

## **Summary of CGL Lab Activities January-December 2008**

The Community Grids Laboratory (CGL) was established in July 2001 as one of Indiana University's Pervasive Technology Laboratories. It is funded by the Lilly Endowment, which provides about one third of the funding with the remainder coming from federal and industry funding. It will continue as part of the Digital Science Center in the new Pervasive Technology Institute. CGL is located in the Indiana University Research Park (Showers) in Bloomington but it is expected to move during 2009 to the new Incubator building near the Wrubel Computer Center. Its staff includes Director Geoffrey Fox, Associate Director Marlon Pierce, 3 senior (post-doctorate) research associates (Drs. Shrideep Pallickara, Judy Qiu, and Kangseok Kim), 2 associate researchers (Drs. Sangmi Pallickara and Yu Ma), 2 web application developers, one software engineer and 10 PhD candidates. We have an international visitors program and 2 Korean scholars visited in 2007-2008 supported by their governments for a year. The students participate in Indiana University's academic program while performing research in the laboratory. 18 CGL students have received their PhD since the start of the lab, and we expect 3 or 4 more students to graduate in 2009.

The Laboratory is devoted to the combination of excellent technology and its application to important scientific problems. Fox has worked in this fashion since he set up the Caltech Concurrent Computation Program (C3P) almost 25 years ago. The technologies we use have changed with field. Starting with parallel computing from 1983 until 1995, we then moved to Web-based computing and education with collaborative technologies. Around 2000, we focused on Grids rather broadly defined to include communities and collaboration. Recently our focus has been multi-core programming and applications while the Grid work continues with a Web 2.0 and cloud computing emphasis.

### ***Research and Development Activity***

#### **Grid Architecture**

We continue our core research in Grid and Web services architecture, which acts as a foundation for all our projects. We now believe that practical systems will inevitably mix Web 2.0 and Clouds with Grid/Web services, and this has been a recent focus with the requirements of data driven science very important. We are also exploring the integration of coarse grain parallel computing with Grid workflow looking for possible unified approaches. We term this Parallel Programming 2.0 which integrates also ideas of Hadoop and Dryad frameworks coming from information retrieval field This work has benefited from our strong involvement with the Open Grid Forum where we lead eScience and several study groups.

#### **Parallelism and Multi-core Chips**

The computer industry will be revolutionized by new chip architectures with multiple cores (processing units) on the same chip. This is illustrated by the Cell processor that IBM has developed for gaming and is highlighted in their new Indianapolis Advanced Chip Technology Center. Moreover, even commodity Intel chips now have 4 and will have over 100 cores in 5 years time. These designs require lower power and potentially

offer huge performance increases. However this requires that one learn how to take parallel computing expertise now largely confined to the science and engineering domain and apply it to the broad range of applications that run on commodity clients and servers. We are part of a major effort in this area funded by Microsoft and in collaboration with Rice University, University of Tennessee, and Barcelona Supercomputing Center with our initial work focused on studying a range of AMD and Intel multi-core architectures and their performance. Currently we have installed clusters of machines with from 8 to 24 cores on individual nodes (24 cores is a 4 chip node with each chip having 6 cores). We are looking into a possible universal runtime for the different forms of parallelism and also at parallel data mining algorithms for multicore chips. Initial parallel algorithms for Bioinformatics, Cheminformatics and Geographical Information Systems (GIS) have been developed with a complete performance analysis. The life science work is in collaboration with the School of Informatics and the Center for Genomics and Bioinformatics in the Biology department. The GIS work is collaborative with the POLIS center at IUPUI.

### **Semantic Scholar Grid**

This project was started in 2006 and is exploring futuristic models for scientific publishing by developing Web 2.0 social networks to support the sharing, annotating and semantic analysis of scientific data and papers. We are building Web service tools that allow integration of capabilities of key systems such as del.icio.us, Connotea, CiteULike, and Google Scholar. The initial system has been completed and has undergone extensive testing in 2008. Two PhD theses are based on this work: investigating difficult consistency questions for metadata prepared on different web sites, security and collaboration problems, and overall architecture investigations. One PhD has been awarded during the current reporting period, and we anticipate the second to be awarded in the first half of 2009.

