Open Science Grid Strategy Towards a Coherent US National Cyber-Infrastructure:
The College Campuses
R. Pordes, M. Livny
August 10th, Early DRAFT for discussion, V3

A significant fraction of the current and future intellectual research capital in the US is at the nation's campuses. Delivering the computational capabilities needed by these researchers is of paramount importance. These capabilities increasingly include significant processing and data cyberinfrastructure (CI). Indeed, it is generally acknowledged that now is a critical and opportune time to ensure that CI focuses on enabling the human capital on the campuses. We need ubiquitous use and sharing of CI for all communities.

We see rapid growth in the number of researchers and universities engaging in CI. We see the increasing breadth and depth of the national, multi-user computational facilities. Our goal is to see the local or regional campus "disseminated" facilities integrated with the high-end shared "centralized" national facilities. We aim to enable the full scope of scholarship from single-person research to grand challenge science and from individual student dissertation to nation-wide educational initiatives. In all activities it is crucial that the program is driven through tight collaboration between the research communities and the infrastructure providers.

The Open Science Grid (OSG) Campus Grids program has been part of our activities for more than four years. The goals of this program are to engage researchers and educators on the nations campuses to understand and benefit from the use of distributed computing as an integral part of their activities. Members of the OSG help other scientists to adapt their applications to make use of both (shared) local CI and remote resources on other campuses and national facilities. The national facilities are a critical part of the end-to-end system. The OSG currently includes multi-user facilities like NERSC at LBNL as well as scientific domain facilities like Fermilab. New (in addition to existing) multi-faculty infrastructures such as at Clemson University and the University of Nebraska are becoming accessible to the OSG. The Structural Biology group at Harvard Medical School is a good example of how an initial engagement of a single PI can blossom into providing local support for a community and outreach efforts engaging more users, departments and universities in the area. On the physics side, there are about 30 US LHC "Tier-3" universities participating today, where each has the potential to reach out to other communities locally, and with the number expected to double in the next year or two.

Our goal is to continue and expand this work as far as possible to provide (community driven and policy managed) access across and between the campuses and any nationally shared facility in support of both our current and new science and education communities. We will continue to use current experiences to help more people and campuses in the organizational principles, technologies, services and operational processes they need. We will continue to follow our principle of harmonization between national, community and campus shared cyberinfrastructure. To meet this goal we want to:

- Enable on campus researchers, faculty and students to interact with the national CI – both locally and remotely.

- Organize the campuses for campus wide building, sharing and use of their computational resources locally and with other communities.

- Bring campuses to be full members of the nation-wide shared cyber-infrastructure – bringing additional resources and users to the table.

In working towards our goals we collaborate and partner with many other programs and projects. These include, among others: ESNET Certificate Authority; Internet2 for network tools; more than 30 software groups (including Condor, other DOE SciDAC-2 projects, Globus/CDIGS and the European gLite) contribut-

ing to the Virtual Data Toolkit; the European Enabling Grids for EsciencE; the WorldWide LHC Grid Collaboration; and collaboration with TeraGrid on student training and other activities.

We also work towards expanding and extending our current activities into the future. To do this we lead, participate in and commit to "Satellite" proposals and projects, where the programs of work contribute directly to our vision and the proponents collaborate directly with us. Examples include the Embedded and Immersive Engagement for Cyber-Infrastructure CI-Team project and the "Adopt a High Throughput Computing Cluster " proposal.

Example activities we will continue and aim to expand are:

- Software distributions built, tested, deployed and configured for community CI and end-user needs to support development of applications and use of the distributed resources. Also provide tailored software stacks for on–campus services, resources and users, including software for bridging to other campus and infrastructures. As the number of communities and campus increase we expect there to be a suite of such distributions, each with specific tailoring. Using and extending the NMI/Metronome Build and Test infrastructure is an important component for this activity.

- Organize sharing of campus clusters and storage – using existing models at DiaGrid at Purdue, Clemson, FermiGrid and the Grid Laboratory of Wisconsin (GLOW) at Madison. Help develop and provide templates for designs for the hardware and software to be deployed; help organize on-campus meetings to education and train; architect and provide bridges and gateways that support sharing across multiple campus, including supporting specific communities' needs.

- Provide and support up-load and transfer capabilities across and between campuses, regional, national and international infrastructures, and shared national facilities. Contribute to the worldwide understanding and provision of end-to-end cyber-infrastructure architectures and design, resource requirements and planning processes, inter-domain policy and trust procedures, as well as operational and support activities.

We have made suggestions for broader activities and organizations in other papers (some together with TeraGrid). We include some of our current discussion points:

- *Planning:* Can we broaden gathering of the requirements of and planning for grand challenge science communities, small and large communities of researchers and scholars, education and workforce preparation? Do we continue to do this in our "local" projects or more broadly?

- *Architecture:* Do we continue with our independent articulation of architecture and principles? Can we articulate and gain broader adoption for the model of end-to-end federation of infrastructures (international, national, regional, local) with a hierarchy of interconnecting routers and gateways?

- *Software:* Do we continue on independent paths of acquiring, licensing, developing, supplying and supporting an ad-hoc set of software "as it becomes needed" >? Can we organize a more engineering, design and full-life cycle approach?

- *Integrative Federation:* Do we address up front support across national and local resources? Do we continue on the path of each local project integrating the multi-user and single-community (including DOE) facilities or do this in a broader collaboration and context?

- *Organization:* Do we continue in small pockets of projects or bring some broader vision to the organizational structures. For example, would an "Institute for Campus Computing" with a mission and scope to be an anchor point of support and provisioning for the campus provide be effective as a

partner to the national shared facility organizational structures of the NSF XD program and the DOE leadership class facilities?