TATRC
Telemedicine & Advanced Technology Research Center

Integrated Research Team Final Report HealthGrid: Grid Technologies for Biomedicine

1-2 March 2006

IRT Co-Chairs:

Ms. Mary Kratz
Jonathan Silverstein, MD, MS, FACS
Parvati Dev, PhD

November 2007

Approved for public release. Distribution unlimited.

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

EXECUTIVE SUMMARY

The US Army Medical Research & Materiel Command (USAMRMC) Telemedicine & Advanced Technology Research Center (TATRC) convened an *Integrated Research Team (IRT) on HealthGrid: Grid Technologies for Biomedicine* on 1-2 March, 2006. The proceedings and conclusions from this meeting are presented in the following report. This paper was prepared with the involvement of leading biomedical researchers, computer scientists, engineers, and Information & Communication Technology (ICT) experts. The purpose of this document is to provide a research roadmap towards "Net-Centric Healthcare"[1], all the while reconciling with future USAMRMC priorities. In an effort to produce tools that enable/enhance research, education, and direct patient care under the umbrella of the Department of Defense Tricare System, this report concludes with a call for TATRC (and others) to enable Grid technologies via applied research programs for the Military Healthcare Service (MHS).

This document serves to communicate expectations and requirements to all parties - end users, policymakers, scientists, as well as technology leaders. The roadmap provides a foundation upon which TATRC may organize its strategic priorities and commit to resource allocations.

HealthGrids are computational environments of shared resources in which heterogeneous and dispersed biomedical data is accessed, mined and computed. Advances in computer science allow biomedical researchers to capitalize on *ubiquitous & transparent* distributed systems and a broad set of tools for resource sharing (computation, storage, data, replication, security, semantic interoperability, and delivery of software as services). The foremost opportunity of the HealthGrid is the creation of a research climate that transforms the biomedical sector into a functional knowledge society, as the contemporary equivalent of the original Advanced Research Projects Agency Network (ARPANET)[2] transformed networks to become today's Internet.

To synthesize a roadmap from the expertise of the IRT participants, four break-out groups were convened to delineate milestones and estimate requirements necessary to overcome current implementation barriers. A mix of academic, industry and government subject matter experts addressed both technological and socio-political dynamics. The HealthGrid was catalogued into four discrete functions which interoperate: Computation Grids, Data Grids, Knowledge Grids, and Collaboration Grids. The break-out groups mirrored this taxonomy.

Systems Medicine[3] provided a framework for the HealthGrid IRT and resultant research roadmap. Contributors worked off the predicate that 'functions of life are mediated by biological networks, and human disease occurs when one or more of these networks become perturbed by genetic mutations and/or abnormal environmental signals'. Systems Medicine is driving development of new in-vivo and in-vitro measurement technologies, which, collectively, in the next 5-20 years, will lead to medical care that is predictive, preventative, personalized and participatory (P4 medicine).[4] Grid technologies provide a powerful technology infrastructure and enabler to attain this goal. Derivatively, the HealthGrid aims to foster innovation in how disease is viewed, diagnosed, treated, and prevented.

This decade is witness to major revolutions in our comprehension of the human genome, infectious disease, environmental influences, cancer and aging. The advent of electronic health records in combination with Grid technologies and growing interdisciplinary collaborations present the opportunity to participate in new scientific and medical breakthroughs at an ever-accelerating pace, emblematic of our transformation to a 'knowledge society'. This report attempts to refine the broad landscape of the HealthGrid into scalable and 'digestable' targets of opportunity.

ACKNOWLEDGEMENTS

This report reflects not only the scope and diversity of HealthGrid research and development activities but the editorial contributions of many individuals, to whom TATRC extends sincerest thanks.

A special acknowledgement to Dr. Leroy Hood for providing the Systems Medicine framework for this Integrated Research Team and his leadership of shared vision for the HealthGrid.

IRT Co-Chairs:

Ms. Mary Kratz, University of Michigan Medical School Information Services and Telemedicine Advanced Technology Research Center
Jonathan Silverstein, MD, MS, FACS, Associate Director Computation Institute, University of Chicago and Argonne National Laboratory
Parvati Dev, PhD, Director, SUMMIT Laboratory, Stanford University

DISTINGUISHED SPEAKERS:

Michael Ackerman, Assistant Director, National Library of Medicine, NIH
Joel Bacquet, MBA, Scientific Officer, European Commission
Ken Beutow, PhD, Director, NCI Center for Bioinformatics, NIH
Peter Covitz, PhD, Director, Core Infrastructure, National Cancer Institute, Center for Bioinformatics, NIH
Kenneth Curley, MD, Chief Scientist, Telemedicine Advanced Technology Research Center
Don Detmer, MD, MA, President and CEO, American Medical Informatics Association
David Forslund, PhD, Senior Scientist, Cognition Group
Ian Foster, PhD, Professor, Argonne National Laboratory and University of Chicago
Mr. Robert T. Foster, Program Executive Officer, Military Health Service, Joint Medical Information Systems
Peter Highnam, PhD, Senior Advisor to the Director, National Center for Research Resources, NIH
Chris Johnson, PhD, Director, SCI Institute
Carl Kesselman, Director Center for Grid Technologies, USC Information Sciences Institute
Caroline A. Kovac, PhD, General Manager, IBM Life Sciences
Fillia Makedon, PhD, Program Director, Office of Cyberinfrastructure, National Science Foundation
James Wilson, MD, Telemedicine Advanced Technology Research Center, Georgetown ISIS

WORKING GROUP CHAIRS:

Parvati Dev, PhD, Director, SUMMIT Laboratory, Stanford University
Ian Foster, PhD, Professor, Argonne National Laboratory and University of Chicago,
Leroy Hood, MD, PhD, President, Institute of Systems Biology
Chris Johnson, PhD, Director, SCI Institute
Carl Kesselman, Director Center for Grid Technologies, USC Information Sciences Institute
Reagen Moore, PhD, Director, Data and Knowledge, San Diego Supercomputer Center
Tim Pletcher, Director of Applied Research, CMU Research Corporation
Marc Wines, General Services Agency

IRT PARTICIPANTS:

Bob Aiken, Senior Systems and Internet Architecture, Argone National Labs
Brian Athey, PhD, Director/ Principal Investigator, NCIBI, University of Michigan
Joel Bacquet, Scientific Officer, European Commission, DG INFSO/H1
James Bayuk, MD
Howard Bilofsky, PhD, Senior Fellow, University of Pennsylvania
George Brett, BA, MFA, Manager, Information and Knowledge Management, Internet 2
Jack Buchanan, MD, Professor and Staff Physician, Unversity of Tennessee Health Science Center
Todd Carrico, PhD, CEO, Cougaar Software
Timothy Clark, Director of Informatics, Mass General Institute for Neurodegenerative Disease
Donald Cook, MS, President, Shoreland, Inc.
Gary Crane, BSEE, MBA, Director of IT Initiatives, Southeastern University Research Association
Dave Damassa, PhD, Dean for Information Technology, Tufts University School of Medicine

Lori DeBernardis, Director Marketing & Public Relations,Telemedicine Advanced Technology Research Center
Jonathan Dugan, PhD, Knowledge Engineering Consultant, Stanford University
Susan Estrada, President, Firstmile.us
Michael Fitzmaurice, PhD, FAO, Senior Science Advisor for IT, Agency for Healthcare Quality, Department of Health and Human Services
Larry Flournoy, BS, Associate Director, Texas A and M University
Richard Foster, MHA, CPH, Principal Consultant, Joint Medical Information Systems
Robert Foster, Program Executive Officer, Joint Medical Information Systems
Robert Foster, PhD, SES, Director, BioSystems, ODUSD (S&T)
Thomas Hacker, PhD, Associate Director, Research and Academic Computing, Indiana University
Peter Honeyman, Scientific Director, CITI University of Michigan
Ray Hookway, PhD, Senior Member Technical Staff, Hewlett-Packard
Jim Karkanais, Senior Director Health Strategy, Microsoft
Phillip LaJoie , Chief Technology Officer, TriCARE
H. David Lambert, VP and Chief Information Officer, Georgetown University
Yannick Legre, CNRS- HealthGrid
Wilfred Li, PhD, Executive Director, National Biomedical Computation Resource UCSD
Bill Lorensen, MS, Graphics Engineer, GE Research
Harvey Magee, Portfolio Manager, Telemedicine Advanced Technology Research Center
Michael Mauro, BSEE, MS, Chief Engineer, JMISO
Sidney McNairy, PhD, Director, Division of Research Infrastructure
Greg Mogel, MD, Deputy Director, Telemedicine Advanced Technology Research Center
Kevin Montgomery, PhD, Technical Director, National Biocomp Center
Leon Moore, PhD, Interim Chair, BID/USUHS
Regan Moore, PhD, Director, Data and Knowledge, San Diego Supercomputer Center
Stephen Moore, Director, Advanced Research Computing, Georgetown University
Ron Pace, IPA, MHS Chief Enterprise Architect, TMA/TATRC/U of A
Hon Pak, MD, LTC, Director, Advanced Information Technology Group, Telemedicine Advanced Technology Research Center
Michael Papka, University of Chicago/Argonne National Laboratory
Beth Plale, PhD, Faculty, Indiana University at Bloomington
Tim Pletcher, Director of Applied Research, CMU Research Corporation
Kenneth Rubin, Health Enterprise Architect, EDS
Stanley Saiki, MD, Director, Pacific Telehealth
Joel Saltz, MD, PhD, Professor and Chair, The Ohio State University
Jay Sanders, MD, President CEO, The Global Health Group and Telemedicine Advanced Technology Research Center
Vish Sankaran, Acting Program Manager, Office of National Coordinator for HIT
Melvin Sassoon, VP Federal Solutions, Cougaar Software
Michael Sayre, PhD, Program Official, National Institutes of Health
Richard Soley, PhD, Chairman and CEO, Object Management Group
Tony Solomonides, PhD, Reader-Medical Informatics, UWE, Britol/HealthGrid
Anil Srivastava, Chief Strategy Officer, Across World/CTIS
Russell Truscott, InfoLab, AT&T Labs-Research
Daniel Updegrove, PhD, Vice President, Information Technology and CIO, University of Texas at Austin
John Wilbanks, BA, Executive Director, Science, Creative Commons

ROADMAP REVIEWERS and CONTENT CONTRIBUTORS:

Jonathan Dugan, PhD, Knowledge Engineering Consultant, Stanford University
Ken Hall, Manager, BearingPoint
Holly Pavliscsak, BS, MHSA, Medical Writer, Telemedicine Advanced Technology Research Center
Beth Philipson, VMD, MBA, Technical Writer
Jeffery Roller, MD, ret. COL USAF, USAMRMC/TATRC
Tom Savel, MD, Centers for Disease Control, National Center for Public Health Informatics
Rick Stevens, Argonne National Laboratory
Wayne Wilson, University of Michigan

ACRONYMS

ARPANET	Advanced Research Project Agency Network
BIRN	Biomedical Information Research Network (http://www.nbirn.net)
caBIG	Cancer Biomedical Informatics Grid (https://cabig.nci.nih.gov/)
HIPAA	Health Information Portability and Accountability Act
HPC	High Performance Computing
ICT	Information Communication Technology
IRT	Integrated Research Team
PHI	Personal Health Information
PACS	Picture Archive Computer System
MAGIC	Middleware and Grid Infrastructure Coordination
NGI	Next Generation Internet
NITRD	National Information Technology Research Development Program
SII	Scalable Information Infrastructure
TATRC	Telemedicine & Advanced Technology Research Center
USAMRMC	United States Army Medical Research & Materiel Command
VO	Virtual Organization

Outside the Morningside Inn - Photo Credit: J. Harvey Magee

BACKGROUND

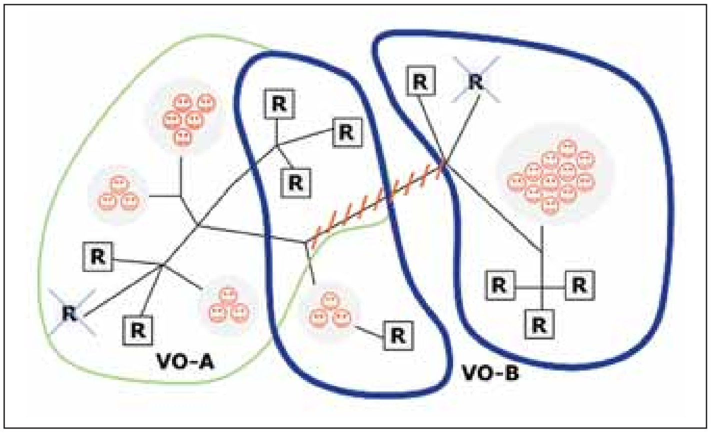
The Telemedicine Advanced Technology Research Center (TATRC) is a laboratory of the United States Army Research and Material Command (USAMRMC). Its mission is to make medical care and services more accessible, reduce costs, and raise the overall quality of military healthcare through the effective application of advanced technologies. TATRC provides technical support to federal and defense agencies; develops, evaluates, and demonstrates new technologies and concepts; and conducts market surveillance with a focus on effective use of emerging technologies in healthcare. It also manages scientific research from congressionally mandated programs, in-house research, Small Business Innovative Research (SBIR), and cooperative research partnerships with international researchers. TATRC projects account for approximately $500 million in government expenditures annually and include over five hundred health science research and development projects.

In its grasp of transformative technologies in virtualization, visualization, service-oriented architecture, High Performance Computing (HPC), distributed computing, Scalable Information Infrastructure (SII) and Next Generation Internet (NGI), TATRC committed to align with global leaders working on the HealthGrid. The MHS 'enterprise' provides an ideal environment to leverage the power of virtual organizations. It is hoped that sponsorship of HealthGrid applications may lead to the implementation of ICT tools useful for other US government agency programs. Inter-agency coordination is a prerequisite for long-term success. In that spirit, TATRC utilized the Integrated Research Team (IRT) to bring together experts from academia, government, and industry in order to formulate tactics & strategies, all sharing a common vector. Subsequent to the HealthGrid IRT, and inter-agency meeting was held at the National Science Foundation on 2 October 2006. The report from that meeting is found in the appendices of this report.

GRID COMPUTING

Cyberinfrastructure is a term originated by the National Science Foundation (NSF) to describe the information technology resources used by researchers, clinicians, engineers and artists to create new knowledge.[5] The HealthGrid is based on Grid computing, which is a specific standard architecture in which new, often high-performance, computational services and applications are created by integrating ICT resources (computer processors, memory, disks, data, algorithms, and people) across distributed geography and organizations.[6] Grid computing is an emergent enabling cyberinfrastructure technology. Grids use a common middleware service layer to establish common interfaces to underlying application resources by integrating multiple networked computers, process execution, datasets, and computational resources across a broad physical infrastructure.

For many applications in the biomedical sector, Grid computing shifts the burden for solving computational issues away from technical resources and onto issues of human collaboration. While there will always be technical challenges in building faster and more capable computer systems, grouping resources into a Grid infrastructure provides us with the ability to build 'virtual organizations'. A virtual organization is a corporate, administrative, not-for-profit, educational or otherwise productive business entity that does not have a central geographical location and exists solely through telecommunication tools[7]. A typical virtual organization comprises a set of (legally) independent enterprise that share resources and skills to achieve shared mission - but most importantly - a virtual organization is not limited to an alliance of just medical organizations, thus enabling the resolve of complex data access and usage issues. The interaction among members of the virtual organization is mainly done over the fabric of the Grid, through computer networks and shared Scientific Commons[8] that enable trust. Grid services are available today to provide security mechanisms[9] to protect data and comply with the regulation of the Health Insurance Portability and Accountability Act (HIPAA) for protection of Personal Health Information (PHI). Computer code can incorporate policy, such as HIPAA regulation into easily deployable Grid services that genuinely scale to the requirements of the biomedical community.



Scholars must be enabled to solve their own particular research challenges independently with transparent assembly of required resources to create virtual organizations. Virtual organizations involve ad-hoc participation in interdisciplinary scientific research teams through local, regional, national and international communities of practice. Such an approach enables many people to work in parallel using shared data and share vast computational power toward solving shared problems.

Grid computing provides more than just a tool to distribute jobs to multiple processors simultaneously. Grid principles apply to all distributed resources and management of innumerable processors. It enables the software developer to focus on the needs of the end user first, and leverage the existing Grid service framework for systems integration and distribution of ICT resources. Dataset that are managed within a Grid infrastructure can be made securely available to remote applications whenever necessary.

THE EMERGING HEALTHGRID

The scope of the emerging HealthGrid is to translate research knowledge to enable better health outcomes within communities and to develop information that "health decision makers" (i.e., individuals, medical practitioners, payers and administrators) can use.