### **Chemical Informatics and Cyberinfrastructure Collaboratory (CICC)**

The NIH-funded CICC project has developed the Web Services, Web portals, databases, data mining and workflow tools that can be used to investigate and mine the abundant publicly available data on drug-like molecules contained in the NIH's PubChem and DTP databases. As part of this effort, we have developed numerous network accessible services, including services for accessing statistical packages, as well as data services and user interfaces that can perform three-dimensional structure searches across the ten million chemical structures (including one million drug-like molecules) currently in PubChem.

Because we have adopted an open service architecture, our services can be easily integrated with external, third party services. A prominent example includes an online docking results service that we also developed, which calculates the ability of the drug-like molecules to attach themselves to much larger proteins. The initial versions of these calculations were used in the inaugural run of Indiana University's Big Red supercomputer. Our work on open architecture data services is described in a paper submitted to CURRENT COMPUTER-AIDED DRUG DESIGN (see publications).

We have also investigated strategies for increasing the performance of these data query services by using parallel, distributed databases connected with distributed messaging brokers (NaradaBrokering). Cloud computing and virtualization strategies are very important for hosting these types of distributed services, so we also investigated the overhead associated with running multiple services in parallel on virtual hosts. This work was presented in a paper at the e-Science 2008 conference (see publications).

Our current focus has been to investigate the application of data mining techniques developed in our multicore research program such as deterministic annealing and dimensional reduction algorithms. The practical result of this work will be a deployed set of data-mining Web services on our lab's new Windows Server cluster.

The NIH funded-phase for this project will be completed in early 2009, but we will continue our chemical informatics research through funding from Microsoft Research. This new project, led by Cornell University, will include Southampton and Cambridge Universities (UK), as well as Pennsylvania State University and Los Alamos National Laboratory. NIH PubChem architects and developers will be volunteer collaborators. Our lab's contributions to this project will include the application of Web 2.0 approaches to metadata management, investigation of new data centric workflow technologies, and development of Cloud Computing-based data management infrastructure.

#### **e-Science Conference Organization and Audio-Video Collaboration**

Fox served as general chair for the e-Science 2008 conference, which was held in Indianapolis from December 7-12. PTL staff organized the conference, attended by over 350 people. In addition to the organization work, we also developed a new Web-based audio-video collaboration system to provide live broadcasts of the conference presentations as well as persistent access to video archives of the presentations. This Web interface integrates YouTube-like video feeds of the speaker and presentation slides (captured using the Tandberg collaboration system) with speaker-editable blogs and live chat streams. This system provided access to approximately 100 workshops, tutorials, and presentations during E-Science.

#### **Minority-Serving Institutions Cyberinfrastructure Outreach Projects**

This initiative will help ensure that a diverse group of scientists, engineers, and educators from historically underrepresented minority institutions are actively engaged in the development of new Cyberinfrastructure (CI) tools, strategies, and processes. Our key strategy was not to identify particular universities to work with but rather interact with the Alliance for Equity in Higher Education. This consortium is formed by AIHEC (American Indian Higher Education Consortium), HACU (Hispanic Association of Colleges and Universities) and NAFEO (National Association for Equal Opportunity in Higher Education) and ensures our efforts will have systemic impact on at least 335 Minority Serving Institutions.

Our current flagship activity is MSI-CIEC Minority-Serving Institution Cyberinfrastructure (CI) Empowerment Coalition, which builds on success of our initial

MSI CI2 (Minority-Serving Institutions Cyberinfrastructure Institute) project. Activities include workshops, campus visits and pro-active linkage of MSI faculty with Cyberinfrastructure researchers. During 2008 we organized three Cyberinfrastructure days. Two (at Elizabeth City State University and the University of Houston downtown) were focused on individual campuses while the event at New Mexico Highlands University was a regional event for the many MSI's in New Mexico. We produced an assessment for each event discussing the campus or regional opportunities for using Cyberinfrastructure in research and education.

