridforum.org/). We envision a rotating Chair, along with a number of working groups that could include the following:

- dynamic and adaptive cyberinfrastructure environments
- sensing and data processing
- modeling frameworks
- streaming/real time data
- metadata generation and protocols
- ontology and cataloging
- distributed data management
- data publication
- data federation
- data and system interoperability
- stewardship of non-digital resources
- data discovery systems
- analysis and visualization
- education and outreach
- networking and telemetry
- intellectual property

The guiding principles for FGIT meetings, which will be held annually, are

- Emphasis on distinct themes rather than projects
- Strong agency involvement
- Travel subsidization
- Involvement of researchers, students, teachers, and operational practitioners
- Practical technology demonstrations
- Significant time for dialog, including informal discussion during extended breaks
- Broad P/R to maximize participation

5. FUNDING AND STATUS

The FGIT Planning Committee met on 16 May and 11 June 2004 and continues to work offline to plan the first meeting, which will be held in the Washington, DC area in spring or early 2005. Funding is being sought from multiple agencies to defray travel costs to maximize participation, especially by students, women and underrepresented ethnic minorities.

6. ACKNOWLEDGMENTS

LEAD is funded by the National Science Foundation under the following Cooperative Agreements: ATM-0331594 (Oklahoma), ATM-

Table 1. The FGIT Planning Committee.

Atmos Sci	Earth Sci	Ocean Sci	Marine G&G	Environ Sci/Hydro	IT/CS	Educ & Outreach	Remote Sensing	GIS/ Viz	Agencies & Societies
Kelvin Droegemeier (Oklahoma)	Tom Jordan (USC)	Peter Cornillon (Rhode Island)	Larry Mayer (New Hampshire)	John Schaake (NOAA/OH)	Beth Plale (Indiana)	Roberta Johnson (UCAR)	Deborah Estrin (UCLA)	May Yuan (Oklahoma)	Steve Meacham (NSF)
Bob Wilhelmson (Illinois)	Chaitan Baru (SDSC)	Dale Haidvogel (Rutgers)	Suzanne Carbotte (Columbia)	David Maidment (Texas/Austin)	Ian Foster (ANL)	Richard Clark (Millersville)	David McLaughlin (Umass/ Amherst)	Sara Graves (Alabama in Huntsville)	Bruce Wardlaw (USGS)
Sepi Yalda (Millersville)					Dan Reed (UNC)	Everette Joseph (Howard)			Ken McDonald (NASA/GSFC)
					Dennis Gannon (Indiana)	Mohan Ramamurthy (UCAR)			Kevin Kelleher (NOAA/NSSL)
						Mary Marlino (DLESE)			Keith Seitter (AMS)
						John Snow (OU)			John Orcutt (AGU)
									Rob Van der Voo (GSA)
									Jon Cole (ASLO)

0331591 (Colorado State), ATM-0331574 (Millersville), ATM-0331480 (Indiana), ATM-0331579 (Alabama in Huntsville), ATM03-31586 (Howard), ATM-0331587 (UCAR), and ATM-0331578 (Illinois at Urbana-Champaign).

7. REFERENCES

Droegemeier, K.K. and Co-Authors, 2004: Linked environments for atmospheric discovery (LEAD): A cyberinfrastructure for mesoscale meteorology research and education. Preprints, *20th. Conf. on Interactive Info. Processing Systems for Meteor, Oceanography, and Hydrology*, Seattle, WA, Amer. Meteor. Soc. (http://lead.ou.edu/pdfs/IIPS LEAD AMS Preprint.pdf).

Estrin, D. and Others, 2003: Environmental cyberinfrastructure needs for distributed sensor networks. Report from an NSF sponsored workshop held at the Scripps Institute for Oceanography, 60 pp. (http://lternet.edu/sensor_report/cyberRforWeb.pdf).

Oldo, J.S. and Others, 2003: Report from the Workshop on Development of a CyberInfrastructure for Research and Education in Structural Geology, 16pp. (http://www.sci.uidaho.edu/cyber/documents/cSIS_report.pdf).

ONR/NSF, 2002: An information technology infrastructure plan to advance ocean sciences. 80pp. (http://www.geo-prose.com/projects/pdfs/oiti_plan_hi.pdf).

NSF, 2003a: Revolutionizing science and engineering through cyberinfrastructure. Report of the NSF Advisory Panel on Cyberinfrastructure. (http://www.communitytechnology.org/nsf_ci_report/report.pdf).

NSF, 2003b: Facilities to empower geosciences discovery 2004-2008, 26 pp. (http://www.geo.nsf.gov/geo/facilities/NSF_Geo2003.pdf).

OSTP, 2004: Federal plan for high-end computing. Report of the High-End Computing Revitalization Task Force (HEC-RTF), 72pp. (http://www.itrd.gov/pubs/2004_hecrtf/20040702_hecrtf.pdf).

NSF, 2004: Cyberinfrastructure for the atmospheric sciences in the 21st century. Report from the ad hoc Committee for Cyberinfrastructure Research, Development, and Education in the Atmospheric Sciences. 54 pp. (http://www.cyrdas.org/report/cyrdas_report_final.pdf).

Walker, J.D., R. Carlson and Others, 2003: Recommendations from the Workshop on Cyberinfrastructure (CI) for the Integrated Solid Earth Sciences (ISES), 6 pp. (http://tectonics.geo.ku.edu/ises-ci/reports/ISESCIreport.pdf).