Providing software as a service requires attention to data, workflow and outcomes. Data management is an essential component that often requires physical experimentation and measurement to enable the digitization of medical data for standard network transit, storage, analysis and visualization. Three classes of applications have proven benefit from Grid computing:

- COMPUTE INTENSIVE applications, including interactive simulation (surgical simulation and modeling), very large-scale simulation and analysis (Public Health Information Networks, systems biology frameworks, NeuroCommons), and engineering (parameter studies and linked component models)

- DATA INTENSIVE applications, including experimental data analysis (computational biology), imaging (Radiology PACS) and sensor analysis (biosurveillance, situational awareness)

- DISTRIBUTED COLLABORATION applications, including online instrumentation (microscopes, mass spectrometry, medical devices), remote visualization (advanced distributed learning (anatomy curriculum, public health population studies), and engineering (large-scale structural testing of distributed systems, biochemical engineering)

The widespread use of grid computing is the next logical step on the ICT development path from computation, to shared datasets, and ultimately to knowledge-based environments[10]. Grid computing offers functionality and methods to build an interaction framework for biomedical application environments. A primary benefit is the ability to re-use Grid services without having to rebuild each time as new user requirement arises. Use of the Systems Medicine framework to form HealthGrids is fundamentally different from how most software is developed today, even in large distributed systems.[11]

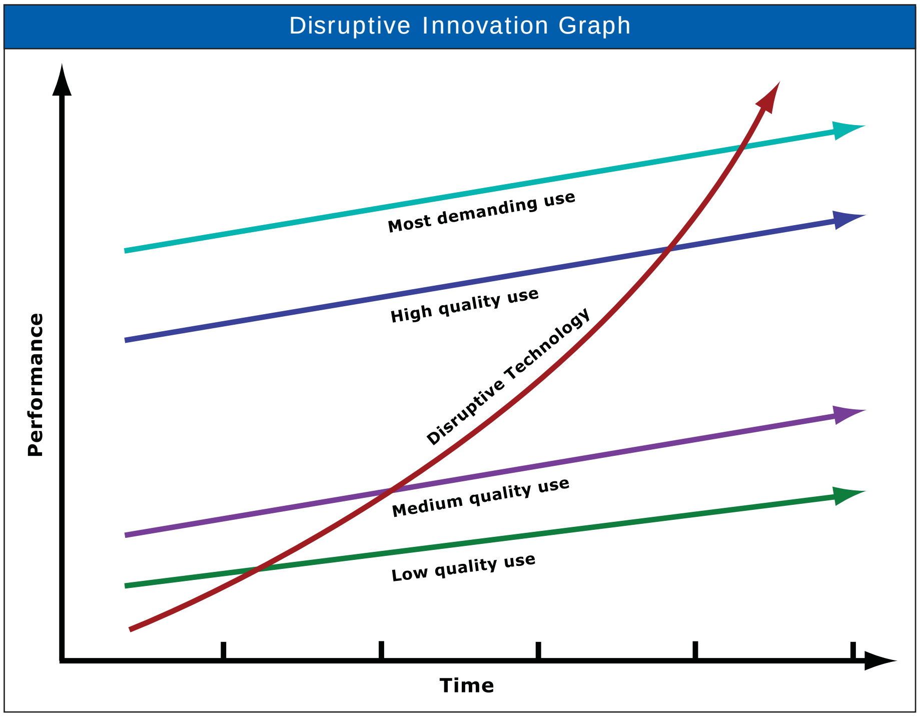
HEALTHGRID IRT - FOUR TOPICAL AREAS

The use of grids is particularly applicable to biomedicine given the enormous quantities of medical data that are being collected and stored in a variety of heterogeneous systems and formats (genomic, proteomic, medical images, medical device output, electronic health records, instrumentation, etc). Advances in medical informatics, text and data mining capabilities, natural language processing, business analytics, visualization, computer-aided drug and therapy design, and simulation have heightened demand for access to and coordination of these heterogeneous data sources. The creation of a common biomedical Grid framework offers promise to advance significantly beyond current scientific and business economic models.

A fundamental goal for the HealthGrid community is to develop the ICT framework for systems-oriented biomedical activities. Akin to Metcalf's Law[26] that stipulates the value of a telecommunications network is proportional to the square of the number of users of the system, the value of HealthGrid will increase proportional to the number of biomedical implementations using standard Grid technical architectures. The end result of access to more data and stronger analytical tools will produce innovative new diagnostics and therapeutics for patients, improving the chances for recovery from some of the most devastating illnesses and over time, and potentially reducing the cost of healthcare through advances in medical research enabling earlier, more precise diagnosis, and more efficient effective therapy.

The economics of disruptive innovation (as defined by Clayton Christensen of the Harvard Business School) is telling in that low end market commodities ultimately emerge as the dominant supply model. Today's vendors of ICT services are already providing Grid based services as low end commodity utilities, such as hosted applications and remote data centers. These vendors are quick to adopt open-source software, fully appreciating the economic value of moving to commodity components, delivering software as services, and avoiding the capital expenditures required to operate many enterprise systems today.

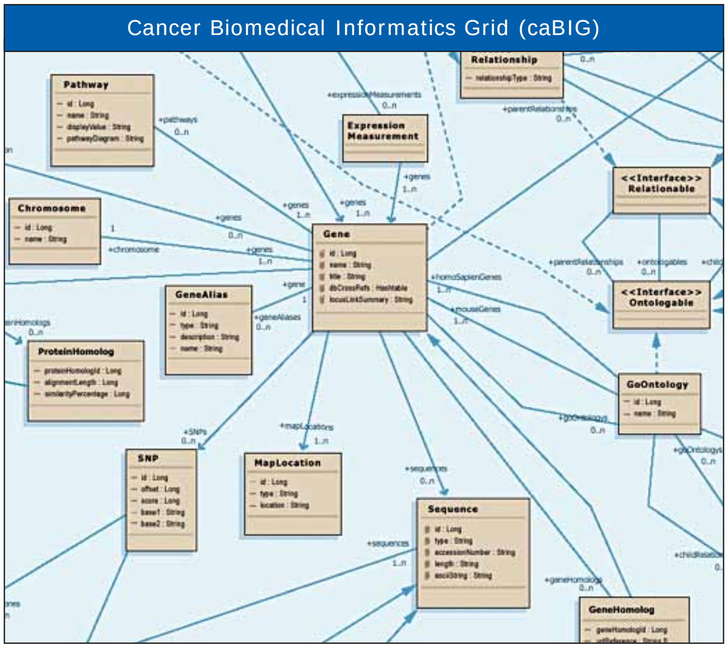
Disruptive Innovation Graph



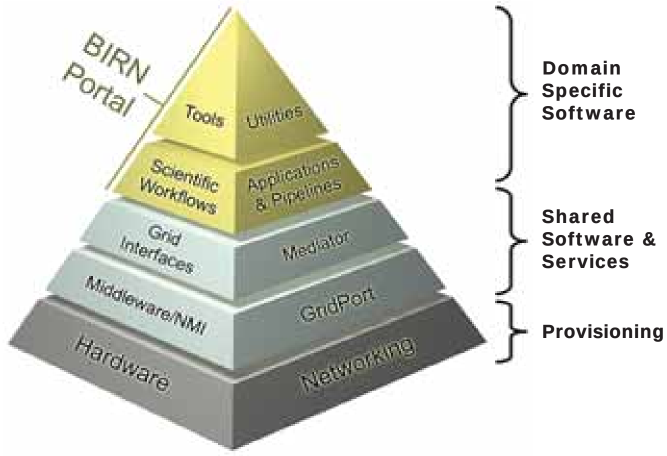
For the purposes of this IRT, four topical areas were delineated:

1. Computational Grids (focus: Life Science Grids):
Computational Grid methods involve vector and parallel processing, data acquisition, digitization, storage, analysis, and visualization. Grid computing today broadly facilitates high performance systems. There are many examples of computational Grids in use today. One is the Cyberinfrastructure for Systems Medicine with a focus on computational biology methods being developed by the TeraGrid[17] program, providing a computational framework to manage, analyze and visualize large datasets. System Medicine is based on computing across billions of individual data points that will require a 'low performance' grid technical infrastructure. Dynamic aggregate of the power of a large number of individual (small) processors represent distributed applications that are simply not possible to implement with a single system. A virtual terascale computer can dynamically aggregate the power of a large number of individual computers (sensors, physiological monitors, personal Electronic Health Record over PDA) in order to provide a platform for advanced applications. Distributed computational functions best done with Grid computer systems involve data acquisition, digitization, simulation, large storage, analysis, and visualization.

2. Data Grids (focus: Translational Research Grids):
This area focuses on the creation of an integrated collection of locally curated data resources, which are seamlessly accessible, interoperable, and usable across a wide (wide) area network. The federation of multiple datasets enables new methods for biomedical science for data mining, ontologies, semantic interoperability, analysis and visualization. Data Grids are characterized by the federation of datasets from locally curated data resources to manage, organize, store and distribute asynchronously and in real-time. The Biomedical Information Research Network (BIRN) http://www.nbirn.net and the Cancer Biomedical Informatics Grid (caBIG) https://cabig.nci.nih.gov/ are two examples of programs that have deployed shared federated datasets over standard grid architecture.



The BIRN shared information technology infrastructure for basic and translational research is available to all researchers from any Internet-capable location.



3. Information and Knowledge Grids (focus: Epidemiological Networks):

A clear theme that emerged from the Integrated Research Team is the concern that U.S. researchers do not have the requisite training, facilities, tools and ICT infrastructure to adequately manage data generated from biomedical activities. Many researchers and clinicians lack the tools and skills to transform aggregate datasets into information, and information into knowledge. Knowledge grids focus on the integration of datasets from multiple disparate sources to achieve, in real time enabling streams of seemingly unrelated data from a variety of sources.

Knowledge Grids are an extension of the capabilities of computational and data grids to provide mechanisms to analyze and publish datasets for effective sharing and reuse. Leveraging computational services and data services, the knowledge services provide for full life-cycle management of biomedical data. Knowledge discovery methods ultimately reduce the dimensionality of complex biomedical datasets. These methods include data categorization (metadata), text mining tools, and interoperable semantics (ontologies). To be effectively integrated, published datasets that can be shared and reused require associated tools and community support.

A focus on biosurveillance and situational awareness drove to recognition of the need for integration of datasets to address the requirements of epidemiological networks. Maintaining data for long term manipulability for future researchers is a critical aspect of data custodianship and curation.[6] There is a need for organizational entities that can act as trusted repositories for the collection and storage of datasets, and provide operational support for public service platforms of both unclassified and classified knowledge. *(See Figure 1 - CDC Graphic on page 40)*

4. Collaboration Grids (focus: Collaborative Analysis, Visualization & Distributed Anatomic Modeling):

This area focuses on enabling communities of practice in virtual organizations. Cooperation by geographically dispersed individuals or groups, as they pursue common goals has led to the realization that virtual laboratories must be enabled for the remote control and management of ICT resources, equipment, sensors, instruments, and shared human interactions.[7]

An overall goal of the HealthGrid is to save lives via information-enabled via just-in-time strategic partnerships. Targets of opportunity include: mass casualty response teams, virtual trauma teams, as well as education and training. Collaboration Grids enable effective multi-user, real-time collaboration and data sharing while maintaining security. Collaboration grids are particularly focused on the aggregate of a great number and variety of actors (biosensors, shared instrumentation, medical devices, HPC, etc.).

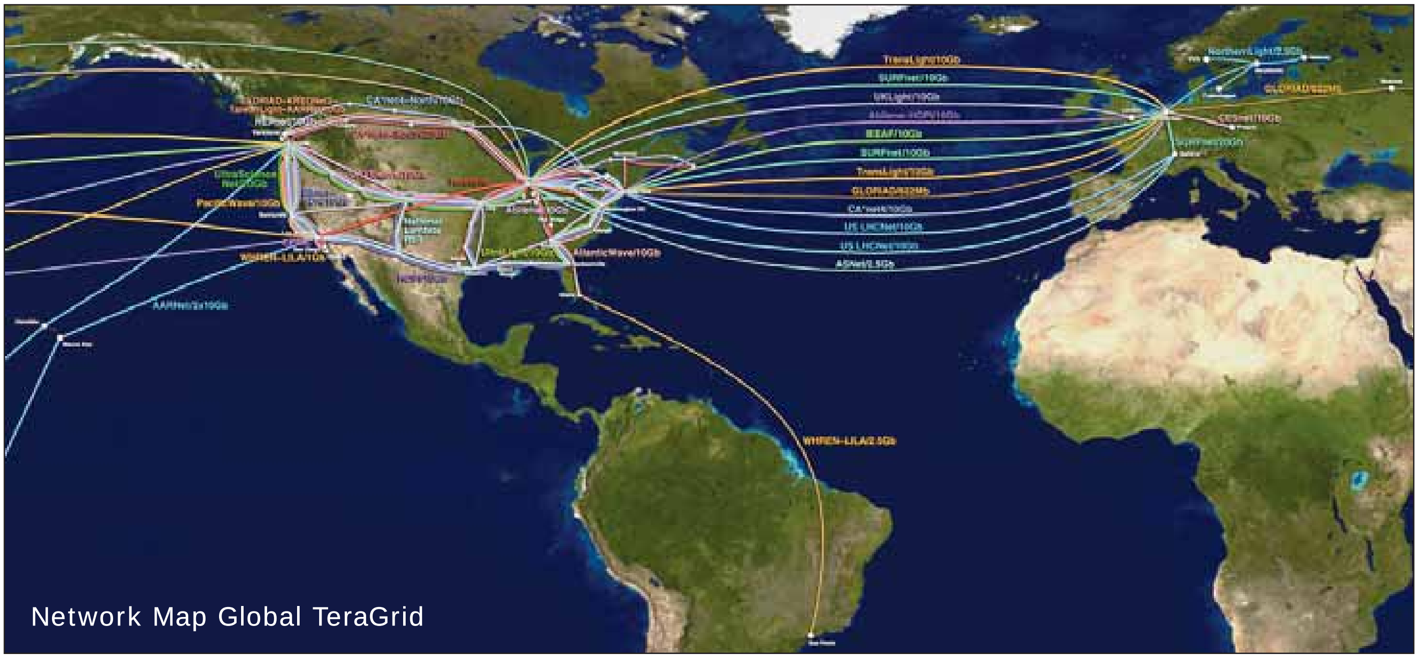
HEALTHGRID RESEARCH ROADMAP STRUCTURE AND TOPICS

A variety of technical, environmental and societal changes propel the development of the HealthGrid. Emerging from the IRT discussions was a catalogue of HealthGrid capabilities. Derivative research topics were identified and prioritized. The capabilities revealed below provide the basis for the formulation of the HealthGrid research roadmap:

- Workforce Transformation
- Scalable Information Infrastructure
- Components and Tools to support Public Service Platforms
- Virtual Organizations and Distributed Communities
- Knowledge Services

Workforce Transformation involves building human capacity through training and education programs. The HealthGrid research roadmap calls for the deployment of Virtual Organizations (VO) to allow scientists to enable unified gateways that provide researchers and the public access to the HealthGrid. A national workforce is necessary to build the HealthGrid that leverages our current infrastructure investments. VOs are the custodians of shared datasets, and curators of local resource aggregations based on expertise within a particular VO or region.

Photo Credits: Jonathan Silverstein MD, MS, FACS, University of Chicago Computation Institute



Network Map Global TeraGrid

Scalable Information Infrastructure (SII) is the creation and utilization of an operational technical infrastructure of high-capacity distributed computing. An investment in SII permits the development of advanced HealthGrid applications and delivery of software as services. SII primarily involves networks (telecommunications), computers (super computers, clusters, desktops), analytic equipment (instruments, sensors, medical devices, and shared equipment), data storage facilities, data center facilities and collaboration equipment (cameras, microphones, video teleconferencing). Ideally, SII provides the technical equipment and engineering necessary to deploy a VO. Grid standards are leveraged to enable deployment of a variety of Virtual Organizations for multiple communities of practice.

Components and Tools to support Public Service Platforms is the creation and maintenance of the information infrastructure and resource management. A Public Service Platform permits the creation and distribution of software for broad application to the benefit of the whole biomedical community. Examples include Open Source visualization tools for advanced distributed learning; application programming interfaces that allow seamless integration with transaction systems (such as HL7 applications); services to scheduling compute processing on shared resources (interactions with shared HPC infrastructure, such as a super computer); basic execution services to manage real-time workflows; and leveraging Web Services over a shared Grid resource framework. The Grid community has already accomplished much in the way of foundational standards, tools and grid service components. The medical sector can quickly leverage this work into biomedical Public Service Platforms that will benefit the broad community. One example is the need for trust authorities to address the various dimensions of identity management. There are many other needs such as a business intelligence platform for dashboards and forecasting tools that put the power of analytics in the hands of the user to reduce healthcare costs. There are many other examples of Public Service Platforms needed by the biomedical community, but an applied research program is necessary to pilot and deploy services.

Image © Scientific Computing and Imaging Institute at the University of Utah

Virtual Organization for Distributed Communities involves the socio-political support necessary to build efficient communities of practice. Organizations typically manage facilities and administrative support within one geographic location. The goal is to network organziatons into Virtual Organizations across national US HealthGrids. Driving forces include: Advanced Distributed Learning to better educate the US workforce; Public Health and Population Medicine to address biosurveillance; human behaviors that improve personal health; correlations to environmental and agricultural impacts to health; the knowledge network for translational research; and cyberinfrastructure resources necessary to bridge basic science research into the clinical domain.

Knowledge Services play a pivotal role in the creation of the knowledge society and need leadership to be achieved. Public-private partnerships between government and non-governmental entities are necessary to enable communities of common interest to maintain expertise, and analyze digital assets. Promoting collaborations that drive the formation of shared knowledge collections rests optimally on a foundation of shared services. Shared collections of knowledge must be managed and monitored effectively. Sustainability and governance of the HealthGrid requires long-term support and foresight to address management challenges and availability of shared data assets and analytic tools. The Knowledge Society is a global community and the HealthGrid is a vital enabling actor in these global efforts.

HEALTHGRID CURRENT STATUS AND OPPORTUNITIES

A central theme for Grid computing in biomedicine was established by the European Union programs on HealthGrid to enable resource delivery through distributed computing technology.[8]

"Although the ultimate goal for e-Health in Europe would be the creation of a single HealthGrid, i.e. a Grid comprising all eHealth resources, naturally including security and authorization features to handle subsidiarity of independent nodes of the HealthGrid, the development path will most likely include a set of specific HealthGrids with perhaps rudimentary inter-Grid interaction/interoperational capabilities. HealthGrid applications are oriented to both individualized healthcare and epidemiology analysis." [9]

The Orange Group - Photo Credit: J. Harvey Magee

The Grid computing community supports a culture of early technology adoption and Open Source sharing of new processes and methods that catalyze outcomes. Knowledge thrives on the free exchange of ideas. Within the healthcare culture, however, new processes are adopted slowly and methodically. Several factors in healthcare account for this, to include concerns for patient safety, high switching costs for providers and reluctance by physicians to adopt new and difficult to understand technologies. Medical knowledge doubles every 3.5 years,[10] and change can take time to embed into the system. A recent study showed an average 17 years delay in adopting widely proven practices in healthcare in the US.[11] Biomedical research provides a bridge between Grid technologies and clinical applications. Rapid technology adoption and innovation in concert with stepwise conservative measures and validation based on standardized peer review processes has great potential.

Collections are important. The Armed Forces Institute of Pathology (AFIP) tissue collection is the 'finest' and unique collection of tissue content available today. Multiple genomes have been sequenced and annotated to varying degrees, resulting in the appearance of substantial functional genomic data collections curated by the National Library of Medicine. Repositories of experimental proteomics and functional genomics results are also being supported. Sharing of all these Federally supported data collections in ways that permit cross correlation of collections to new workflows and algorithms will enable a variety of scientific communities to further knowledge of genomics, proteomics and metabolic pathways. Grid Computing serves as a mechanism to enable data sharing for communities of interest for biomedical research, in turn forging scientific discoveries & breakthroughs.