As part of this project, we host the MSI-CIEC project wiki ([http://www.msi-ciec.org/eduwiki/index.php/Main\\_Page](http://www.msi-ciec.org/eduwiki/index.php/Main_Page)) and have developed the MSI-CIEC Portal (<http://www.msi-ciec.org>). This portal is designed to combine the Web 2.0 concepts of social networks with online bookmarking and tagging. By using the portal and services, researchers can bookmark URLs (such as journal articles) and describe them with simple keyword tags. Tagging in turn builds up tag clouds and helps users identify others with similar interests. User profiles provide contact information, areas of research interest, tag cloud profiles, and RSS feeds of the user's publications. The value of social networking sites depends directly on the amount of data and users, so to populate the portal's database, we imported NSF database information on previously awarded projects (available from <http://www.nsf.gov/awardsearch/>) and from the TeraGrid allocations database. This information was converted into tags and user profiles, allowing users to use tags to search through awards by NSF directorate, find the top researchers in various fields, and find networks of collaborators.

During the current reporting period, we completed a redesign of the system's user interface to simplify content management and extensibility. We also added the entire Grants.gov database as a searchable and taggable RSS feed. As part of our backend infrastructure, both the Grants.gov and NSF award notices are made available as RSS syndication feeds. We have also continued our work on building the service infrastructure and algorithms to provide open-source recommendation systems. These rely upon many of the same algorithms used in our multicore research projects. We are researching their application to social networks that can be used to identify (for example) MSI researchers with similar interests. This work spans Web 2.0, portal, and multicore activities.

### **Earthquake Crisis Management in a Grid of Grids Architecture**

This DoD phase II SBIR was led by Anabas with CGL and Ball Aerospace as subcontractors and created an environment to build and manage Net-Centric Sensor Grids from services and component Grids. CGL technologies including our GIS and NaradaBrokering systems are used, and CGL also supplied non-military applications including earthquake crisis management. The project focused on integrating wireless sensors (webcams, RFID, GPS, Lego Robots with environmental sensors) that are integrated and managed using lightweight Linux computers (Nokia N800). The initial system supported initial deployment and dynamic real-time management of Collaborative sensor Grids. The project concluded in July with a very successful final review in Dayton (home of Wright Patterson Air Force Base) and we are planning follow-on work.

### **Particle Physics Analysis Grid and Composable Data Driven Applications**

This DoE phase II STTR aims at an interactive Grid using streaming data optimized for the physics analysis stage of LHC data grids. This differs from the mainstream work of the Open Science Grid and EGEE which concentrates on the initial batch processing of the raw data. We have come up with a novel concept (“Rootlets”) that provides a distributed collaborative implementation of the important CERN Root analysis package. We have built a prototype based on CGL’s NaradaBrokering and the Clarens software from our collaborators at Caltech. It allows collaborative data analysis from multiple distributed repositories and can be applied to any of a class we call “composable” data analysis approaches. Interestingly this includes information retrieval applications and we have compared our NaradaBrokering environment with one based on open source Hadoop implementation of Google MapReduce. This work will be extended to allow use of Dryad environment from Microsoft and a detailed discussion of architecture and performance of software to support composable applications.

### **Polar Grid**

This activity stems from our MSI-CIEC collaboration with Elizabeth City State (ECSU an HBCU) in North Carolina and also has strengthened our collaborations with the Research Computing division of IU’s University Information Technology Services (UITS). We are working with the Center for Remote Sensing of Ice Sheets (CReSIS), an NSF Science and Technology center led by Kansas University, to define and implement Cyberinfrastructure to support modeling and remote sensing of ice-sheets. The recent dramatic evidence of the impact of Climate Change on the Polar Regions makes this an urgent project of great societal importance. We were awarded an NSF Major Research Instrumentation (MRI) grant for this work, which deploys in-the-field, high-performance data processing of sensor data that is linked to dedicated analysis systems (Linux clusters) at Indiana University and ECSU. The first stage of this work has focused on data analysis with parallel SAR (Synthetic Aperture Radar) algorithms in the field in Greenland (Summer 2008) and Antarctica (Winter 2008-2009). The second stage will focus on a new generation of simulation models for glaciers and their melting. These will exploit data gathered by CReSIS and analyzed on Polar Grid.

The IU team led by Prof. Fox was also invited to participate as a full member of CReSIS in the center renewal application to the NSF.

