Guest Editorial
Introduction to the Special Section on BioGrid: Biomedical Computations on the Grid

I. Background

RESEARCH in life sciences increasingly relies on globally distributed information and knowledge repositories. The quality and performance of future computing and storage infrastructure in support of such research depends heavily on the ability to exploit these repositories, to integrate these resources with local information processing environments in a flexible and intuitive way, and to support information extraction and analysis in a timely and on-demand manner. Modern grid technology represents an emerging and expanding instrumentation, computing, information, and storage platform that allows geographically distributed resources, which are under distinct control, to be linked together in a transparent fashion [1], [2]. The power of the grids lies not only in the aggregate computing ability, data storage, and network bandwidth that can readily be brought to bear on a particular problem, but also on its ease of use. Nowadays, the grids have been moving out of research laboratories into early adopter production systems, such as the computational grid for computation-intensive applications, the data grid for distributed and optimized storage of large amounts of accessible data, as well as the knowledge grid for intelligent use of the data grid for knowledge creation and tools to all users. Specifically, recent research in aspects of grid-enabled infrastructures, test beds, management, and security has demonstrated the value of modern grid techniques in support of areas including, but not limited to, the following:

1) computational genomics; computational proteomics;
2) systems biology, biological information integration;
3) storage of biomedical information;
4) retrieval of distributed biomedical information;
5) biomedical modeling and simulation;
6) biomedical image processing and simulation;
7) distributed medical database management/integration;
8) mining and visualization of biomedical data;
9) tele-systems for diagnostic, prognostic, and therapeutic applications;
10) computerized epidemiology;
11) pharmaceutics and clinical trials (CTs);
12) collaborative and proprietary health networks;
13) social health care.

In the late 1990s, the grid was proposed as a distributed computing infrastructure, allowing to couple distributed resources and offer consistent and inexpensive access to resources irrespective of their physical location or access point. It enables sharing, selection, and aggregation of a wide variety of geographically distributed computational resources (such as supercomputers, computing clusters, storage systems, data sources, and instruments, etc.), thus allowing them to be used as a single, unified resource for solving large-scale computing and data-intensive computing applications, a requirement of which is the efficient management and transfer of large amounts of data in distributed computing environments. The main focus of grid technologies has been the definition of protocols for allowing the integration of distributed systems with emphasis placed on heterogeneity, scalability, and fault tolerance. Representative projects include the Globus [3]–[6], Legion [7]–[10], and UNICORE [11]–[14] among several others. Globus and Legion are aimed at building a generic computational grid, with support for specific applications added on. The UNICORE focuses on uniform batch job submission and monitoring. Several regional biogrid initiatives are emerging into global infrastructures.

II. Road Ahead

After more than a decade's research effort, the grids are now becoming a viable solution to certain computation- and data-intensive applications [15]–[17]. However, a few fundamental issues for the success of grid technologies in support for life-science-related research work have not been fully settled although various solutions have been suggested.

A. Ethical Issues

The deployment of grid technologies will inevitably foster the sharing of information from molecular, individual to population levels. Releasing personal genomic data, even with consent, implies a de facto release of information pertaining to related individuals. Protocols generally agreed upon are yet to be worked out. In addition, the uniqueness of personal genotype often renders anonymity of the information source difficult. Strict regulations need to be devised to keep such information from being abused.

B. Interoperability

A fundamental issue for the success of grid technologies supporting health care research and practice will be the interoperability at the levels of health data format, middleware, and the system architectures. At present, these issues have not been settled although various solutions have been suggested, such as the Open Grid Service Architecture (OGSA).[1] The compatibility

Digital Object Identifier 10.1109/TITB.2008.918585

[1]Grid Service Specification, Feb. 2003 (http://www.globus.org).

of diverse security models and the translation of different high-level protocols, which specify actions in the grid, are the critical elements for interoperability.