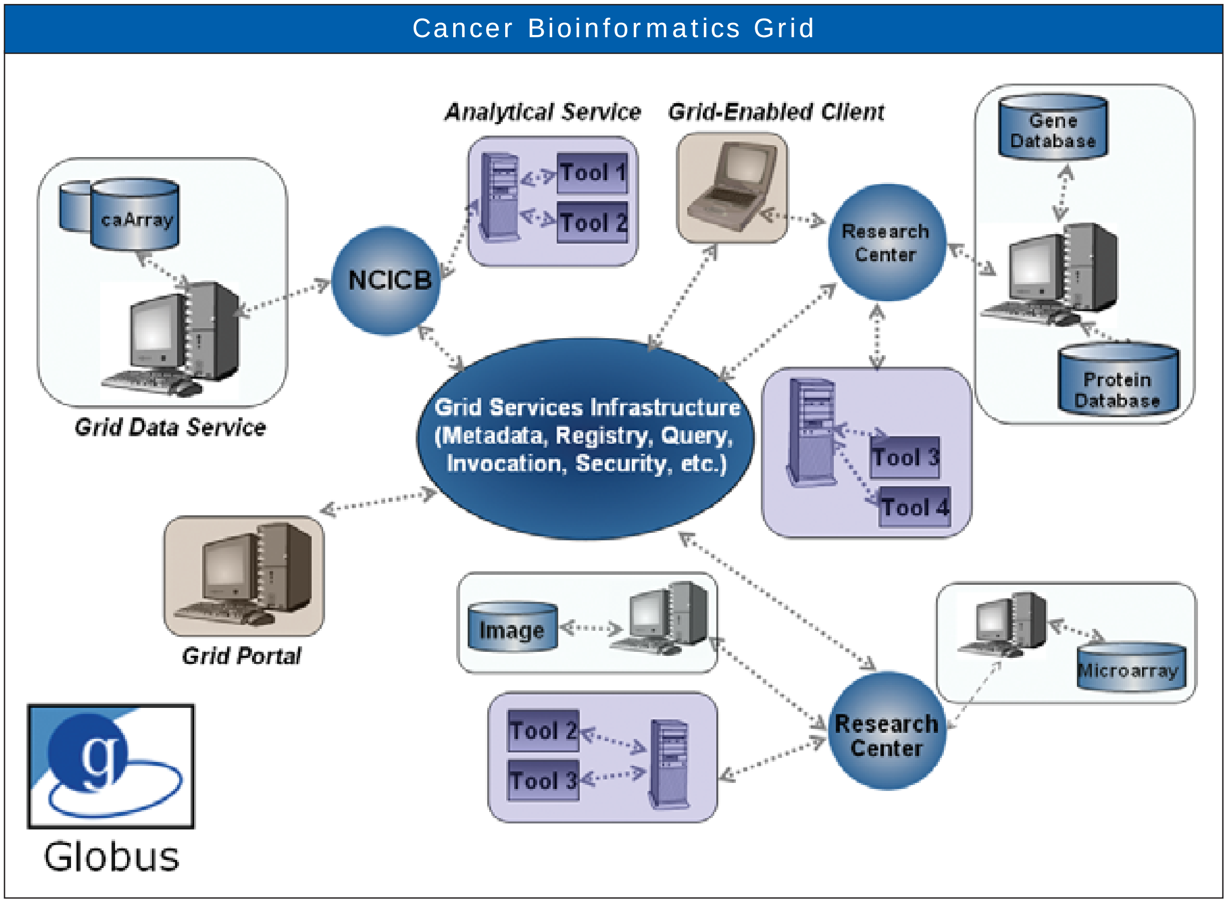
HealthGrid addresses the following:

- The needs for multiple laboratories to collect data for analysis in up-to-date, shared, but competitive environments
- The large computational 'horsepower' required to scale biology and medicine beyond an individual hospital or academic center or laboratory
- Easy accessibility to datasets and collaboration and analysis tools for scientists, medical staff and consumers
- Usability of information (molecular, personal, population based) and knowledge tools necessary to understand and scale knowledge to a wide variety of communities

As clinician and patient demand for knowledge-driven information grows in urgency, the priority for advancing interoperability between biomedical research and clinical practice is recognized by leaders across government, industry and academia. A common approach is necessary to effectively address the benefits of applying ICT to biomedicine. HealthGrid provides this common ICT fabric for the future knowledge society.

BARRIERS AND RESEARCH CHALLENGES

Today, few biomedical applications are built to function within the Grid Computing paradigm. The capacity of Grids for biomedicine is thus presently much underutilized. What are the factors that limit the adoption of Grid computing in biomedicine? First, users are familiar with batch transaction systems, whereas the interfaces to Grid software are often complex and different from those in batch systems. Second, users are used to having transparent file access at the local desktop, which Grid software does not conveniently provide. Third, efforts to achieve widespread coordination of computers while solving the first two problems is hampered when clusters are on private networks. What is needed is a variety of software that allows users to transparently use Grid resources as if they were local resources while providing transient access to files within a virtual organization, even when private networks intervene.[12]



Current health related computer applications lack semantic integration. This negatively affects the adoption of Grid platforms, particularly in clinical settings, where data content is not mediated across a consistent context.

While there are many biomedical applications that could leverage the high-throughput approach of distributed computing, the added level of service provided by Grids is not typically understood. The Grid approach stands foreign to today's normal application development practices and resulting operational environments. Potential uses in the clinical practice of medicine are today largely untapped. Further study is required to assess the full impact and potential of the HealthGrid.

Although some large research centers do provide computing resources for systems biology, most users must make a substantial effort to maintain redundant collections of algorithms and biological databases on local machines or use dedicated web portals, each with specific, limited functionality. Virtual Organizations are not the norm for today's enterprise based corporate administrative structures.

As these issues are resolved, a dramatic expansion of the HealthGrid's utility and function is envisioned, in turn changing the practice of medicine and biomedical research. The HealthGrid, consonant with the economics of utility-based computing, helps to shift ICT budgets from capital to commodity expenditures.

PUTTING FEDERAL RESEARCH INTO ACTION

"…Grids are the integrated platforms for all network-distributed applications or services whether they are computationally or transactionally intensive." [13]

Current US Government priorities as seen through the National Institutes of Health Roadmap call for advancing collaboration in biomedical research and using biomedical data and information to improve the quality and outcomes of health care delivery.

President George W Bush has spoken clearly in setting a goal to make an electronic health record available for most Americans by 2014 and to develop a Nationwide Health Information Network under the leadership of the U.S. Department of Health and Human Services. One could not ask for a more propitious occasion to collaborate on advanced HealthGrid projects, showcasing the possibilities for and the power of interagency sharing, open solutions and innovation. The grid community can contribute significantly to advancing heathcare quality to achieve higher levels of performance and interoperability, while reducing costs and increasing patient safety.

This report is the first evolution towards a high-level secular plan to facilitate communications across intergovernmental agencies and organizations via exploiting the advantages of shared computational power. An environment of open collaboration should drive further actions toward connecting US Government agencies with private and public organizations, ultimately leading to a more strategic utilization of national resources.

The establishment of an action plan for inter-agency cooperation and identification of mutual interests between and across current programs will enable productive use of cyberinfrastructure investments and inter-agency collaborations to develop and sustain US cyberinfrastructure for biomedicine. Specific targets include:

- Data movement across networks, collection, annotation and provenance
- Training; human capacity building on grid technologies and HealthGrid
- Visualization
- Simulation Models
- Utilization of computational grids (Teragrid and the Petascale Facility) in biomedicine
- Situational Awareness
- Systems Biology
- Translational Research
- Knowledge tools
- Making grid services real across domains

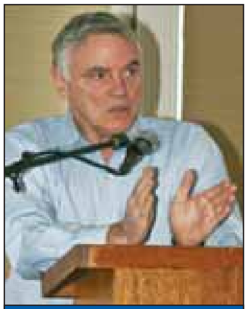
The goal for expanding communities of specialized practices must be to support efficient cyberinfrastructure environments for scientific collaborations and effective knowledge sharing across biomedicine, the healthcare industry and other domains.

The panel recommends the formation of a HealthGrid Core Strategic Planning Group representing diverse areas of Grid technology across the U.S. Government. Participation from key agencies including DoD [TATRC] (applied research); DoD [Health Affairs] (electronic health record AHLTA data availability for public good and architecture); Veterans Health

Administration (SOA activity and VistA); National Science Foundation Office of CyberInfrastructure; National Institutes of Health; and NASA are essential. The following is a brief list of areas within the biomedical domain where Grid technology is being used.

- Genetic Linkage Analysis
- Molecular Sequence Analysis
- Determination of Protein Structures
- Identification of Genes and Regulatory Patterns
- Biological Information Retrieval
- Biomedical Modeling and Simulation
- Biomedical Image Processing and Analysis
- Data Mining and Visualization of Biomedical Data
- Text Mining of Medical Information Bases

FUTURE CONSIDERATIONS


Leroy Hood, M.D., Ph.D.
Photo Credit:
J. Harvey Magee

Adoption of HealthGrid principles and practices is not a trivial undertaking. Sharing of clinical data has many policy and legal implications related to intellectual property and data ownership that need to be addressed. On the one hand, some of the most active users of the TeraGrid are computational biologists, so to some degree high performance computing in general is moving rapidly from a focus on the physical to the biological sciences. Tracking the high performance computing community will provide a good indication of use. On the other hand, there is an enormous gap between the capabilities and use of HealthGrid principles and practices between advanced research communities as compared to individual researchers and clinicians. Establishing a broad base of assessment activities to appreciate the adoption of grid technologies in biomedicine is needed:

- Establishment of a workgroup to identify functional requirements and/or potential uses of Grid systems for use by the different U.S. Government agencies
- Identify potential organizations in the private sector to collaborate on the research, development, testing and use of
- Grids in healthcare, e.g. industry and professional societies
- Conduct a feasibility study into the use of grids
- Investigate changes in the business economics and ICT workflow processes that may need to be made in anticipation of utilizing grid technology
- Initiate and fund pilot project(s) and complete a detailed cost benefit analysis
- Coordinate HealthGrid activities through NITRD, Subcommittee on Large Scale Networking and Information Technology, Middleware and Grid Infrastructure Coordination (MAGIC) Team.

One limiting factor for HealthGrid remains the infrastructure of networks and computational resources available for high quality and reliable operational systems. In the context of biomedicine, this limitation is now largely a question of investment and provisioning management. One aspect to address for biomedical application is intermittent high demand users. Today much capacity is wasted without a coherent shared provisioning strategy. Highly reliable, fast, high quality networks can be built, but require attention in design toward specific use cases and execution so as not to let them deteriorate to lowest common performance based on the actions of poor local choices (a variant of Gresham's Law). Resources should focus on supporting efforts to provision networks and other ICT infrastructure at the highest level available. Conceivably, military and biomedical grade networks would have similar performance characteristics, for scalability across the spectrum on biomedical environments.

Additional efforts should be applied toward creating data stores that offer the possibility for diverse application areas to leverage Public Service Platforms. While there are existing applications, the choice of which projects to fund should include criteria for those that create data and make it available to future applications.

It is important to focus research on questions that connect directly to compelling social and economic models of the future knowledge society. These needs drive industry adoption and technology transfer from the research arenas into the marketplace. There is a tendency in medicine to reinvent the wheel. We should acknowledge the existence of powerful tools and insist on interaction with the developers of those tools to adopt them for biomedical applications.

It is important to early-on establish a management structure to move the process towards achieving the over-arching goals. In addition to dedicated government personnel, it is necessary to establish some stable organizational infrastructure for the US to participate in the global HealthGrid community. Exchange and transfer of ideas are best facilitated via meetings and collaborative interactions with professional societies. A global HealthGrid office now exists in Europe, with associated organizational constructs necessary to build the global HealthGrid. The US can and should play a pivotal role in the global HealthGrid community.

Photo Credit: Parvati Dev, PhD,
Stanford University Medical Media and Information Technology (SUMMIT)

HEALTHGRID ROADMAP

Category Of Capabilities	Goals	1-3 Year	3-5 Year	5-10 Year	Priority
Workforce Transformation	Increase Human Capacity and build awareness of Grid model. Create communities of practice through sustainable workforce training programs.				
	Human Capacity Building on Grid Technologies for Biomedical sector via training and education programs	U.S. Workforce Development Program to build HealthGrid competencies for 2-5 sponsored Virtual Organizations and fellowships/post doc positions	U.S. Workforce Development Program to build HealthGrid competency for 5-10 Virtual Organizations sponsored fellowships/ post doc positions	U.S. Workforce Development Program for HealthGrid competency for 10-20+ Virtual Organizations sponsored fellowships/ post doc positions	High
	Establish the US as a leader in global Grid technologies and services.	HealthGrid Training Program for 2-5 workshops	HealthGrid Training Program for 5-10 workshops	HealthGrid Training Program for 10-20+ workshops	High
	Contribute to the scientific, technological and subsequently economic competitiveness of the US globally.	Support development and deployment of HealthGrids for Biomedical Sector and contributions to EGEE	Demonstration of how grid technologies enable analysis and data management	Support HealthGrid Public Service Platforms and contributions to Cyberinfrastructure/ eScience	High
	Strengthen US economic and social cohesion with global partners.	Visiting Scientists Program for 2-5 sponsored participants in US HealthGrid projects	Visiting Scientists Program for 5-10 sponsored participants in US HealthGrid projects	Visiting Scientists Program for 10-20+ sponsored participants in US HealthGrid projects	Medium
BUDGET	**Total: $10m**	**Total: $2m**	**Total: $5m**	**Total: $3m**	
Scalable Information Infrastructure	Create and maintain up-to-date, high-capacity distributed computing technical infrastructure permitting the development of advanced applications and services for research, education and clinical care.				
	Scalable Network infrastructure	Advance national research networks Network backbones (Internet2/NLR; PHIN, NECTAR) Regional and local networks (RONs, RHIOs, First Mile, etc.)	Enable Network connections between DREN and national networks Optical Technology research, such as logical routing (content based routing) WAN to rural areas and international peers	Scale Network technology advances	High
	Computational Grids	Increase computational capacity for the US Super Computers and Compute clusters	Scale up computational capacity for parallel and vector processing	Scale up deployment of computational infrastructure	Medium
	Analytical Output	Grid enable Scientific Instrumentation and sensors	Identification of middleware services needed to access and use Grid resources	Scale up deployment of analytic capabilities	Low

Category Of Capabilities	Goals	1-3 Year	3-5 Year	5-10 Year	Priority
	Data Storage Grids	Install generic infrastructure to coordinate federation of distributed data repositories	Scale up deployment of distributed storage	Scale up deployment of access and storage capabilities	High
	Develop vigorous SBIR Program to enhance the HealthGRID	1-3 SBIR topics for Military medicine	3-5 SBIR topics for Military medicine	3-5 SBIR topics for Military medicine	Medium
BUDGET	**Total: $15m**	**Total: $5m**	**Total: $5m**	**Total: $5m**	
Components and Tools to support Public Service Platforms	Create and maintain information infrastructure for resource management, which permits the development of advanced applications for research, education and clinical care.				
	Visualization Platform	Create and support toolkits that allow multiple users to control computation and virtual collaborative environments for real-time and near-time applications.			High
	Open Source Tool Development Platform	Open Grid Services Architecture (OGSA) development support	Development support for additional OGSA tools/services	Development support for additional OGSA tools/services	High
		Simple API for Grid Applications (SAGA)			
		Job Submission Description Language (JSDL)			
		Basic Execution Services (BES)			
	Web Services Resource Framework WSRF	Conventions for application discovery, inspection, and interaction with stateful resources in standard and interoperable ways.	Development of registries for structured data to increase value for the community		Medium
	Data Analysis Business Intelligence Platform	Analytics Framework for dashboards, forecasting, data mining and other analysis capabilities			Medium
	Security Services	Interoperable security services for PHI and, to address the "core" use cases for scientific/technical applications			Medium

Category Of Capabilities	Goals	1-3 Year	3-5 Year	5-10 Year	Priority
	Develop vigorous SBIR Program to enhance the HealthGRID	1-3 SBIR topics for Military medicine	3-5 SBIR topics for Military medicine	5+ SBIR topics for Military medicine	High
BUDGET	**TOTAL: $ 30 m**	**$10 m**	**$10 m**	**$10 m**	
Virtual Organizations for Distributed Communities	Build Communities of Practice by supporting the socio-political and technical engineering for operational HealthGrid deployments of National Virtual Observatory applications.				
	Advanced Distributed Learning	Enable general anatomic modeling in support of health applications Enable grid computing systems to apply anatomic information in a rational, systematic way based on standards and supporting interoperability.	Mulit-scale models framework for anatomy Define how to evaluate results of research for its value in medical practice, and how to translate research knowledge into actual practice	Interoprable modeling Define benchmarks for virtual organization testbeds Automation of Anatomy System	
		Framework for Dynamic Creation of Collaboration Grids Create a software support infrastructure that makes creating teams and virtual organizations simple and easy.	Development of collaboration testbeds	Understand how collaborative environments change workflow and outputs	
		Create and support a set of data visualization tools that allow multiple users to view and analyze, and manipulate data in real time.	General availability of one or more set of grid-enableD tools for real time collaborative data visualization	Sustained use of visualizatino tools in 2 or more biomedical domains other than the original domain	
	Public Health and Population Medicine	Enable grid computing systems to apply public health information in a rational, systematic way based on standards and supporting interoperability	Framework for Dynamic Creation of Epidemiology Grid Enable general public health sensor networks and communications over collaboration grid systems	Enable grid computing systems to apply public health information in a rational, systematic way based on standards and supporting interoperability.	
		Create a support infrastructure for an Infectious Disease weather map that makes data collection, structure and aggregation possible in real-time			

Category Of Capabilities	Goals	1-3 Year	3-5 Year	5-10 Year	Priority
	Biocomputation and the Life Sciences	Work with USG agencies to create a community of researchers conducting real time collaborative biocomputation	Create a community of researchers using real-time collaborative simulations for training	Build awareness of this new computational model.	
	Clinical Transformation and Interoperability	Enable use of research results in multiple care sites including clinical trials, image processing, treatment planning, personalized, and community based systems medicine	Support discovery and access to data from health clinics to enable research. Support the clinical research forums	Enable use of research results in private practice	
		Partner with professional societies to develop Computer-assisted Clinical Decision Support	Build CDS models and testbeds for military programs	Deploy CDS for military programs, for example TBI/PTSD	
	Develop vigorous SBIR Program to enhance the HealthGRID	1-3 SBIR topics for Military medicine	3-5 SBIR topics for Military medicine	5+ SBIR topics for Military medicine	High
BUDGET	**TOTAL: $30 m**	**$10 m**	**$10 m**	**$10 m**	
Knowledge Services	Provide a pivotal role in the creation of Knowledge Society and provide US leadership for HealthGrid deployments to enable systems medicine. Establish Collaborations with other Government and nongovernmental entities.				
	Assets	Human Capacity Building via the development of communities of common interest that will maintain the expertise needed to interpret and analyze the digital holdings	Address HealthGrid Data Challenges by focused efforts across USG agencies	Develop groups of expertise that will drive the research agenda	
	Advancement	Ontology models and Terminology Mediation	Promote collaborations that drive the formation of shared collections Promote interactions with USG agencies to encourage use of research results	HealthGrid Monitoring and Manageability Challenges	
	Foresightedness	Make structured data available for analysis via VO Promote international collaborations based on open source and virtual organizations	Address Virtual Organization Management Challenges	Promote sustainability and governance mechanisms for long-term support	
BUDGET	**TOTAL: $15 m**	**$5 m**	**$5 m**	**$5 m**	

CONCLUSION

Panel members are participants and technology enablers in today's complex environment of biomedical research and healthcare delivery. We view the current ecosystem as a 'perfect storm' where the front of escalating demand for equitable services collides with that of ever-tightening constraints on the resources to provide these services. We conclude that exploitation of the profound opportunities presented by computation, applied as HealthGrids, is a critical component in harnessing this storm. Ignoring these inescapable technology trends incurs a substantial opportunity cost. Instead, a prudent investment strategy by government, industry, and the academy in the pursuit of public health, should reap dividends for decades to come.