### **QuakeSim and GIS Grid Project**

The QuakeSim project (formerly known as SERVOnet) is funded through NASA’s AIST and ACCESS programs. The AIST funding continues work led by Dr. Andrea Donnellan at NASA JPL to build the distributed computing infrastructure (i.e. Cyberinfrastructure) begun under previous NASA AIST and CT program grants. The Community Grid Lab’s focus in this project is to convert the QuakeSim portal and services into an NSF TeraGrid Science Gateway. We have updated the QuakeSim portal to be compliant with current Java and Gateway standards. We have also developing workflow and planning services based on the University of Wisconsin’s Condor-G software that will enable QuakeSim codes such as GeoFEST and Virtual California to run on the best available NSF and NASA supercomputers. We have spun this work off as a

separate project, the Swarm Service, which we are packaging through our OGCE project (see below). The Swarm Service, although originally developed for submitting earthquake modeling applications, has been successfully applied to protein docking and gene sequence determination, both of which require supercomputing facilities.

The NASA ACCESS project is a joint project that combines team members from the QuakeSim project with the NASA REASoN project. Our work here is to develop and exchange portal components and Web Services with the REASoN team. Exchanged components include GRWS (a GPS data service developed by UCSD/Scripps), Analyze\_tseri (portlets and services developed by CGL and adopted by the REASoN team), and RDAHMM (GPS data mining services developed by CGL using JPL codes and adopted by the REASoN team). The RDAHMM portlets and services are currently being expanded to allow historical analysis of network state changes in the SCIGN (Southern California) and BARD (Northern California) GPS networks. We have also developed services and portlets for interacting with real-time GPS data streams from the California Real Time Network (CRTN). This stream management was based on CGL's NaradaBrokering software, and we demonstrated its scalability to networks 20 times the size of the current CRTN. During the current reporting period, we have spun the GPS-centric services and portal off as a separate, dedicated instance of the QuakeSim activity. We have significantly revised the infrastructure for processing the daily GPS data, developing new services and portlets that greatly increase user interactivity as well as services for consolidating GPS system state information. We have also developed animations of the GPS network state evolution as separate services.

The QuakeSim project's NASA AIST funding was recently renewed for an additional three years.

### **Open Grid Computing Environments (OGCE)**

The OGCE project provides downloadable, generic portal software for building scientific Web portals and gateways. This NSF-funded project is a consortium of several universities and is led by CGL. The OGCE project won a major continuation award from the NSF Office of Cyberinfrastructure in 2007, allowing us to continue the work initially begun under the NSF Middleware Initiative program in 2003. The OGCE website is <http://www.collab-ogce.org>.

During the current reporting system, we continued our software release reorganization. Releases (from November 2008) include

- The OGCE Portal, version 2.4. This is our flagship software release and is described in previous reports. The major addition to the current release is a file-browser applet that allows users to interact with NSF TeraGrid and other Grid file systems, simplifying the process of transferring files between the user's desktop and the TeraGrid. The release also included numerous bug fixes and minor enhancements, many contributed by third party developers.
- OGCE Services, Development Release. This is a packaging solution for the numerous services developed by the project. OGCE services are packaged and downloaded as a single release and can be deployed in a one-step installation.

Services include the Resource Discovery and Resource Prediction services. Several additional services will be added in the next six months.

- GTLAB Tag Libraries, Development Release. GTLAB provides building blocks that allow developers of portlets and gadgets to rapidly build new applications. We significantly revised the core of GTLAB and repackaged it into a one-click installation (following the same conventions used in other OGCE projects) during the current reporting period.
- OGCE Gadget Container, Development Preview. We are making the transition from the portlet standard to Google-style gadgets. The gadget container is a replacement for portlet containers used in the OGCE portal's 2.x releases.

Pierce and the OGCE team (particularly Suresh Marru) organized full-day tutorials at TeraGrid 2008 and Supercomputing 2008. These presentations are listed below.

Finally, the OGCE team led the third Grid Computing Environments workshop (GCE 08) at Supercomputing 2008. This year's workshop featured 13 peer-reviewed papers. Proceedings will be published by the IEEE.