The Red Group - Photo Credit: J. Harvey Magee

APPENDICES

FORWARD TO THE APPENDICES

Over the course of the two-day meeting, leading technical, medical and information systems experts from industry, academia and government assembled to discuss the vision for a HealthGrid research roadmap to form the baseline for success metrics and foresights into the future knowledge society.

To achieve this goal, four Working Groups were established from all participants in the IRT. The Working Group participants where balanced to include representation of government agencies, industry and academia. The Working Group participation was also balanced to provide equal subject matter expertise from biomedical sector and computer science/engineering sectors.

The Working Groups where each assigned one content area:

- Computational Grids with a focus on the Life Sciences
- Data Grids with a focus on Translational Research
- Knowledge Grids with a focus on Epidemiological Networks
- Collaboration Grids with a focus on Distributed Anatomic Modeling

Each group was tasked with a discussion format.

INTRODUCTION and BACKGROUND:

- *Defining the HealthGrid*
  - *Clarify the concept of HealthGrid*
- *Relevant background and lessons learned from history/experience*
- *Outline uture directions*

SERIES OF SPECIFIC OPPORTUNITIES

- *Delineation of HealthGrid driver*
  - *Identify biomedical issues that could be better addressed TODAY using GRID technologies*
- *Timescale for opportunities; Using the HealthGrid drivers as a starting point, create a roadmap to the future*
  - *Where are we today? Where do we need to be tomorrow (5-10 years future)?*
  - *How do we get there?*
- *Analysis of technical challenges*
- *Scale/Scope/Importance/Prioritization*
- *Budget parameters*
- *Don't want to miss any big opportunities, but need to articulate what is available*
- *Recommendations for government leadership*

STRATEGIC PARTNERSHIPS

- *The Nature of Collaborative Relationships*
- *The biomedical landscape*
- *Cross-disciplinary research opportunities*
- *Learning to speak the language*
  - *Of medicine and biology*
  - *Of Grid/Cyberinfrastructure*

TRIPLE HELIX/Public Private Partnerships

- *Virtual Organizations to Collaborative Federal Repositories of knowledge*
  - *Strategies from Academia*
  - *Strategies from Industry*
  - *How can government/funding agencies facilitate the HealthGrid?*
- *Creation of new markets in the knowledge economy*

EVALUATING OUTCOMES; ASSESSMENT PARAMETERS FOR HEALTHGRID

- *Measurements*
- *Evaluation criteria*

CONCLUSIONS: FINDINGS AND RECOMMENDATIONS

REFERENCES

The attached appendices contain the information aggregate collected and summarized by each Working Group. The information collected was analyzed, scrutinized and amalgamated into the HealthGrid research roadmap detailed in the previous section of this report.

LIFE SCIENCE COMPUTATIONAL FOCUS GROUP

Group Leaders:

Leroy Hood and Carl Kesselman

Group Members:

Brian Athey
Joel Bacquet
Ken Beutow
Howard Bilofsky
James Cassett
Timothy Clark
Don Detmer
Robert Foster
Caroline Kovac
Yannick Legre
Wilfred Li
Sidney McNairy
Hon Pak
Beth Plale
Joel Saltz
Wayne Wilson

BACKGROUND

Life sciences research is increasingly focused on genomics and systems biology, both of which are increasingly dependent upon high performance computation. Today, via the TeraGrid[14] and other grids, virtual clusters of computers dynamically aggregate the power of thousands of individual computers in order to provide a platform for advanced high-performance and/or high-throughput applications that could not be tackled by a single system. This synergy between the scaling up of life science research and computation enables a qualitative change in the way future medicine may be perceived and practiced.

One view of a fully realized health grid is the enablement of so-called systems medicine: predicative, preventive, personalized, and participatory. In a typical scenario, a person's genome may be quickly sequenced using nano-sequencers, the markers for different genetic loci may be analyzed using haplotype analysis, and the results form the basis of personalized medicine.[15] After personal information is associated with clinical phenotype (via data grid techniques), it is possible to enable predictive medicine (via grid-based analytics). Once predictions may be made, then preventative medicine may be used to correct a person's deficiencies through mechanisms such as gene therapies, gene replacements, etc. With this information in hand, a patient becomes an active participant in this medical treatment, through collaboration with the physicians by providing realtime feedback (via collaboration grids), and contributes back to the knowledge grid indirectly.

Current approaches to the representation of disease states in systems biology[16] include the use of networks and perturbations to help understand the progression of diseases such as prion disease, which is a "rare progressive neurodegenerative disorders that affect both humans and animals." This type of systems based approach to disease analysis may lead to more comprehensive finger print analysis libraries. These may be diagnostic of diseases based on markers or sets of markers with comprehensive context information. In fact, preliminary studies have shown that blood tests for organ specific secreted transcripts are possible - organ specific blood finger prints. The collection of blood samples is a routine procedure during routine healthcare and therapies, and hence rapid development of this diagnostic tool is important. Challenges include: development of the ability to make noninvasive measurements at the single cell level; the ability to overcome the cell type, developmental stage and other heterogeneities; the development of a reproducible systems network; and the development of diagnostic markers.

The availability of all data on all parts in the genome (genes) makes possible the modeling and simulations at the systems level based on experimental data. Predictions can be made that may change how we represent healthy and diseased states and how we develop and test interventions. As parallels are being drawn between complex systems in engineering (cars and airplanes) and biology (genes and humans), it is possible to transfer the tools from engineering and apply them to systems biology.[17] The successful modeling of different levels of activity from the level of protein functions, to the organ level, to clinical observations of electrocardiogram (ECG) is one proof that such activities are feasible and suggest they are essential to create the future of predictive medicine.[18]

Several efforts are being funded by other agencies, especially the National Institutes of Health, which complement, and indeed, provide essential components for the HealthGrid.

- The Biomedical Informatics Research Network (BIRN) is funded by NCRR to build, maintain and mature a national information infrastructure to enable and advance biomedical research. One of their experiments in the Morphometry BIRN is to support diagnostics using high performance computing through morphometric analysis to correlate human neuroanatomical data with clinical data. Such analyses may be used to diagnose diseases such as Unipolar depression, Alzheimers's, mild cognitive impairment.

- The MIDAS (Models of Disease Agent Study)[19] project is supported by NIGMS to model infectious disease agents to help make decisions by developing mathematical models of the infectious diseases.

- The National Centers for Biomedical Computing (NCBCs) grew out of a report of a working group of the Advisory Committee to the Director of NIH issued in 1999.[20] The first recommendation of that report was the NIH should establish between five and twenty National Programs of Excellence in Biomedical Computing devoted to all facets of this emerging discipline, from the basic research to the tools to do the work. It is the expectation that those National Programs will play a major role in educating biomedical-computation researchers. This recommendation was finally satisfied when the establishment of these centers became one of the initial activities of the NIH Roadmap.[21] Four centers were funded in Fiscal Year 2005 and 3 more in Fiscal Year 2006. Activities range from the application of the principles of Physics to biological problems to the solution of the problem of bringing advanced biomedical techniques to the bedside.

- National Institute of General Medical Sciences (NIGMS) has funded five "glue" grants over the past several years. The goal of the glue grants is to solve an overarching problem in biomedical research through collaboration in ways that would be impossible to achieve by research groups working alone. A significant part of each of these efforts is a large computational/ informatics core. The topics range from developing methodology to study cell motility to finding markers to predict outcomes in burn and trauma patients.

- A key initiative aimed at personalized medicine is the Pharmacogenetics Research Network: http://www.nigms.nih.gov/Initiatives/PGRN/. The goal is to understand the genetic basis of the varied response to pharmaceutics. The network is linked by a database, PharmGKB, http://www.pharmgkb.org/network/members/pharmgkb.jsp.

- The National Cancer Institute's cancer Biomedical Informatics Grid, or caBIG™, is a voluntary network or grid connecting individuals and institutions to enable the sharing of data and tools, creating a World Wide Web of cancer research. The goal is to speed the delivery of innovative approaches for the prevention and treatment of cancer. NCI funding is supporting the development of core caBIG infrastructure, which includes standard data models and a caGrid platform, based on Globus technology.

While the use of high performance computational grids is approaching the normal state for life sciences research, the state of use of grids in the health care information technology space is lagging far behind with few examples of grid based storage networks and few if any computational grids being significantly utilized for processing. Pushing toward creating the future of P4 medicine will tend to close the gap and is one viable approach toward advancing information technology in health care - from the side of the biomedical spectrum more research focused, moving technology toward the side of the biomedical spectrum more operationally focused.

BARRIERS AND RESEARCH CHALLENGES

The success of P4 medicine is critically dependent on bringing the information typically obtained from bioinformatics to the bedside. Typically this would include genetic and gene expression data. Such data obtained on the individual patient will be essential to therapies.

Meeting this overall objective will require focus in several areas:

Data sharing: Mechanisms are in place for sharing genetic data. Similar mechanisms need to be established for sharing clinical data. Importantly, as we begin to understand the relationship between genetic variation and disease outcomes it becomes necessary to share data for individuals. Necessary safeguards will have to be developed to ensure patient privacy. The issue is exacerbated because genetic data, themselves, are sufficient to reveal the identity of a patient.

Development of knowledgebases to link genetic and medical data: It is essential to capture not only the data themselves but also the knowledge that is obtained from the mining and analysis of these data.

Ontologies: Scientists and physicians from different disciplines and subdisciplines often use different terms to mean the same thing. Even scientists within a specific subdiscipline may use different terms depending on different traditions, education, and experience in a field. Hence the development of widely accepted ontologies is essential to join the fields of medical and bioinformatics. Much work has already been done on ontologies, and it is essential not to reinvent the wheel, even if existing ontologies may have problems.

Development of easy-to-use software: The scientists and physicians who will have to make the links between medical and bioinformatics, are unlikely to be trained computer scientists, nor will they have the time or patience to wade through cumbersome, non-intuitive interfaces. Hence considerable thought has to be given to the human computer interface.

FUTURE CONSIDERATIONS

GOAL: A fully realized goal of the Health GRID is systems medicine, predictive, preventive, personalized and participatory.

Objective: (1-3 years)

Establish a management structure to ensure the future of the HealthGrid and map strategies for the future

- Task: In addition, to assigned government personnel, establish an office to coordinate the collaborations and fund necessary meetings and workshops.
- Task: Establish an Advisory Committee to provide advice to the government on an on-going basis. This committee would meet annually. The membership of this committee would come from representatives of the triple-helix, government, industry and academia. In addition, because of the need for "participation" of the patient, lay membership would be essential. Such lay membership could be from patient advocacy groups. In addition, it is essential that the membership of this committee reflect the diversity of this nation.

Objective: (1-3 years)

Establish Collaborations with other Government and non-government entities working and funding components that might be incorporated into the HealthGrid.

- Task: Align the HealthGrid with the NIH Roadmap in appropriate areas
- Task: Coordinate HealthGrid activities with other similar efforts

Objective (1-3 years)

Coordinate the development of Ontologies that will facilitate the use of the HealthGrid

- Task: Coordinate the development of Ontologies that will facilitate the use of the HealthGrid
- Task: Identify gaps and working with other interested federal agencies, develop funding mechanism to fill those gaps.

Objective (3-5 years)

Develop a vigorous SBIR Program to develop software that enhance the HealthGrid.

- Task: Since TATRC funds for SBIR grants are limited, develop collaborations with other federal agencies (principally NIH) with a greater capacity for these grants.
- Task: With input from the advisory committee develop and enhance topics that will be especially suitable for these programs.

TRANSLATIONAL RESEARCH USING DATA GRID TECHNOLOGY

Group Leaders:
Reagan Moore and Marc Wine

Group Members:
Jack Buchanan
Dave Damassa
Don Detmer
David Forslund
Ray Hookway
Mary Kratz
Phillip LaJoie
Leon Moore
Ron Pace
Ken Rubin
Stanley Saiki
Michael Sayre
Jonathan Silverstein
Anil Srivastava
John Wilbanks

BACKGROUND

A data grid is an implementation of the generic software infrastructure that organizes and manages data for grid computing applications. Today, data grids manage shared collections that span multiple institutions and that are international in scope. Grid software infrastructure is middleware in that it is neither application nor basic resources like databases or computer processors. A data grid uses the identical grid middleware to other grids, with other areas of focus, but focuses specifically on management of massive and/or distributed data collections. Within the healthcare arena, HIPPA patient confidentiality requirements require the authentication, authorization, and auditing services typical of grid middleware. Specifically, data grids support controls on access to data and metadata, access audit trails, encryption of data except when decrypted by the end client, and data storage specifications. The production systems that support shared collections include Petabytes of data and hundred of millions of files, similar to those expected for medical information collected across health grids.

BARRIERS AND RESEARCH CHALLENGES

Most data grids are tailored to specific groups of people working on shared problems. These groups establish standards (sometimes formally, other times in de-facto systems) that enable effective data sharing. These standards typically cover:

- Semantics - standard attribute names to describe physical quantities, which are used to support browsing and discovery
- Data encoding formats - standard structures for the data sets that are deposited into the shared collection
- Services - presentation and analysis mechanisms for parsing the standard data encoding formats using the standard semantic labels.

For most application areas, the data in question grow and change over time - often very rapidly. As new knowledge is generated over time, the standards for how each data grid operates must adapt to support the new data requirements. In response to this dynamic environment, data grids must exhibit the following characteristics:

- They must employ extensible schema that can accommodate the creation of new semantic terms.
- The data management system must be flexible in order to adapt to the development of new encoding formats handling new types of data.
- The data grids must incorporate new access methods as new services are created.
- It must be possible to move the data collection onto new vendor products as new storage systems are developed and obsolete technology is replaced by more cost effective products.

A generic translational health grid environment that meets the above requirements must:

- support the formation of data grids
- publish data in the digital libraries that provide the access services
- preserve the data in persistent archives.

Substantial collections of medical records and bioinformatics research data have already been created by the medical and research communities. These resources are distributed across geographic sites, institutions, and federal agencies. Often these communities create standards for data semantics, encoding and services independently.

In practice, applications frequently must access and apply data from different communities. Therefore, different standards need to be integrated for the applications to function. Data federation is the concept of allowing two independent data sources to share name spaces so that they can both be applied in the same application. Federation can work is several different ways:

- Peer to peer data grids. Researchers are able to read and publish over each other's data.
- Master-slave grids. A central data grid serves as the authoritative source for all registered data, which is organized into a single overall conceptual or executable (semantic) framework. Slaves receive and distribute data to and from the master grid.
- Central archive data grid. Peer data sources deposit their data into a central archive. No central management paradigm exists such that data remains semantically unconnected among sources.
- Hub-spoke data grids. Peer to peer data sources register only their name spaces into a federation hub, not the details of their data. Researchers can then access all of the peer data through authorization by the hub. Peer data grids only establish a trust relationship once with the hub data grid, so administration is simplified.

Shared collections allow the authoritative data source to remain at the original institution. The data grid is a logical data collection that provides unifying name spaces across institutions. Standard services applied on top of the data grid enable access to the original data sources, no matter what storage technology has been selected by a particular site.

There are several groups currently providing grid infrastructure:

- The digital library community
  - Metadata Encoding Transmission Standard (METS) for encapsulating and structuring metadata
  - Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for transporting data
  - Dublin Core for explicit metadata schema

- The preservation community
  - Open Archival Information Standard (OAIS)
  - Archival Information Package (AIP) for mechanisms to encapsulate metadata and data for long term preservation
  - Producer Archive Submission Pipelines for the authoritative import of data into a shared collection or preservation archive.

The medical and research communities already have created substantial collections of medical records and bio-informatics research data. These resources are distributed across geographic sites, institutions, and federal agencies. The creation of an authoritative collection can be approached from two perspectives, both of which recognize that the actual data records remain distributed and controlled by the originating institutions:

- Mediation (archiving approach), linking independent data resources. Explicit mediators are developed to map from the semantics and access mechanisms used by a particular database to the interoperability mechanisms used by the mediation environment. This enables discovery across independent collection, requires the registration of the independent resources into a registry, and enables creation of standard portals such as the bio-informatics workbench for providing uniform services across the independent collections.

- Shared collection (master-slave approach), explicit registration of selected data into a common collection. Application level servers are installed at each site where data resides to map from the data grid protocol to the local storage protocol, and to support mapping from the data grid logical names to the local physical file names. This approach enables the controlled sharing mandated by HIPAA patient confidentiality. This also enables the use of standard semantics within the shared collection (independently of the semantics used by the local resource), and the porting of standard services on top of the data grid. The control of the registered data remains with the originating institution. This approach also enables federation, the formation of coordinated strategies for ensuring authenticity across multiple independent shared collections.

The data grid community is not the only group that is providing infrastructure to assemble authoritative shared collections. The digital library community is developing standards for encapsulating and structuring metadata (METS - Metadata Encoding Transmission Standard), for transporting metadata (OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting), and explicit metadata schema (Dublin Core). The preservation community is developing the Open Archival Information Standard (OAIS) and mechanisms to encapsulate metadata and data for long term preservation (AIP - Archival Information Package). The preservation community is also developing Producer Archive Submission Pipelines for the authoritative import of data into a shared collection or preservation environment.

For the medical community to deploy data grids upon the infrastructure that has become available, several initial decision points need to be addressed across the organizations engaged.

These include:

- Specification of criteria for when data can become public or should remain as intellectual property.
- Agreement on authentication mechanisms. An example is use of the Grid Security Infrastructure for managing identity certificates stored in Certificate Authorities.
- Standard semantics for metadata attributes that will be used to support browsing and discovery
- Standard data encoding formats such as the DICOM format for medical imaging
- Standard services for manipulating data, such as metadata extraction routines for parsing DICOM headers and loading the metadata into a collection
- Standard identity tag to support correlation of records and treatment across institutions. This may be a one-way hash of the true patient identity, or a Globally Unique Identifier.

Given consensus on these points among entities participating in a data grid, it is then possible to design a data grid and even data grid federations that enable research or direct patient care. An example is federation of Regional Health Information Organizations (RHIOs) to build larger collections of authoritative records. Each RHIO could choose the set of metadata and records that would be registered into the shared collection. Anonymized records could be replicated to build specific collections intended to address a particular research issue. Logical shared collections could be created that correspond to the same characteristics as a selected individual, enabling through association the creation of virtual personalized clinical trials.

An effective way to demonstrate the power of data sharing is to assemble testbeds that provide access to shared collections for either research or personalized medicine. This will require linking health providers or biomedical researchers through use of data grid architecture. Explicit projects are needed that address a specific commonly shared goal that links data resources across currently independent environments.

An example is the development of shared collections between the DOD and the VA for longitudinal patient records. This requires a system that supports the strong authentication required for use of DOD structured information applied to VA records. This also requires the development of mechanisms to enable the development of structured information subset of the VA data for use in research.

A second example is the use of Access Grid technology to manage information distribution during the SARS epidemic in Taiwan. Access Grid technology (interactive multi-media systems) was installed in each hospital that had managed patients with SARS. This enabled the sharing of information and also verbal communication between doctors without the risk of personal contact (a temporary virtual organization based on collaboration and data sharing grids).

The set of services that enable access and use of shared collections can be quite extensive. When faced with the broad range of existing services, most communities develop simplified interfaces that expose to the user only those operations that they will need. Portals are used to control the presentation of the services, and simplify the management of HIPAA requirements.

The desired result is the formation of testbeds that demonstrate the functionality that can be achieved through development of shared collections. The specific actions that need to done include:

- Gap analysis of current technology.
- Promotion of community standards based open source tools
- Development of training material on use of the standard tools by the medical community
- Development of a primer on the principles and practices for use of grid technology, tuned for the medical community system developers
- Selection of resources that should be federated such as patient medical records, genome databases, proteomics databases, research publications, environmental databases, and genealogies.
- Possible resources that could be used in the testbed will depend upon the scale of analysis that is desired. For very large analysis, the NSF Teragrid is used to manipulate 10-100 Terabyte collections. For smaller analysis, institutional clusters will be adequate (1-10 Terabtyes).

The timetable for development of initial testbed plans is a single year. This will require developing strategic partnerships between the institutions that provide data resources and institutions that support the extended data analysis.

Desired result

- Promote interdisciplinary research projects
- Interoperability among VAMC, MHS, and academic affiliates (medical schools)
- Interoperability between research topic focuses (academic health centers, caBIG, Clinical Trial Organization - expand test community to broader community, Indian Health Service)
- Across medical practice communities (Clinical Trial Organization - expand test community to broader community, Indian Health Service - epidemiology studies)

Actions

- A major action item is the analysis of the support requirements for protected health information versus research information. The shared collection will need to differentiate between research information and private health information. At the moment, the vast majority of data in private health care are not accessible. In order to use this data for research, the community will need the ability to:
- Create a mechanism to de-identify data. This has been done in the NIH Biomedical Informatics Research Network, and includes the modification of MRI images to remove pictures of the patient's face.
- Create a mechanism to assign a unique person identifier to data (one-way hash) that allows correlation between data sets while maintaining patient confidentiality.
- Development of a standard publication mechanism for selecting information for use by research across a federation of repositories. The expectation is that each institution will retain control of their information, and will choose the subset that will be published into a shared collection addressing such complexities as k-anonymity. The DSpace

digital library provides a workflow environment for managing the addition of data and the creation of appropriate structured metadata. A similar mechanism can be developed for the medical community.
- Development of a standard set of semantics and services. Two approaches can be tried:
  - Bottom up through creation of shared collection
  - Top down through a coordinated federal initiative
- The Semantic Web and Semantic Grid communities may provide technology to simplify semantic interaction between collections. An assessment of current technologies is needed for implementations that could be used today.
- Development of standards for authoritative annotation (tags, new metadata, hyperlinks, provenance).

Resources

The major resource requirement is the identification of funding support for the development of cross-agency testbeds. This will require creating alliances and shared research initiatives that can drive the formation of the shared collections.

Existing grids can serve as testbeds on which the grid services are demonstrated. They are effective exemplars for the use of shared collections. An example of such a data grid is a project in Italy that uses data grid technology to build a shared collection across multiple PAAC archives, with mining of DICOM metadata to support browsing and discovery.

Existing computation and storage resources include the NSF Teragrid, NIH BIRN project, DOE Open Science Grid, and international grids such as EGEE. Possible groups for using the technology include the military medical system, healthcare delivery system, and the Indian Health Service.

Strategic partnerships should develop alliances that can drive both national and international research project. Opportunities include:

- International health study group
- Participation on international workshops (India, EU, International Council on Technology, CODATA, Asian Development Bank, Agency for International Development, International Council for Science, World Data Centers, Global Health Office, ONC)
- Relevant profession societies
- Gates Foundation
- Neutral facilitator to encourage collaborations
- Grid communities - EU HealthGrid, EGEE, …
- Digital libraries
- Persistent archives

Actions

- Build MPOP - Medical Point of Presence systems that simplify participation in a data grid, while integrating with care delivery systems. Analogous systems have been developed for deployment of oceanographic research vessels by Scripps Institution of Oceanography and for creating the BIRN data grid. The combination of hardware and software enables easy connection into a national data grid, with the cost of such systems being as low as $1200 per Terabyte for systems that store 20 Terabytes of data. The systems include an operating system, a Gbyte of memory, a RAID disk controller and a Gigabit-Ethernet network connection.
- Promote last-mile connections to health clinics, preferably with wireless networks.
- Develop standard clinical protocols for building structured data of sufficient quality for inclusion in a national HealthGrid.
- Apply in Global Trials projects to support global health issues such as avian flu.

Resources

The use of modular systems to enable health clinics to build and manage their own data grids is the foundation on which national shared collections can be constructed. The use of data grid federation technology would then enable the creation of virtual collections that can be designed to meet specific health issues. Groups would be able to participate in national scale issues, while retaining local control over information needed to address local issues.

Strategic partnerships

- Inter-agency partnerships with NIH. One goal will be to tie the AHRQ last-mile systems into research systems
- Public Health Service, USDA, FCC universal services support, FDA are organizations that can define national-scale research goals that depend upon the aggregation of medical records across health clinics.

Actions

- Identify interactions with personalized medicine
- Identify grid interactions with large scale drug-trials. This assumes that grid technology will be used for large studies.
- Identify common models for promoting sustainability on an international scale

FUTURE CONSIDERATIONS

GOAL: Install generic infrastructure to coordinate federation of distributed data repositories

Objective: (1 -3 years)

- Development of structured data to increase value for the community.
- Demonstration of how grid technologies enable analysis and data management.
- Identification of middleware services needed to access and use resources.

GOAL: Enable progress in science and health care

Objective: (1-3 years) Enabling rapid progress in science and health care are:

- Define benchmarks for how testbeds would work
- Define how to evaluate results of research for its value in medical practice, and how to translate research knowledge into actual practice
- Develop management policies that promote use of research in actual practice

Tasks:

- Identify and build shared collections that represent intellectual capital enabling research
- Identify standards on which collections are based and used (semantics, encoding format, services)
- Complete a planning document
- Map mission and objects to other international efforts
- Identify best practices for driving joint research
- Establish a workshop on sharing clinical trial results through data grids
- Establish a demonstration pilot that illustrates capability of data grids

GOAL: Enable use of research results in multiple care sites including public community health clinics

Objectives (3-5 years) Enabling use of research results in multiple care sites are:

- Support discovery and access to data from health clinics to enable research. This includes the promotion of bi-directional data exchange from health clinics back to research institutions
- Support the clinical research forum. This includes the promotion of the involvement of primary care providers and patients in

driving the research agenda. Feedback is needed from the consumers of medical practice on the impact of research.

- Enable use of research results in private practice. Mechanisms for assessing the quality and worth of the research results will be needed. Mechanisms will also be needed to validate the transformation of research knowledge into actionable medical practice. A consensus process is required for assessing the utility of the research results. This can be driven by existing panels that assess knowledge and relevance of research findings.

Tasks:

The desired results include the establishment of communication networks that link health clinics into a national healthgrid backbone. The "last-mile" problem of connecting facilities to existing national communication backbones is similar to the rural electrification effort. While, wherever possible, extensions of regional and metropolitan optical networks will support the most potential for future growth and network evolution, the advent of high speed wireless communications networks offer significant potential for those most remote sites not well served by existing infrastructure. Examples include the use of wireless networks to link Indian reservations, remote observatories, and environmental monitors in San Diego.

GOAL: Promote international collaborations based on open source and virtual organizations

Objective: Ensure open access while seeking interoperability across the following exemplar projects:

- BioSimGrid - example of an international research data grid
- caBIG UK MRC
- Global Trial Bank
- NLM malaria program

The major challenges are related to the use of different patient confidentiality requirements within each nation. This can be turned into an advantage by noting that European trials provide earlier results than US trials. Thus, an international data grid could provide information to the US that would improve understanding and early access to trials results that can impact the US research agenda.

The communities that are developing data management tools are international in scope. Thus, efforts at creating generic Semantic Grids, Digital Libraries, and Persistent Archives are being conducted in the US, the European Union, the UK, Taiwan, and Australia. Freely available tools will enable open collaborations that span national boundaries.

GOAL: Promote sustainability and governance mechanisms for long term support

Objective: (5-10 years) -Ensure the continued support of the digital holdings that represent the intellectual capital on which future research decisions are based. The ability to assemble authoritative holdings that have been validated for accuracy, have the appropriate data structure (both semantics and data encoding format), and that represent the most current understanding of medical practice is essential. These requirements have analogies in the preservation community in the concepts of authenticity and integrity. The ability to track provenance of data and the ability to insure the integrity of the data are two of the preservation principles driving development of persistent archives. The third preservation principle is infrastructure independence, the demonstration that the choices of storage and information management technology have not introduced any restrictions on the ability to migrate the archives onto new, more cost-effective technology. The systems that provide these capabilities are constructed today on top of data grid technology.
The explicit objectives are:

- Promote collaborations that drive the formation of the shared collections
- Promote interactions with agencies to encourage use of research results
- Develop groups of expertise that will drive the research agenda
- Develop communities of common interest that will maintain the expertise needed to interpret and analyze the digital holdings.

Long-term sustainability requires the formation of a market model that will provide incentives for sustaining the shared collections. The challenge is to create a market that will sustain the shared collections. This requires:

- Defining a self-sustaining economic model for supporting the shared collections
- Building a self-sustaining economy that both enables use of the research results and drives the development of future research results. An example is personalized medicine that builds upon the creation of virtual personalized clinic trials based on the set of persons with similar genetic and environmental factors.
- Creating the ability to participate in multiple virtual organizations. The ability to use the same clinical records in multiple health initiatives will decrease the reliance on a single sustainability model. This requires analyzing the economies of scale for promoting the addition of new information, and incentives for growth of the shared collections to make them attractive to a broader range of research questions.
- Defining models for publishing data ("pay" on publication, "pay" on access, pay as part of the research initiative, or pay as institutional infrastructure). Examples include the public utilities. One can envision a health data grid to which all medical institutions subscribe, with the health data grid maintained as part of the shared infrastructure for enabling medical practice and medical research.
- The expectation is that this will require development of new "economic" incentives related to quality of life.

While this goal is at least a five-year effort, it needs to be started immediately to ensure that a health grid is part of the future US infrastructure.

EPIDEMIOLOGICAL NETWORKS WITH A FOCUS ON KNOWLEDGE GRIDS

Group Leaders: Timothy Pletcher and Ian Foster

Group Members:

James Bayuk
Todd Carrico
Gary Crane
Donald Cook
Susan Estrada
Michael Fitzmaurice
David Forslund
Thomas Hacker
Peter Highnam
Mary Kratz
Joe Mambretti
Kevin Montgomery
Tony Solomonides
Russ Truscott

BACKGROUND

Conventional epidemiology requires extensive collections of data concerning populations, health and disease patterns, and environmental factors such as diet, climate, and social conditions. A study may focus on a particular region or a particular outbreak, or it may take as its theme the epidemiology of a condition across a wide area. The range of data required will vary with the type of study, but certain elements persist. A degree of trust in the data is essential, so it 'provenance' has to be assured and the standards of clinical practice under which it was obtained have to be above a certain threshold. Where the data has been gathered under different clinical regimes, it must be possible to establish their semantic equivalence, to ensure that aggregation or comparison of datasets is legitimate. Ethical issues may also arise if data collecting in the first place in the source of individual healthcare is to be used for research.[22]

Epidemiological science is like a puzzle, of which disparate players each have a piece. The big picture does not become clear until all the pieces have been brought together and assembled. Technology is the glue that holds the pieces of information together.

The top level goal of epidemiology is to improve health by detecting disease (both known and unknown) and managing start to finish including the collection, analysis, integration and response and detecting trends throughout the process. Epidemiology also aims to improve treatment, predict disease proliferation, develop policy for disease prevention and respond appropriately to events to mitigate the event and minimize damage.

Epidemiological science is the consummate application for the Grid. Since its inception, epidemiological science has brought together virtual communities of like-minded people who want to prevent disease. Shared goals that are too big for one group alone drive collaboration and the formation of virtual organizations. Technology is enhancing scientists' ability to collect statistical information, which can be combined with data from human sources, but it is also complicating the analysis of the data for predictive and combative purposes. More data means more calibration needed among data sources resulting in more time needed to analyze the data.

BARRIERS AND RESEARCH CHALLENGES

In general, the goal of epidemiological work is to identify, predict and react to disease trends. Grid use will help in identifying emerging diseases and modeling the spread of the disease. Simulations will be more easily accomplished driving clinical and public health cures and policies. The present difficulty is that the level of IT support is not consistent within and among virtual communities.

In order to define the Grid, it is necessary to study where the intersection of epidemiology networks and computer science lies. Active discussions between the groups in the two areas of expertise are needed to bridge the gap between the two disciplines.

CURRENT EXAMPLES OF EPIDEMIOLOGY GRIDS ARE:

- Genetic epidemiology Grids for the identification of genes involved in complex diseases
- Statistical studies: work on populations of patients. One example is the tracking of resistance to therapeutic agents. This is most notable in relation to antibiotic resistance in common bacteria in nosocomial and community settings
- Drug assessment: drug impact evaluation through population analysis
- Pathology follow-up: pathologies evolution in longitudinal studies
- Grids for humanitarian development: Grid technology opens new perspectives for preparation and follow-up of medical missions in developing countries as well as support to local medical centres in terms of tele-consulting, tele-diagnosis, patient follow-up and e-learning.
- Surveillance has taken on a much higher level of importance due to the increased attention being paid to global disease tracking and bioterrorism.

Public health addresses the overall health of a community based on population health analysis and intervention. The scope of any impact to a community's heath may range from a small handful of people to the case of a pandemic, involving whole continents; nevertheless, with increased mobility due to air travel and the open exchange of goods and services worldwide, even the smallest groups may ultimately affect population health on a global scale.

Public health activities (or functions) can be thought of as falling into four general domains: identification, analysis, intervention and communication. In today's increasingly interconnected world, the identification and analytic activities within the field of public health require the rapid collection of ever-growing volumes of disparate data elements concerning populations, health and disease patterns, and geospatial coordination and collaboration. Environmental factors, such as diet, climate, and social conditions, to name only a small set, further highlight the amount of detail and complexity to be managed. Such massively aggregated and distributed data requires the construction of complex models and the use of sophisticated statistical tools.

On the horizon, genomic analysis will significantly extend the variables under study, the computational burden to be undertaken and the range of expertise to be coordinated and employed. Traditional shoe leather approaches to collection, process, analysis, collaboration and intervention, while enduringly essential to every Public Health activity, are confronting an onslaught of digital measurements, thus creating excessive, distributed volumes of data. At the same time, traditional computer systems and architectures are proving inadequate to support these critical public health activities. The world of public health data is fragmented within separate application silos and offer limited, if any, interoperability and knowledge sharing.

One promising solution under consideration is grid architecture. Modeled after the Internet, it has matured into a highly distributed, interconnected network of computer resources, offering data, computational, service-oriented and collaborative access worldwide. Grid offers a powerful, flexible, scaleable and extensible infrastructure capable not only of sustaining the public health agenda today but enabling it to grow and flourish in new ways, especially in the sciences of epidemiology and informatics, under whatever extreme load it will most certainly face in the future.

In order to provide a clear framework for understanding the value of a HealthGrid for the domain of public health in general, this discussion focuses mainly upon the epidemiologic activities and their unique challenges. Epidemiology encompasses a large enough segment and traverses other sub-domains sufficiently within public health to be representative and informative for the whole.

The science of epidemiology, for instance, involves the following:

- Collecting information for identifying, characterizing, and monitoring events and conditions of public health importance;
- Processing information about events and conditions of public health importance by standardizing and compiling data from various sources in a timely manner so that these data can be analyzed with commonly available analytical tools, compared, and, if appropriate, linked;

- Analyzing information from various sources to identify, characterize and monitor events and conditions of public health importance and evaluate the impact of the public health response;
- Using the analyzed information to evaluate existing and new public health and clinical interventions, as well as impact public policy.

The time-honored activity of public health surveillance is the ongoing, systematic collection, analysis and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those responsible for prevention and control.[23] Given the increased attention being paid to global disease tracking and possible bioterrorism, new concepts, such as situational awareness and early event detection, have come into the lexicon. These new concepts, in contrast to classic, systematic public health surveillance, attempt to address the issue of extremely rapid ad-hoc information collection and analysis.

Although clinical medicine and public health both use the science of epidemiology to improve the health of the population, these classically disparate domains are beginning to converge in many ways. In the wake of advancing technology, healthcare providers now have the ability to quickly examine their clinical population, similar to the activities classically performed in public health departments. Specifically, through the maturing field of medical informatics-for example, the increasing adoption of data standards and electronic health records (EHR)-clinicians are now able to perform these new "public health" tasks. One example is the ability for a clinical practice or hospital network practitioner to perform an ad hoc query on the average hemoglobin A_1C level in their population to access quality of diabetic care. Either in the old model this would be performed by public health or by clinical research studies- both time consuming activities.