### **NaradaBrokering and Granules Projects**

NaradaBrokering's content dissemination capabilities underlies many lab projects. During the current reporting period, we released version 3.3.2 of the software. This release incorporates support for hierarchical topics. Hierarchical topic spaces provide greater flexibility in the selecting data streams that a consumer would be interested in. Hierarchical streaming simplifies the process of registering interest in content. Without support for hierarchical streaming, every consumer would need to be aware of every finer-grained description of content. We have incorporated support for explicit, and implicit, wildcard-based subscriptions within these hierarchical spaces. NaradaBrokering's core software is written in Java, but we also supply a C++ bridge for non-Java applications such as the Rootlets application. We have released version 2.0.0 of this C++ bridge summer 2008.

We have also begun a new, related project, Granules, to investigate Cloud Computing research issues. The Granules runtime for Cloud Computing can orchestrate millions of computations concurrently on a cloud of computers. At a given machine, Granules can interleave the concurrent execution of thousands of computations.

Granules also incorporates support for the map-reduce programming model, which facilitates the concurrent processing of large datasets. Here, large datasets are split into smaller more manageable sizes which are then processed by multiple *map* instances. The results produced by individual map functions are then sent to *reducers*, which collate these partial results to produce the final output. A clear benefit of such concurrent processing is a speed-up that is proportional to the number of computational resources.

Typical MapReduce stages look like a directed acyclic graph with the MapReduce execution progressing in monotonically increasing stages. Besides the basic support for MapReduce, we have incorporated support for variants of the MapReduce framework

that are particularly suitable for scientific applications. This includes support for iterative and recursive implementations of the framework. This feature is particularly useful in clustering algorithms where the computed (reduced) results need to be refined (to meet a targeted error rate) by the map stages.

There is clearly a pressing need for runtimes like Granules. Traditional batch systems tend to rely on executing a large computational task that runs for several minutes-to-hours; during this time, no other tasks can be scheduled on the resource. The proliferation of pervasive environments, sensors and other communications-enabled devices has resulted in the need for an infrastructure that can orchestrate and satisfy the real-time processing requirements of millions of computations whose CPU-bound times during a given execution would vary from a few seconds to several minutes. This is the type and scale of computations that Granules manages.

An internal release, version 1.0.1, of Granules was made available at the end of 2008.

### **OMII Software**

We were funded by the UK Open Middleware Infrastructure Institute (OMII) to develop core Web Service (Grid) support for reliable messaging (FIRMS) and notification (FINS). Both these software packages have been successfully deployed within the latest version of the OMII Container. These projects are complete.

### **Collaboration Grids**

This project generated important input for the audio/video transport component of NaradaBrokering, and we are focusing on improving the core infrastructure and the application to e-Sports for sharing and annotating real time video between trainers and athletes. This is supported by an ongoing research effort that focuses on developing algorithms and software for the annotation of streams. Currently, we have developed content handlers for streams based on video, audio, text, whiteboards and images. In the case of video and audio we also have support for different multimedia codecs. For video, we have support for JPEG, H261, H263 and MPEG4 codecs, while in the case of audio we support ULAW, GSM, G723, and DVI codecs.

A stream can be annotated with any another type of stream. Thus, a video stream can have multiple audio (or video) streams associated with it. During playbacks, a client may choose specific annotation streams that will be played-back alongside the original stream.

The system also incorporates support for collaboration. Authorized entities can visualize annotations as they are being made remotely. All streams within the system utilize NaradaBrokering for its content dissemination needs

## ***Educating the Residents of Indiana and Beyond (includes outreach)***

The Community Grids Laboratory has major activities in outreach to Minority Institutions faculty and students. These efforts are motivated by the observation of a Dr. Richard Tapia, Rice University Professor and distinguished Hispanic American scientist:

*“No first-world nation can maintain the health of its economy or society when such a large part of its population remains outside all scientific and technological endeavors.”*

Our work hinges on the observation that Cyberinfrastructure and its underlying Grid technology inherently bridges the Digital Divide and can broaden participation in science and provide better education and business opportunities. Currently our activities are focused with the Navajo Nation in providing education and health applications for their Grid and with HBCU’s Elizabeth City State and Jackson State Universities. As described above, IU and ECSU are partners in the Polar Grid project. We hosted an undergraduate student this summer from Jackson State as part of IU’s university-wide HBCU initiative for the second year. There are clear ways that our work could be extended to K-12 education but proposals in this area have not been successful.