The science of epidemiology often requires the integration of multiple, discrete, pieces of information. Classically, this data collection has meant significant manual effort (e.g., the "shoe leather" approach). The field of healthcare informatics, in facilitating the increased adoption of standardized electronic data and information, is helping to reduce the burden involved in this data collection process. A provider now collects clinical data continuously over time in a digital format. In a paper-based office, access to this type of data would require a visit by an epidemiologist. With electronic health records, these data can now be extracted, consolidated and analyzed remotely without an onsite visit. The challenge of public health informatics is to leverage the use of electronic clinical data (EMR/EHR) to be used for public health purposes. When this challenge is multiplied throughout cities, states and regions nationally and globally, the size of the problem grows exponentially. Until now, the field of information technology has provided no consolidated way of dealing with this growing problem of information overload as it may be assimilated, shared and analyzed worldwide.

Epidemiological science, therefore, is the consummate application for the Grid. A Grid-based architecture facilitates both the time-critical situational awareness requirements, as well as classical ongoing public health surveillance activities. Given the scope of population health (the focus of epidemiology) in the age of globalization, the Grid addresses the core issues of vast disparate data sources, a wide variety of stakeholders, and an ever increasing scale.

Public health touches the world on the scale of populations at all levels of society-the citizen of each nation; the disenfranchised and marginalized in every location; local, state, federal (CDC, HHS) and international (WHO) government agencies; academic and research organizations; as well as private and public corporations. A globally extensible public health grid supports the objective of public health and preventive medicine to improve the health of individuals, populations and nations. In an increasingly porous environment, where health concerns do not stop at national borders and diseases not only disregard all boundaries but also can easily spread to the other side of the planet on a jet or ocean vessel, the need for a common, standards-based platform that enables jurisdictional interconnectedness, interoperability and collaboration becomes crucial to population health and analysis. A global vision for a public health grid demands the following:

- Local health departments in every state, region and country become connected to each other;
- Hospitals, metropolitan hospital systems, pharmacies and labs become connected similarly on the same global network;
- Regional health information systems become connected to each other and to federal agencies within the same framework;
- Academic and research organizations become connected to all of the above entities.

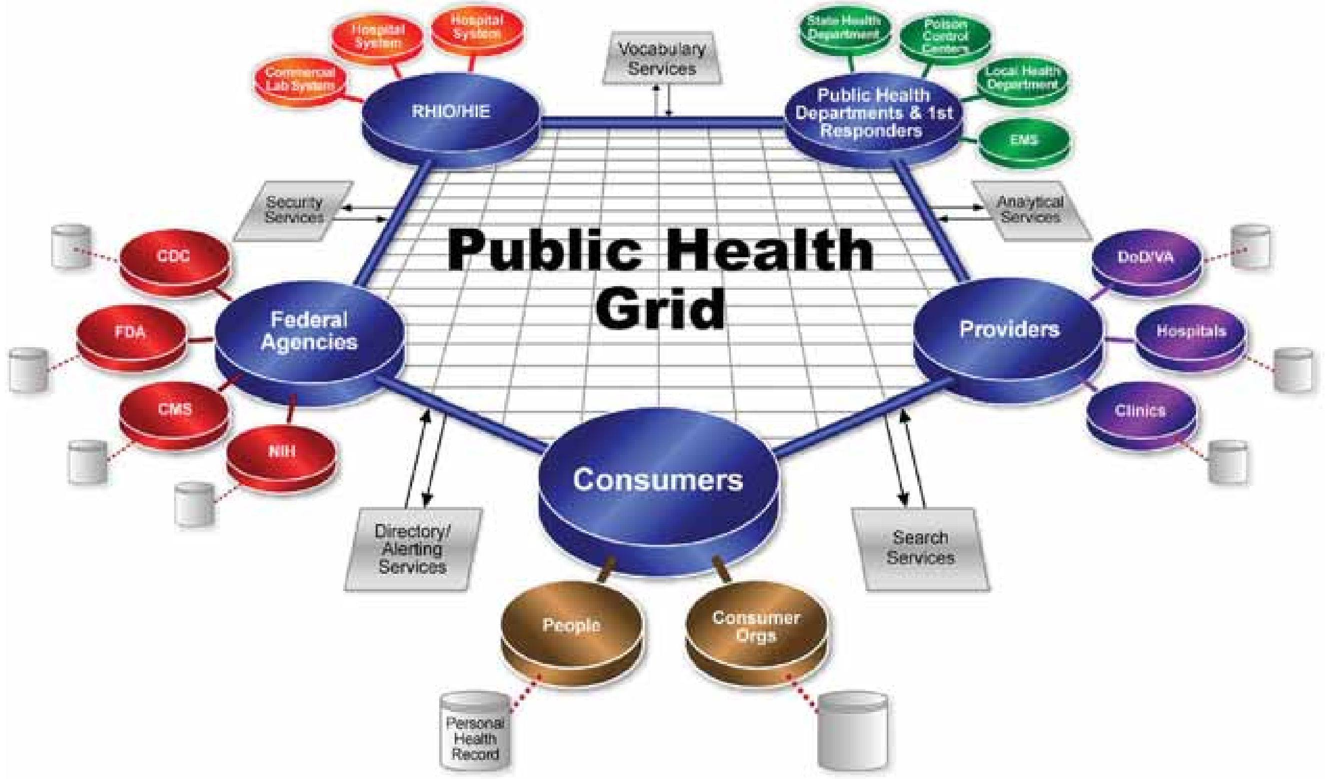
Thus, this vision encompasses nothing less than the interconnection and interoperation of local and state public health departments, hospital systems, academic and research organizations, regional health information systems and federal agencies across all national and natural boundaries.

Regardless that this notion is extremely idealistic and grand, the fact remains that a proven technological framework (which has not existed) exists today to inaugurate such a vision, thus making the decision to move in this direction by public health policy makers the responsible, essential and reasonable thing to do. If only one percent of this vision is implemented, public health gains in the end.

No connecting computer framework for public health has ever existed. The science and dedication of public health officials has provided the strongest linkage to date. Technologically, public health has depended too long upon individual, independent software applications, separate databases without any form of interconnectedness and lock-in from traditional vendors selling software for solving complex statistical problems, among other things. The Internet has certainly added a level of interconnectedness both socially and technically, but the Internet does not solve the larger public health computational dilemma. A public health grid builds upon the model of the Internet by providing the tooling and infrastructure necessary for application, data and computational interconnectedness and interoperability. This then becomes a primary requirement for public health to impact population health on a global scale.

In order to get a better picture of this grand notion, it may be helpful to reduce the scale of this vision to a view of the national landscape. The following simplified diagram represents a model of the US Public Health Grid (See Figure 1).

Figure 1 - CDC Graphic



In this thinner scenario, a national public health grid would interconnect Public Health Departments, RHIOs, Providers and all of HHS. Such a configuration would form a grid of services based upon national standards, providing collaboration resources, semantic and syntactic data exchange, computational and application resources. This rich computational environment will include a variety of services based upon Service Oriented Architecture (SOA). A small sample of services that will become essential components in the public health grid toolkit is listed:

- Security Services
- Architectural Services
- Vocabulary Services
- Analytical Services
- Visualization Services
- Search Services
- Directory and Alerting Services
- HL7 Transformational Services

These services will provide the necessary building blocks for all those involved in public health activities. As services on the grid they will be accessible by all grid participants; therefore, for example, a local public health epidemiologist in Attica, New York may use the same Period Prevalence Analytical Service that a CDC epidemiologist uses in Atlanta, Georgia; each supplying different sets of data yet utilizing the same statistical tool. A researcher at the University of Washington may run a query against public health data located within Washington, Oregon and Idaho to develop a tuberculosis outbreak simulation within these three states. Research findings could be shared through the public health access grid with other researchers at other institutions throughout the country. The applications and capabilities are innumerable.

The national public health grid community would participate in and adopt interconnectivity and interoperability standards being developed at the national / federal level. These standards would ensure that the public health grid would be able connect and operate with hospitals, regional health organizations, health information exchanges and other entities not directly connected to the public health grid. Services could be developed to interact with each of these entities to support basic data exchange and interoperability. Specifically, for example, the public health grid would help facilitate recommendations from the AHIC Population Health and Clinical Care Connections Working Group relative to priority areas such as public health case reporting, bi-directional communications, response management and adverse event reporting.

As a way to further understand the implications of a national public health grid, it may help to sharpen the view by looking at a potential first step toward building a public health grid through the implementation of an epidemiological weather map. The science of epidemiology requires the integration of multiple, discrete, pieces of information, leveraging the critical steps of identification, analysis (prediction), communication and intervention. The grid facilitates these steps. The use of grid resources-computational, data, analytical and collaborative-can help in identifying emerging diseases and modeling the spread of disease. Simulations can be more easily accomplished, thus driving clinical and public health cures and policies.

Some of the theories in epidemiology data collection and analysis parallel those of coastal ocean prediction research. Grid research and development has grown dramatically among academic institutions. Universities have discovered the mature nature of grid technologies in their research agendas to solve large scale computational and data problems. The SURA Coastal Ocean Observing and Prediction (SCOOP) program, for example, is integrating diverse efforts and empowering a virtual community of scientists with the tools, resources, and ideas to promote the effective and rapid fusion of observed oceanographic data with numerical models and to facilitate the rapid dissemination of information to operational, scientific, and public or private users. One of the goals of the SCOOP program is to "deploy the technical infrastructure to create an environmental prediction system that can be used as a research tool and handed off to the responsible entity that will use it to support the decision-making activities that benefit society."[24] They were able to predict the storm surge height of Katrina approximately twenty minutes before it made actual landfall; and this was in research mode with minimal operational resources; imagine the predictive capabilities once it's handed over to an operational entity to monitor. An epidemiological weather map project can leverage the research, tools and lessons learned by the SCOOP project to begin to build out the national public health grid infrastructure. Ocean scientists have been focused on creating a preoperational prediction model that seems similar to a virus spread prediction model.

Universities and research organizations across the globe have been working for many years to successfully solve large-scale computational problems such as this one by utilizing a grid framework. Many of these types of research problems have direct analogs with the computational and distributed data problems faced by the national public health community. Pardon the pun, but the perfect storm seems to be forming where the accumulation and confluence of grid research by academicians to solve large scale computational and data issues is cresting at the exact moment when similar large scale problems being faced by the public health community need answers. Ideally, CDC's Public Health Information Network (PHIN) community will help facilitate this initiative, at first as a research activity with the Public Health Informatics Centers of Excellence, then gradually growing it into an open collaborative effort involving NACCHO, ASTHO, CSTE, APHL, NAPHSIS, NAPHIT, state and local health department partners and internal subject matter experts in the National Center for Public Health Informatics and the National Center for Health Marketing.

To help visualize what an epidemiological weather map might actually look like, an interesting paper was presented at the Infovis 2005 Conference in Minneapolis. The paper was titled "Visual Correlation for Situational Awareness".[25] The authors described a novel visual correlation paradigm that takes advantage of human perceptive and cognitive facilities in order to enhance users' situational awareness and support decision-making. The following graphic represents a biological threat analysis but could be easily applied to an epidemiological weather map.

In order to protect US citizens from the threat of chemical and biological attacks, the Department of Homeland Security proposed a program called BioWatch to detect and report the presence of harmful agents. One of the most significant obstacles BioWatch faced was how information about possible attacks is quickly and effectively communicated to control centers.

This information includes:

- Detected agent or agents,
- Probability of their existence
- Where they are detected
- How the probable presence of such agents changes over time.

Understanding all of the dimensions of this information is critical to developing response plans in order to protect people in danger, and to create defense strategies against future possible attacks.

For example, if an analyst wants to understand how symptoms and treatment correlate to the possible presence of an agent, the following graphic (Figure 2) would be viewed.

Without going into detail and simply applying this map analogically to an epidemiological weather map, data gathered across the national public health grid from public health and clinical health care data repositories throughout the country would populate such a map, as in this example, to provide real-time epidemiological conditions in specific locations throughout the United States. Certainly, the cyberinfrastructure needs to be developed for this to occur, as well as the application of public health science to inform the services along the grid to provide such a view-however, an operational grid framework already exists. The obstacles facing the public health community at this point in the "perfect storm" are the dissemination of knowledge about the potential of a public health grid, the ability to communicate this vision in a way to be easily understood and appreciated, and the commitment by public health policy makers to move in this direction.

Figure 2 - SCI Graphic

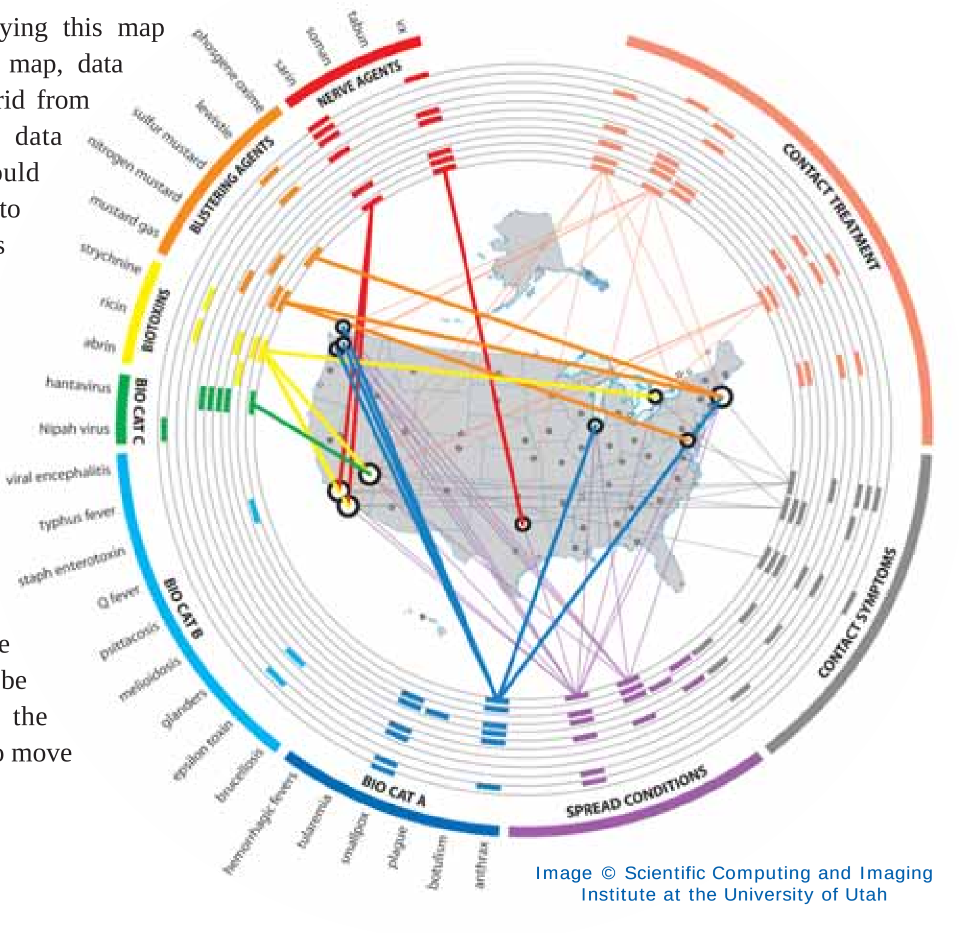


Image © Scientific Computing and Imaging Institute at the University of Utah

Some of the core concepts that make this vision compelling are as follows:

- The grid framework is based upon an open-source, collaborative development model;
- The framework has matured within the academic research community over many years;
- Many existing grid resources may be leveraged and reused to build out the national public health grid;
- The grid community is committed to a standards-based approach to the framework and to all services on the framework;
- The grid framework provides all of the necessary components for doing public health on a national and global scale: computational, data, analytical and collaborative;
- The public health grid is agnostic to users, whether they are states, hospital networks, clinics, solo practitioners or third world countries (A grid node may be run on a simple inexpensive desktop computer);
- The public health grid leverages economies of scale and minimizes redundancy of effort-build once, share with all.

The notion of a global public health grid is extremely idealistic and grand, but the fact remains that a proven technological framework (which has not existed before) exists today to inaugurate such a vision, thus making the decision to move in this direction by public health policy makers the responsible, essential and reasonable thing to do. Again, if only a small percentage of this vision is implemented, public health gains in the end.

Challenge Area	**Grid Impact**	**Current Maturity of technology that supports the space** *Immature, in process, mature* *Does it exist?*	**Cost -- the initial amount of money it will take to make a difference/ minimum level of purpose- Phase 1** *High, medium, low*	**How long it will take to make a difference/reach minimum level of purpose *(years)***	**Barriers**
Data collection, structure and aggregation (Capture and organization of the data)	High	Immature to in process	Medium to high	1-5 years	Organizational barriers to sharing data, political will, ontologies, human capacity
Disease outbreaks	High	Mature - technology exists but it is not deployed	Medium to low	1-5 years	Lack of funding, lack of political will, international political issues, resource issues, confirmations versus indirect evidence - sensor calibration
Longitudinal studies e.g. BIRN brain studies	High	In process	Medium	1-5 depending on the type of study	Lack of funding, demonstrated value, IP, comparability of data/calibration among data collection devices
Strategic decision support	High	In process	Medium to low	1-10 years	Organizational linking - use of tool by people making decisions, flexibility of computational and network resources, linkage between research and tactical decision-makers
Tactical decision support	High	Immature to in progress	Medium	1-5 years	Credibility of management, media and govmts; organizational linking;
Modeling normal activity	High	In process	Medium	1-2 years for baseline, 2-3 more years to establish criteria	Need robust disease outbreak solutions in place

Dynamic reallocation of computational resources (e.g. changing queue structures in Teragrid, moving data, moving programs to data, etc.)	High	In process with regional maturity	Low to medium	1-5 years	Lack of standards, lack of policies, lack of expertise, lack of funding
How to support in countries/localities with underdeveloped infrastructures (i.e. people and technology that supports that) Loss of infrastructure-electricity, bandwidth, etc. both disaster backup and underserved areas	Low	In process to mature	High	5-10 years	Lack of funding, organizational barriers, sustainability of deployed technology
Developing and maintaining social networks to insure solid data collection	Low	Immature	Medium with high variance	1-5 years	Education of human capital, development and deployment of incentives, lack of national health info infrastructure - lack of passive data collection
Viz of data -- Scale of viz and metaphors - normalizing the visualization (color schemes, etc. to standardize the output)	Medium	In process	Low	1-5 years	Effective collaboration between CS and bioscience
Capture good current practices from solid, well documented existing studies - leverage overdeveloped environments, microcosm	High	In process	Low	1-3 years	Identifying the examples and communicating to wider groups, visibility

FUTURE CONSIDERATIONS

Goal: Framework for Dynamic Creation of Collaboration Grids
Create a software support infrastructure that makes creating teams and virtual organizations simple and easy.

Objective: (1-3 years) Understanding existing technology and what is possible

Task: Create and maintain Resource Inventory of trusted tools
Task: Cookbook / HOW TO for creating Collaboration systems
Task: Existing standards and interoperability overview
Task: Examples and use cases

Objective: (1-3 years) Deployment of collaboration testbeds

Task: Identify and fund collaborative environments
Task: Disseminate best practices from working collaboration systems
Task: Identify logistics planning and service support organizations
Task: Create expertise discovery mechanism

Objective: (5-10 years) Understand how collaborative environments change workflow and output

Task: Fund and implement a study on tools and processes necessary for successful implementation of grid-enabled dynamic collaborations.
Task: Follow up / revision of Cookbook for creating Collaboration systems

COLLABORATION GRIDS IN SUPPORT OF DATA VISUALIZATION AND DISTRIBUTED SIMULATIONS

Group Leaders: Parvati Dev and Chris Johnson

Group Members:

Michael Ackerman
Bob Aiken
George Brett
Todd Carrico
Jonathan Dugan
Larry Flournoy
Richard Foster
Peter Honeyman
Bill Lorensen
Harvey Magee
Michael Mauro
Melvin Sassoon
Richard Soley
Tony Solomonides
Anil Srivastava

BACKGROUND

Collaborative environments deployed over grid architectures, if appropriately developed and used, have the potential to change the methods and practice of science. Currently scientific collaboration for data analysis is typically asynchronous. Emails or data files are sent to individuals who perform analysis alone followed by return messages, review, discussion and the process repeats. The promise for collaboration systems over grid architectures is that many actions we typically think of as solitary, such as data analysis and visualization, will be done in real-time collaboration with other members of a team in remote locations. For example, viewing research data, running data analysis, and running a simulation or a model will all be performed with the input and experience of many different people contributing simultaneously to the process from distributed geographic locations.

Real-time distributed collaboration coupled with multi-directional voice and video communications enables multiple people to work together to view and manipulate data in real time, share ideas, test hypotheses and communicate. The people engaged will participate from disparate and even arbitrary geographic locations.

Biomedical research and health care present prime opportunities for use of these new technologies because these activities commonly depend upon complex multidisciplinary interactions among highly distributed teams. Though they are few to date and are pioneering in this area, there are already distributed research teams utilizing related technologies and environments for systems biology and neuroscience research as well as anatomic education, intensive care unit monitoring, and health care team simulation and training. As grid technology matures, such deployments and use will become more common.

BARRIERS AND RESEARCH CHALLENGES

Collaboration Grids are being used to create virtual descriptions and connections for people and groups, linking them in long lasting or ephemeral communities, for research collaboration. If these technologies evolve into standards of practice, they have the potential to revolutionize not only the process of scientific investigation, but also medical practice, emergency management, and learning environments. Data and Computation Grids further support the virtualization of these collaboration communities, by virtualizing the location of *resources*, so that sophisticated algorithms and huge data files are accessed and manipulated in community interactive sessions without regard to their physical maintenance. Visualization capabilities, from desktop to theater size, from simple to immersive, also bring people together in a *virtual space*, to talk, to manipulate and analyze data jointly, to practice multi-person medical and surgical procedures, and to practice and teach with visualizations from microscopic biologic detail to human-scale anatomy to city and region-scale disaster and emergency management scenarios.

DYNAMIC VIRTUAL ORGANIZATIONS

Many situations, particularly in biomedicine, involve creating dynamic teams of people interacting in real time to address specific issues. Such teams do not have rigid organizational structures or rules for how people interact, and they typically assemble in an ad hoc fashion, persist for arbitrary time periods, and may even disband at the end of each interaction. These are referred to by the grid community as *virtual organizations* because they are usefully thought of as arbitrary, temporary functional overlays of real organizations.

Grid computing enables the creation of ad-hoc groups of people to focus on specific problems. These groups typically exist for short periods of time, and their membership can either come from existing collaborations or from just-in-time groups forming to address a specific issue. Types of activities these organizations can conduct are also enabled by applications that run on the grid. Grid computing facilitates technology delivery among groups of people separated geographically.

For example, in the realm of epidemiology, virtual teams need to be dynamically assembled in real time to address specific concerns such as outbreak data. Grid services can enable dynamic formation of such teams with people appropriate to the problem at hand. In the case of an emerging outbreak, such a team might include local physicians, government health experts from different countries, and infectious disease specialists with topical coverage of the suspected pathogen.

In pre-hospital emergency care, early distributed computing efforts are bringing together collaborative teams over municipal wireless networks. By enabling emergency medical technicians in the field to simultaneously share audio, video, and physiological data with hospital emergency medical care providers, regional emergency medical network leaders, and even local incident commanders (at multi-ambulance field incidents), the data to manage complex or low-frequency (rare or non-protocol circumstances for emergency medical technicians) events will be available to all parties for real time shared critical decisions.

COLLABORATIVE COMPUTATION: SIMULATION AND TRAINING ENVIRONMENTS

Collaborative computing systems can also support real-time interaction to steer computation tools, both scientific simulation and real-world simulations for training. In these different cases, the computation may be different, but several common issues arise: the speed, stability, and rapid provisioning of the network, human communication and dynamics, and resource controls and data availability.

The complexity of creating collaborative environments is high enough that today service organizations exist to support these tools, much like NMR or Mass Spec service centers exist within biomedical research environments. Firewalls, for example, are essential in many organizations, but can be major barriers to the implementation of collaborative environments. Therefore, it may be necessary to create support services to assist and train groups in making the transition to grid-enables collaboration. For example Internet2, SURA, and the Research Channel have created workshops, cookbooks, and support services to assist their community in collaboration with videoconferencing. Their early efforts have now created a community that is at the forefront of the use of high-resolution video for communicating breakthrough scientific visualization, such as real-time, high definition video of the deep sea bottom.

With the advent of newer and faster networks exploiting the research and education community's investments in owned fiber optic infrastructure and with completely new architectures, (e.g. the fully optical networks), new network standards and protocols will emerge. This may have an impact on much prior work that is based on current network protocols. Collaboration, in particular, may be very dependent on protocol details such as multicast and high throughput transport protocols. As these new capabilities are investigated, it is important that the subject of HealthGrids, and its uses, be brought to the attention of the developers of these new optical networks.

DATA VISUALIZATION

Many areas of biomedicine leverage extremely complex data. Tools exist today such as VTK, SCIRun, ITK, Paraview, CAVE, geowall, clview, and Chimera to create visualization environments for data analysis. However, these tools were designed primarily to enable a single researcher sitting in front of a workstation to manipulate and view data by themselves. Increasingly, data are coming from many different sources for scientific research and clinical practice, and often those sources are geographically disparate. Many different experts must come together with the data to make sense of how it all fits together. Increased ease of communication, enabled by the Internet and tools such as Access Grid, enables geographically disparate groups to connect. However, there are significant technical barriers and substantial expertise required in creating and maintaining collaborative visualization environments, and the often amazing tools that do exist are known only to few bleeding edge groups.

Many different areas can benefit from distributed visualization:

- Trauma simulation
- Radiological data visualization for surgical planning and education
- Bringing investigatory learning into the K-12 classroom
- Doc in a tent in desert
- Injured soldier in a field
- Bioinformatics data, vector fields, genomics data
- Time series data for biological pathway analysis
- Ecological data and evolutionary trees
- Syndromic surveillance and outbreak data
- Geographic data coupled with epidemiologic datasets
- Anatomic data for education and physiology simulation for radiation treatment
- Health care team simulation for education and crisis/medical error research

ANATOMIC INFORMATION

Anatomic data is ubiquitous in healthcare. Radiological studies such as CT and MRI are critical to modern health care across nearly all specialties. For example, such varied domains as cancer, heart disease, neurologic disease, and sports medicine depend critically upon interpretation of anatomic data. Simultaneously, the teams involved in providing this care are increasingly expansive. Grid computing technologies will enable future systems that access, distribute, and support analysis of anatomic information thereby enabling significant new capabilities in healthcare delivery. Specifically, HealthGrids can bring together, across nearly limitless geographies and computational resources, such diverse data and services as: teaching and patient education models and four-dimensional simulations; patient specific volumetric visualizations for treatment planning; standard and new analysis methods such as computer-aided diagnostics.

Current challenges to distributed, anatomic modeling come primarily from three areas: First, the quality of available data is dependent largely on market forces that drive what data is acquired. Typically, this data is tailored to workflow patterns developed with solitary physicians analyzing data, for example a surgeon reviewing CT results or a radiologist reading an MRI. However, computer-supported collaborative environments create the potential for use of anatomic data and functional data that are current not available. For example, data that links the anatomy to the specific physiology of the patient, and a semantic understanding of which measurements apply to which anatomical part will enable remarkable capabilities across education, diagnostics, and therapy.

Second, several models exist that describe anatomy. Typically, these systems are limited in scope to selected organ systems or subsets of biology - and work on only one level of detail. To make a general system for managing anatomic data, such systems need to be integrated and work at multiple levels of abstraction. There is a connection gap between many currently existing data sources that provide intelligent information about physiology and function and existing software systems that manipulate anatomic data.

Third, Health grid technologies depend upon adaptation of computer technologies being driven from the defense, consumer, business, and entertainment sectors, such as messaging standards and video gaming systems, and large scale grid-based data management systems and services. Such general technologies need to be adapted and modified for biomedical applications requiring domain experts, computational experts, and informaticians working in concert to bridge the gap and make significant qualitative change in biomedicine. Early deployments of picture archiving and communications systems over grid architectures have been successful technically, as have shared distributed visualization systems and physiological modeling systems, but integrating these technical achievements with specific organizational programs for sustaining long-term outcomes has been limited as specific investment in deployment is needed, not just technical research, or direct clinical care.

FUTURE CONSIDERATIONS

Advances in large-scale computing and web and grid services technology now enable scientists to tap into powerful remote computational resources.

However, remote computation as it is currently practiced limits a scientist's ability to interact with the results of the computation, which may be generated a great distance away and may impose large storage requirements that cannot be met at the scientist's end.

When using remote resources for visualizing large datasets, therefore, scientists often revert to batch mode visualization, in which images and short animations are created and downloaded offline for later investigation. This approach may be the scientist's only choice. However, scientists using this method are often frustrated by their inability to dynamically interact with the visualization and analyze the data.

Interactive remote visualization is a relatively new research area that aims to bridge this gap and provide interactive visualization over the Internet. Interactive remote visualization splits the visualization process between the scientist's local machine and the remote computational end where the data reside. It is the task of the remote visualization infrastructure to connect the user's input, the rendering engine, and the data itself, and deliver the results to the user's display.

Several factors make remote visualization especially challenging. First, the large distances between the data and the display introduce long latencies that impede and may even prevent dynamic interaction. Second, bandwidth limitations reduce the amount, rate, and quality of visualization that can be streamed to the user. Third, the visualization resources at the two sites may not be compatible or may vary dramatically from one session to another. These resources limitations can range from slow CPUs and limited memory to slow and outdated graphics hardware. Finally, network reliability issues such as error rate and bandwidth inconsistencies or error rate and lack of throughput reliability may reduce the quality of the interaction.

The need for interactive, sometimes real-time interactivity, of many biomedical research and clinical applications is a major challenge in current Grid research and development, because most current Grid services are targeted towards static application partitions and are often not appropriate for visualization grid services that tend to be more heavyweight in nature.

A significant research and development opportunity is to create a uniform framework for grid-enabling and partitioning more heavyweight services such as remote visualization. Current challenges to providing Grid services for visualization or other interactive needs include:

- Grid visualization resources are limited. We must find an easy way to grid-enable existing packages.

- Grid resources are "wildly" heterogeneous with regard to Visualization (graphics, displays, I/O, plus usual). Programs should be performance-portable (sense resources at load time) and configure themselves in some "optimal" way.

- Grid resources are dynamic. Programs need be adaptive to dynamical adjust to changing resources - might need user decision (e.g. accuracy versus frame rate)

- Applications that provide heavyweight grid services must be resource-aware.

In addition, to Grid visualization services, there needs to be more research and development in Grid visualization algorithms, for example:

- Compression (wavelets, multiresolution, mpeg, etc.) algorithms
- Importance based rendering algorithms
- View dependent algorithms
- Image based rendering algorithms

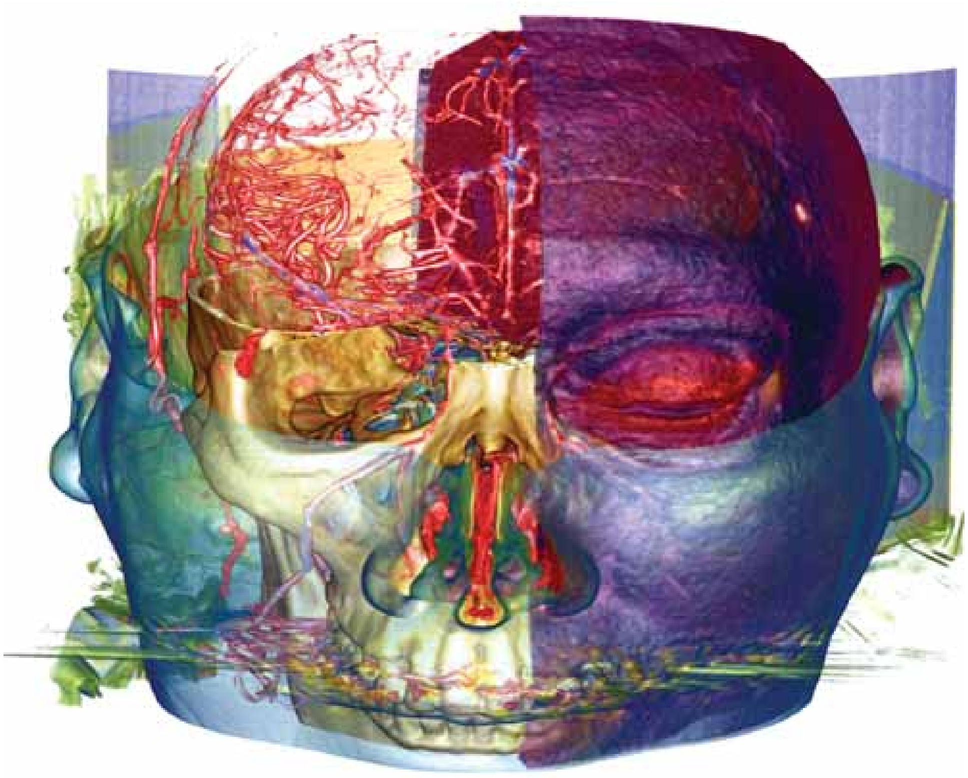


Figure Caption: Volume Rendering of Computerized Tomographic Angiography in SCIRun/BioImage using Multiple 2D Transfer Function Cutting Planes

Image © Scientific Computing and Imaging Institute at the University of Utah

Objective: (1-3 years) General availability of one or more set of grid-enabled tools for real time collaborative data visualization

- Task: Identify 2-5 sets of tools as possible candidates, and use in test beds
- Task: Select 1-2 sets of tools for further development
- Task: Implement user conference for further dissemination

Objective: (5-10 years) Sustained use of visualization tools in 2 or more biomedical domains other than original domain

- Task: Fund and implement multiple domain-specific testbeds for demonstration
- Task: Select a few domains for expanded support and increase number of collaborators

Goal: Real-time collaborative biocomputation and simulation

Create and support computation toolkits that allow multiple users to control computation and virtual collaborative environments, to support training and biocomputation.

Objective: (1-3 years) Build awareness of this new computational model

- Task: Develop white paper on existing biocomputation and simulation methods, and the steps necessary to enable collaboration using grids
- Task: Identify computational architectures, and network issues (firewalls, extensions of optical networks to critical remote facilities), that will be common in this transition
- Task: Identify middleware requirements, compare with those currently available standards and tools, and focus on the development of extensions to support this class of biomedical research.
- Task: Develop toolkits to support this transition

Objective: (3-5 yrs) Create a community of researchers conducting real-time collaborative biocomputation (other than BIRN)

- Task: Identify relevant areas in biomolecular research, infectious disease modeling, pharmaceutical development, and others
- Task: Support their use of the toolkits developed above

Objective: (3-5 yrs) Create a community of researchers using real-time collaborative simulations for training

- Task: Identify relevant areas in surgery, mass casualty training, operating room team training, and others
- Task: Support their use of the toolkits developed above
- Task: Investigate the impact of collaboration on the delivery of training and on the efficacy in learning.

HEALTHGRID STRATEGIC PLANNING SESSION: U.S. GOVERNMENT AGENCY COORDINATION

October 2, 2006

Introduction

On October 2, 2006, the HealthGrid Core Strategic Planning Group, composed of subject matter experts from key U.S. Government agencies, met to discuss the intersection of priorities related to the development of the U.S. HealthGrid; where biomedicine meets grid technologies. This addressed strategic planning interests of individual Federal agency programs related to the HealthGrid. Thirty HealthGrid Strategic Core Group leaders were invited to participate in an introductory planning session held at the National Science Foundation.

Mission and Purpose

The meeting was to follow-up to the, "HealthGrid: Grid Technologies for Biomedicine" Integrated Research Team, held by the Telemedicine Advanced Technology Resarch Center (TATRC)/US Army Medical Research Materiel Command, of March 1-2, 2006 that laid out the roadmap for the planning and development of the U.S. HealthGrid. The scope of the emerging HealthGrid is to translate research knowledge to enable better health outcomes within communities and to develop information that "health decision makers" (i.e., individuals, medical practitioners, payers and administrators) can use.

The purposes of this meeting were to: 1) recommend plans and tools to inventory various USG activities; 2) begin to identify priorities to realize effective utilization of resources and; 3) establish a community of practice to support collaborations for Grid computing.