Prof. Fox and CGL staff members frequently lecture on their research and broader topics as part of seminars and courses offered at Indiana University and IUPUI. A comprehensive list of publications is available from <http://grids.ucs.indiana.edu/ptliupages/presentations/>. The following presentations highlight our outreach seminars and lectures to students, minority-serving institutions, and general (non-technical, non-specialist) audiences:

- Geoffrey Fox, “Cyberinfrastructure and PolarGrid”. Presented to the North Carolina Research and Education Network during NCREN Community Day (via video conference) December 5 2008.
- Geoffrey Fox, “Cyberinfrastructure: An Opportunity for UHD”. Presentation at Cyberinfrastructure day, CI@UHD, at University of Houston Downtown November 7 2008
- Geoffrey Fox, “Technology Futures and Lots of Sensor Grids”. Presentation at Association of Pacific Rim Universities (APRU)’s 4th Earthquake and Tsunami Symposium August 21 - 22, 2008 University of California, Davis (USA).
- Marlon Pierce, “QuakeSim Project: Portals and Web Services for Geo-Sciences”, Association of Pacific Rim Universities (APRU)’s 4th Earthquake and Tsunami Symposium August 21 - 22, 2008 University of California, Davis (USA)
- Geoffrey Fox, “Research Issues and IUPUI Informatics”. Presentation to IUPUI Informatics Department, Bradford Woods August 15 2008.
- Alex Ramirez (Hispanic Association of Colleges and Universities), Geoffrey Fox (Indiana University), Al Kuslikis (American Indian Higher Education Consortium), Richard Alo (University of Houston-Downtown), Karl Barnes (National Association for Equal Opportunity in Higher Education), Diane Baxter (San Diego Supercomputer Center), and Julie Foertsch, “Engaging Minority-Serving Institutions (MSIs) in Cyberinfrastructure (CI) through CI Days” The 3rd annual TeraGrid Conference, TeraGrid '08 All Hands Meeting Riviera Hotel and Casino in Las Vegas Nevada, June 11 2008.

## ***Accelerating Economic Growth***

The Community Grids Lab partners with several small business ventures. These are described in greater detail in the previous sections.

- We completed a DOD Phase II SBIR to dynamically build and manage Grids. CGL partners with Anabas (a small startup company that leads the project) and Ball Aerospace.
- We are continuing work on a DOE Phase II STTR with Caltech and Deep Web Technologies .
- Anabas and CGL won an additional DOE Phase I STTR for collaborative visualization systems for Plasma Physics.

Our work with both Anabas and Deep Web technologies has focused on building collaborative sensor and data grids. We have presented this work to Boeing and Nokia.

- Geoffrey Fox and Alex Ho, “Collaboration, Grid, Web 2.0, Cloud Technologies” Boeing Seattle Presentation August 14 2008.
- Alex Ho “Anabas: Taking Collaboration to the Next Level”. Presentation at Nokia Research Laboratory Palo Alto July 13 2008.
- Geoffrey Fox and Alex Ho, “SBIR Final Meeting: Collaboration Sensor Grid and Grids of Grids Information Management”. Presentation at Ball Aerospace July 9 2008.

## ***Bringing Distinction to Indiana University and the State of Indiana***

Geoffrey Fox continues as Vice President responsible for e-Science for the Open Grid Forum. He was general chair of the annual e-Science conference, which was hosted by Indiana University at IUPUI in December 2008. The conference featured over 150 papers and posters and approximately 350 attendees, including a strong international presence.

An interesting milestone is that Fox has reached a total of 58 Ph. D. theses supervised. Fox has also been given courtesy positions at the University of Southampton (UK, renewal), University of Houston Downtown and the Alliance for Equity in Higher Education to recognize importance of collaborative work.

Marlon Pierce continued his organization of the GCE workshop at Supercomputing 2008. This was the workshop’s fourth year and was attended by over 50 people and featured 13 peer-reviewed papers.

## ***Lab Research Outlook for 2009***

### **Grid Architecture**

We continue this foundation activity focusing on interaction of Grid, Web 2.0 and digital library technology.

### **Parallelism and Multi-core Chips**

We expect this activity to grow in importance with a focus on applications that are likely on future multicore clients. We expect to install a 64 node parallel cluster with 16 to 24 core nodes. This application work will be in conjunction with research in new run time systems. Our major application focus will be Bioinformatics focused on Multiple Sequence Alignment.

### **Semantic Scholar Grid**

This Web 2.0 activity will be augmented by a broader range of projects looking at ways of using Web 2.0 approaches in file transfer, document sharing and people networking in science and education. We will also complete our security infrastructure.

### **Chemical Informatics and Cyberinfrastructure Collaboratory (CICC)**

We will complete the NIH-funded phase of this project by deploying Web service-wrapped data-mining algorithms developed in our multicore research project on our new Microsoft Windows Server cluster. These will be applied to NIH PubChem and related data sets. This project will then be transitioned to the Microsoft Research-funded e-Chemistry project. We will examine Web 2.0 metadata formats, security issues, and data cloud computing.

### **Open Grid Computing Environments (OGCE)**

We will add a number of services to our OGCE Services release. These will include the Grid Portal Information Repository, Swarm Mass Job Submission, and XMC-Cat Metadata Management services. We will enhance GTLAB to provide support for RSS feed ingestion and OpenID security. The OGCE Container will be improved to support the Google-led Open Social specification for building collaborative gadgets.

### **Minority-Serving Institutions Cyberinfrastructure Outreach Projects**

We will continue our campus visit programs and follow up on areas (student research opportunities and cyberinfrastructure enabled education) identified as critical by earlier visits.

We will revisit the current architecture of the MSI-CIEC portal to exploit more open standards, including Open Social and Facebook programming interfaces. We will also exploit numerous programming APIs to backend cloud services such as YouTube, Google Calendar, Google Docs, and Blogspot. We expect to participate in an NEH (National Endowment for Humanities) funded activity to provide Web 2.0 portals supporting user annotation of Smithsonian Natural History and American Indian digital artifacts.

### **Earthquake Crisis Management in a Grid of Grids Architecture**

The continuation of this work is still being discussed with a probable focus on sensor Grids integrated with lightweight computing devices.

### **Polar Grid**

We will develop a prototype PolarGrid portal in collaboration with ECSU and the CReSIS team. We will base this on Web 2.0 approaches, including KML, GeoRSS, and Open Social-style gadgets. We will also support the CReSIS team's high throughput signal image processing requirements for field-collected radar imagery of glacier depths.

### **NaradaBrokering, Granules, and Particle Physics Analysis Grid**

During this period we are planning on releasing software that incorporates production implementations of our scheme for the scalable tracking of distributed entities. An additional capability would be to support the voluminous replay/recording of multimedia streams produced within the eSports project.

We also plan to release a prototype version of the framework, using NaradaBrokering, which will be used within Clarens to discover, and load-balance accesses to, services that are available to physicists for analyzing, and collaborating over, data produced in particle physics experiments.

We will continue the development of the new Granules project with a focus on novel Cloud Computing applications.

### **QuakeSim and GIS Work**

We will add services to support updated versions of GeoFEST and Virtual California applications. Both will require access to Supercomputing facilities at NASA and the TeraGrid. We will use the Swarm service to support this work. We will also begin the transition to the new AIST project, which will focus on prototyping QuakeSim data and computing services in a Cloud Computing infrastructure. This will rely in part on new Eucalyptus-based Cloud services at Indiana University.

### **E-Science Audio-Video Collaboration**

We will complete the archiving phase of this project so that the presentations will be made available through this persistent Web interface. We will also look for opportunities to embed this technology in other Tandberg broadcasts in Indiana University.

**Collaboration Grids and e-Sports:** We plan to do a release of this software in February 2009. Subsequent releases will be aimed at addressing issues related to organizing annotations, building redundancy into the system and support for additional codecs.