Expected Near and Long-Range Strategic Outcomes

The outcomes of this meeting should provide: 1) the establishment of an action plan for inter-agency cooperation and; 2) identification of mutual interests between and across current programs; 3) strategies to make productive use of cyberinfrastructure; 4) inter-agency collaborations to develop and sustain US cyberinfrastructure for biomedicine;

Specific areas to be addressed in the plans and initiatives of the HealthGrid Research Roadmap include:

- Data movement across networks, collection, annotation and provenance
- Training; human capacity building on grid technologies and HealthGrid
- Visualization
- Simulation Models
- Utilization of computational grids (Teragrid and the Petascale Facility) in biomedicine
- Situational Awareness
- Systems Biology
- Translational Research
- Knowledge tools
- Making grid services real across domains

Background

The HeathGrid is the medical domain equivalent of the U.S. science and technology Cyberinfrastructure agenda. HealthGrid is necessary for the U.S. to support the application of exemplary computer science and engineering to biomedicine.

As clinician's and patient's demand for knowledge-driven information grows in urgency, the priority for advancing interoperability between biomedical research and clinical practice at the point of care, is recognized by leaders across government, industry and academia. A common approach is necessary to effectively address the benefits of applying Grid technologies and services to biomedicine.

A high-level strategic plan should be established for facilitating communications across intergovernmental agencies and organizations. A process for open collaboration should be followed that will lead to actions to connect U.S. Grid agencies and organizations, points of contact and potential partnerships in order to identify strategic utilization of resources.

The goal for expanding communities of specialized practices must be aimed at supporting efficient cyberinfrastructure environments for scientific collaborations and effective knowledge sharing across biomedicine, the healthcare industry and other domains.

Goals for HealthGrid Collaboration

- Provision of a high performance network across the medical domain; HealthNet.
- Enhancement of current visualization environments, and creation of new tools to provide understanding of the huge amount of data being generated in biomedicine (Systems Biology to Advanced Distributed Learning and Simulation)
- Discover linkages of data over the long term by capitalizing on public and private data sources for Situational Awareness. Knowledge, once discovered can be embedded back into production systems for future surveillance, early warnings, and prediction of the long term costs of supporting the future healthcare system, as well as better health care overall.
- Demonstrate that disease based Biomedical Information Grids (BIRN, CaBIG) scale to broader everyday areas of application. The approach is to provide a Systems Biology environment where large and rich data sets can be made available for longitudinal study, leveraging the tools coming out of the disease specific projects, and deploy grid services that, when brought together look like an application. Systems Biology will enable predictive, preventative, personalized, participatory medicine; deemed to be the future of medical practice.
- Enable Translational Research with the national capacity for data storage and access protocols, generic Grid functions applied to the medical domain, such as ubiquitous Physician Credentialing.

Uses of U.S. HealthGrid for Sharing and Collaboration

Grids are being used in healthcare in a number of ways. There are now at least three types of grids that have emerged. There are: (1) COMPUTATIONAL GRIDS being used to solve large-scale computation problems in healthcare research; (2) DATA GRIDS that don't share computing power but instead provide a standardized way to access data internally and externally for data mining and decision support; and (3) KNOWLEDGE GRIDS that let dispersed users share information and work together on extremely large data sets to build knowledge environments.

The HealthGrid Core Strategic Planning Group represented diverse areas of Grid technology across the U.S. Government. The following is a brief list of areas where Grid technology is being used in biology, medical and health

related fields by participating key agencies including TATRC (applied research); MHS (AHLTA data availability for public good and architecture); VHA (SOA activity and VistA); NSF OCI; NIH
(NCRR, NLM, NCI, NIGMS); and NASA:

- Genetic Linkage Analysis
- Molecular Sequence Analysis
- Determination of Protein Structures
- Identification of Genes and Regulatory Patterns
- Biological Information Retrieval
- Biomedical Modeling and Simulation
- Biomedical Image Processing and Analysis
- Data Mining and Visualization of Biomedical Data
- Text Mining of Medical Information Bases

Discussion and Next Steps

- Participants shared different views on the need for exploring, analyzing and identifying their projects and plans that might be considered for sharing.
- The group was invited to contribute inventories of agencies' projects and plans to a HealthGrid Resources Sharing Database. The need to make data registries a priority highlighted the need for resource inventories.
- A model Web-based database was presented and demonstrated through TATRC that may be considered for use as a tool for maintaining a project inventory or data registry to facilitate collaboration.
- If agencies move forward toward identifying sharing opportunities, priority consideration should be given to open solutions that represent best practices, and will contribute measurable products to the development of the HealthGrid. Overall, sharing opportunities may focus on data, information, systems, and sources of knowledge.
- Consider establishing a workgroup to identify functional requirements and/or potential uses of Grid systems for use by the different agencies.
- Identify potential organizations in the private sector to collaborate on the research, development, testing and use of Grids in healthcare, e.g. medical schools, vendors.
- Conduct a feasibility study into the use of grids and select potential pilot projects.
- Investigate changes in business and IT processes that may need to be made in anticipation of utilizing grid technology.
- Initiate and fund pilot project(s) and complete a detailed cost benefit analysis.
- HealthGrid activities will be coordinated under NITRD Group of the Subcommittee on Large Scale Networking and Information Technology, Middleware and Grid Infrastructure Coordination (MAGIC) Team.
- A Collaboration Expedition Workshop on Grid Computing is being planned for June 19th 2007 at the National Science Foundation. The workshop will be an opportunity for leaders and managers from different fields in the government and private sectors to come together in order to share perspectives and future directions in an overview fashion.

Session Co-Chairs:

Mary Kratz, MT(ASCP) USAMRMC/Telemedicine Advanced Technology Research Center University of Michigan Medical School Information Systems	Mike Sayre, PhD NIH/NCRR

Invited Participants, HealthGrid Core Strategic Group:

Michael Ackerman, NLM, NIH
ackerman@mail.nih.gov

Cathy Angotti, NASA
Cathy.angotti@nasa.gov

Dan Atkins, NSF OCI
Datkins@nsf.gov

Ken Beutow, NCI, NIH
buetowk@mail.nih.gov

James Cassatt, TATRC consultant
jcassatt@worldnet.attt.net

Peter Covitz, NCI , NIH
covitzp@mail.nih.gov

Theresa Cullen, IHS
Theresa.Cullen@ihs.gov

Jonathan Dugan, BioContact
dugan@biocontact.org

Michael Fitzmaurice, AHRQ
Mfitzmau@ahrq.gov

Helen Gill, NSF CISE
Hgill@nsf.gov

Gail Graham, VA
gail.graham@hq.med.va.gov

Chris Greer, NSF OCI
cgreer@nsf.gov

Carl Hendricks, DoD MHS
carl.hendricks@tma.osd.mil

Peter Highnam, NCRR, NIH
highnamp@mail.nih.gov

Robert Kolodner, VA/ONC
rob.kolodner@hq.med.va.gov

Mary Kratz, TATRC
mkratz@umich.edu

Betti Joyce Lide, NIST
bettijoyce.lide@nist.gov

Peter Lyster, NIGMS, NIH
lysterp@mail.nih.gov

Michael Marron, NCRR, NIH
marronm@mail.nih.gov

Grant Miller, NITRD
miller@nitrd.gov

Eric Noji, USUHS
enoji@cdham.org

Hon Pak, TATRC
hon.pak@amedd.army.mil

Michael Sayre, NCRR, NIH
sayrem@mail.nih.gov

Edward Sondik, CDC
esondik@cdc.gov

Ram Sriram, NIST
sriram@nist.gov

Steven Steindel, CDC
sns6@cdc.gov

Simon Szykman, NITRD
szykman@nitrd.gov

Paul Tibbits, DoD MHS
patibbits@US.MED.NAVY.MIL

Syed Tirmizi, VA
syed.tirmizi@va.gov

Susan Turnbull, GSA
susan.turnbull@gsa.gov

John Whitmarsh, NIGMS, NIH
whitmarj@nigms.nih.gov

Marc Wine, GSA
marc.wine@gsa.gov

BIBLIOGRAPHY

Baru C, Moore R, Rajasekar A, Wan M, The SDSC Storage Resource Broker, Proc. CASCON'98 Conference, Toronto, Canada, November 30 - December 3, 1998, 5.

Foster I, Kesselman C, *The Grid: Blueprint for a New Computing Infrastructure*, Chapter 5, "Data Intensive Computing," Morgan Kaufmann, San Francisco, 1999.

Goble C, Kesselman C, Sure Y. (Editors), Semantic Grid: The Convergence of Technologies, Dagstuhl Seminar proceedings 05271, July, 2005.

Hansen CD, Johnson CR. *The Visualization Handbook*, Edited by C.D. Hansen and C.R. Johnson, Elsevier, 2005.

Hansen C, Johnson CR. *Graphics Applications for Grid Computing*, Guest Editors, IEEE Computer Graphics and Applications, 2003. (23):2.

Johnson CR, MacLeod RS, Parker SG, Weinstein DM. Biomedical Computing and Visualization Software Environments, In Comm. ACM, 2004. (47):11:64-71.

Johnson CR, Moorhead R, Munzner T, H. Pfister, P. Rheingans, Yoo TS. NIH-NSF Visualization Research Challenges Report, IEEE Press, 2006. Website: http://tab.computer.org/vgtc/vrc/index.html - Accessed August 20, 2007

Johnson CR, Weinstein DM. Biomedical Computing and Visualization, In Proceedings of the Twenty-Ninth Australasian Computer Science Conference (ACSC2006): Conferences in Research and Practice in Information Technology (CRPIT), Edited by Vladimir Estivill-Castro and Gill Dobbie Hobart, Australia, (48):3-10. 2006.

Johnson CR. Top Scientific Visualization Research Problems, *In IEEE Computer Graphics and Applications: Visualization Viewpoints*, July/August 2004. (24):4: 13-17.

Macleod RS, Weinstein DM, de St. Germain JM, Brooks DH, Johnson CR, Parker SG. SCIRun/BioPSE: Integrated Problem Solving Environment for Bioelectric Field Problems and Visualization, In Proceedings of the Int. Symp. on Biomed. Imag., Arlington, Va, pp. 640--643. April, 2004.

Moore R, Baru C, Rajasekar A, Ludascher B, Marciano R, Wan M, Schroeder W, and Gupta A, Collection-Based Persistent Digital Archives - Parts 1& 2, *D-Lib Magazine*, April/March 2000, Website: http://www.dlib.org - Accessed August 20, 2007.

Moore R, Baru C, Virtualization Services for Data Grids, Book chapter in Grid Computing: Making the Global Infrastructure a Reality, New York, John Wiley & Sons Ltd, 2003. 409-436.

Moore R, Building Preservation Environments with Data Grid Technology, *American Archivist*, June 2006.

Moore R, editors S. H. Koslow and S. Subramaniam, Persistent Collections, book chapter in *Databasing the Brain*, John Wiley & Sons, 2005.

Moore R, JaJa J, Rajasekar A, Storage Resource Broker Data Grid Preservation Assessment, SDSC Technical Report TR-2006.3, Feb 2006.

Moore R, Marciano R, edited by Julie McLeod and Catherine Hare Technologies for Preservation, book chapter in *Managing Electronic Records*, Facet Publishing, UK, October 2005.

Moore R, Marciano R, Prototype Preservation Environments, "Digital Preservation: Finding Balance" *Library Trends,* Johns Hopkins University Press. (54)1: 144. Summer 2005.

Moore R, Operations for Access, Management, and Transport at Remote Sites, Global Grid Forum, GFD.46, December 2003.

Moore R, Rajasekar A, Wan M, Data Grids, Digital Libraries and Persistent Archives: An Integrated Approach to Publishing, Sharing and Archiving Data, *Special Issue of the Proceedings of the IEEE on Grid Computing*, March 2005. (93):3: 578-588

Moore R, Rajasekar A, Wan M, Storage Resource Broker Global Data Grids, NASA / IEEE MSST2006, Fourteenth NASA Goddard / Twenty-third IEEE Conference on Mass Storage Systems and Technologies, April 2006.

Moore R, The San Diego Project: Persistent Objects, *Archivi & Computer*, Automazione E Beni Culturali, l'Archivio Storico Comunale di San Miniato, Pisa, Italy, February, 2003.

Moore R, Wan M, Rajasekar A, Storage Resource Broker: Generic Software Infrastructure for Managing Globally Distributed Data, Proceedings of IEEE Conference on Globally Distributed Data, Sardinia, Italy, June 28, 2005.

Parker SG, Weinstein DM, Johnson CR. Edited by Arge E, Bruaset AM and H.P. Langtangen. The SCIRun Computational Steering Software System, *In Modern Software Tools in Scientific Computing*, Birkhauser Press, Boston pp. 1-40. 1997.

Rajasekar A, Moore R, Berman F, Schottlaender B, Digital Preservation Lifecycle Management for Multi-media Collections, Lecture Notes in *Computer Science*, November 2005. (3815): 380-384.

Rajasekar A, Moore R, Data and Metadata Collections for Scientific Applications, High Performance Computing and Networking (HPCN 2001), Amsterdam, Holland, June 2001, 72-80.

Rajasekar A, Vernon F, Lindquist K, Orcutt J, Lu S, Moore R, Accessing Sensor Data Using Meta data: A Virtual Object Ring Buffer Framework, DMSN 2005, Trondheim, Norway Aug 2005.

Website: http://drops.dagstuhl.de/portals/index.php?semnr=05271 - Accessed August 20, 2007

REFERENCES

1) Laxminarayan, Swany and Luis Kun. The Many Facets of Homeland Security: An overview from Guest Editors. IEEE Engineering in Medicine and Biology Magazine. January/Feburary 2004. 04:1-11.

2) ARPANET, From Wikipedia Website: http://en.wikipedia.org/wiki/ARPANET - Accessed on August 14, 2007.

3) Auffray C, Imbeaud S, Roux-Rouquie M, Hood L. From functional genomics to systems biology: concepts and practices. C R Biol. 2003 October - November; 326 (10-11): 879-92. PMID.

4) Weston AD, Hood L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. J Proteomics Res. 2004 Mar-April; 3(2):179-96. PMID 15113093.

5) The Globus Alliance Website: http://www.globus.org/grid software/ecology.php Accessed August 13, 2007.

6) Information Technology Laboratory-Computer Security Division, Computer Security Resource Center-CSRC, National Institute of Standards and Technology Cryptographic Toolkit, http://csrc.nist.gov/CryptoToolkit/ - Accessed August 15, 2007.

7) Carr N, The End of Corporate Computing, Management of Information Systems, MIT Sloan Management Review, Reprint 46313; Spring 2005, 46(3):67-73.

8) Gilder G, Telecosm, How Infinite Bandwidth will Revolutionize Our World, 1st Edition: New York, NY: Free Press; 2000.

9) The HealthGrid White Paper Website: http://whitepaper.healthgrid.org/ - Accessed August 14, 2007.

10) Frietas R, Doubling of Medical Knowledge. Website: http://discussforesight.org

11) President's Information Technology Advisory Committee, Revolutinizing Health Care Through Information Technology," June 2004.

12) Klous F, Son, Taim, Roy, Livny and VaDenBrand "Transparent access to Grid resources for user software", Concurrency and Computation: Practice and Experience, 2006, 18:787-801.

13) Strong P, Interesting Times for Distributed Datacenters, Ebay Research Labs, May 14, 2007.
- Accessed August 14, 2007.

14) The TeraGrid, Website: http://www.teragrid.org - Accessed August 14, 2007.

15) L. Hood, J. R. Heath, M. E. Phelps, and B. Lin, "Systems biology and new technologies enable predictive and preventative medicine," Science, vol. 306, pp. 640-3, 2004.

16) CDC definition of Prion Disease Deverse engineering of biological complexity," *Science*, vol. 295, pp. 1664-9, 2002.

17) Csete M, Doyle J, "Reverse engineering of biological complexity," *Science*, vol. 295: 1664-9, 2002.

18) Noble D, "Modeling the heart-from genes to cells to the whole organ," *Science*, vol. 295: 1678-82. 2002.

19) The MIDAS (Models of Disease Agent Study) Website: http://www.nigms.nih.gov/Initiatives/MIDAS/
- Accessed August 17, 2007.

20) The Biomedical Information Science and Technology Initiative, Charge to the working group on Biomedical Computing, June 1999. Website: http://www.nih.gov/about/director/060399.htm - Accessed August 16, 2007.

21) OPASI: Office of Portfolio Anaylsis and Strategic Initiatives. National Institute of Health. NIH Roadmap for Medical Research Website: http://nihroadmap.nih.gov/ Accessed August 15, 2007

22) HealthGrid A Summary - A joint white paper from the HealthGrid Association and Cisco Systems Website: http://whitepaper.healthgrid.org/30260HealthgridWPv5.pdf - Accessed August 16, 2007.

23) Centers for Disease Control: Comprehensive Plan for Epidemiological Surveillance. Atlanta GA, 1996

24) http://scoop.sura.org/- Accessed December 6, 2007.

25) Livnat, Yarden, et al., "Visual Correlation for Situational Awareness",
Website: http://www.sci.utah.edu/publications/yarden05/VisAware.pdf - Accessed 2005.

26) http://en.wikipedia.org/wiki/Metcalfe's_law - Accessed December 7, 2006

RESOURCES

List of grid projects around the world	http://www.ibm.com/developerworks/grid/library/gr-gridorgs/index.html
What is the Grid?	http://www-fp.mcs.anl.gov/~foster/Articles/WhatIsTheGrid.pdf

Southeastern Universities Research Association - SURAgrid	http://www.ibm.com/developerworks/grid/library/gr-gridorgs/index.html
Cyber-enabled Discovery and Innovation Initiative, National Science Foundation	http://www.nsf.gov/news/news_summ.jsp?cntn_id=108366
National Science Foundation Cyberinfrastructure Report, March 2007	http://www.nsf.gov/od/oci/CI_Vision_March07.pdf
Globus Consortium provides Grid Communities of Practice	http://www.globusconsortium.org/about/index.shtml
Open Grid Forum	Meet the international community accelerating grid adoption by providing an open forum for grid innovation and developing open standards for grid software interoperability. http://www.ogf.org/
Globus Alliance	http://www.globus.org/
Globus Toolkit	http://www.globus.org/toolkit/
Grid Today	http://www.gridtoday.com/grid/
IBM on Grid	http://www-03.ibm.com/grid/
International Virtual Data Grid Laboratory	http://www.ivdgl.org/
Service Oriented Science	http://www.sciencemag.org/cgi/content/full/308/5723/814?ijkey=aqCCmCFix8Ll.&keytype=ref&siteid=sci
Global Grid Forum	The objective of the Global Grid Forum is to promote and develop Grid technologies and applications via the development and documentation of "best practices," implementation guidelines, and standards with an emphasis on rough consensus and running code. http://www.gridforum.org/