C. Legal and Liability Issues

The issue in regards to the determination of the person(s)/institute(s) liable in case of medical accidents or errors pertaining to the use of health grid while providing health care to a patient is crucial. For an international virtual organization enabled by the health grid, such issues become far more complicated. As an initial step toward the determination of jurisdiction, the European Union has adopted the Council Regulation (EC) No. 44/2001 of December 22, 2000 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters [18].

D. Security

Most ongoing grid developments have emerged from a high-performance computing context. However, a large number of biomedical applications rely on the sharing and exploitation of large amounts of globally distributed data and information repositories as opposed to computation resources. Besides, data management and replication mechanisms proposed by the current grid middleware mainly deal with flat files. Data access control is handled at a file level. In certain data grid projects, user authentication relies on the asymmetric key-based Globus Grid Security Infrastructure (GSI) layer [19]. File access is controlled through access control lists (ACL). This infrastructure does not take metadata into consideration. Note that metadata plays an important role in the health care database management systems, and an effective abstraction of the health data is essential in its storage, access, organization, and authentication in the health grid environment.

E. Others

In addition to the aforementioned solutions, research problems arising from areas such as data migration strategies for distributed life science data, health grid economics, protocols for sharing of biomedical data, etc. are all awaiting further exploration.

III. About This BioGrid Special Section

All submissions were reviewed by at least three experts. Research results reported in the selected papers include the deployment of modern grid technologies in several important life science applications [20]–[24], as well as the development of novel system frameworks for life science research and practice [25]–[30].

Article [20] combines high-performance computing and grid computing technologies to accelerate multiple executions of a biomedical application that simulates the action potential propagation on cardiac tissues. First, a parallelization strategy was employed to accelerate the execution of simulations on a cluster of PCs. Then, grid computing was employed to concurrently perform the multiple simulations that compose the cardiac case studies on the resources of a grid deployment, by means of a service-oriented approach. Article [21] proposes a solution that adapts the Digital Imaging and Communication in Medicine (DICOM) protocol to the Globus GSI and utilizes routers to transparently route traffic to and from DICOM systems. Thus, all legacy DICOM devices can be seamlessly integrated into the grid without modifications. A prototype of the grid routers with the most important DICOM functionality has been developed and successfully tested in the MediGRID test bed, the German grid project for life sciences. Article [22] presents the application of a component-based grid middleware system for processing extremely large images obtained from digital microscopy devices. Parallel, out-of-core techniques for different classes of data processing operations employed on images from confocal microscopy scanners are developed. Article [23] reports a human neuroimaging collaboratory enabled by the *Biomedical Informatics Research Network* (BIRN). The BIRN has developed a federated and distributed infrastructure for the storage, retrieval, analysis, and documentation of biomedical imaging data. The infrastructure consists of distributed data collections hosted on dedicated storage and computational resources located at each participating site, a federated data management system and data integration environment, an Extensible Markup Language (XML) schema for data exchange, and analysis pipelines designed to leverage both the distributed data management environment and the available grid computing resources. Article [24] deploys the grid technology for a distributed, Internet-based collaboration to address one of the worst plagues of our present world, malaria. The first step toward this vision has been achieved during the summer 2005 on the European Enabling Grids for E-Science in Europe (EGEE) grid infrastructure where 42 million ligands were docked for a total amount of 80 CPU years in six weeks in the quest for new drugs.

Article [25] presents the design and implementation of a semantics enabled service discovery framework in the SIMDAT Pharma Grid, an industry-oriented grid environment for integrating thousands of grid-enabled biological data services and analysis services. The framework consists of three major components: the OWL-DL-based biological domain ontology, OWL-S-based service annotation, and semantic matchmaker based on the ontology reasoning. Built upon the framework, workflow technologies are extensively exploited in SIMDAT to assist biologists in (semi-) automatically performing in silico experiments. The *Domain Ontology Oriented Resource System* (DOORS) and *Problem Oriented Registry of Tags And Labels* (PORTAL) are proposed in Article [26] as infrastructure systems for resource metadata within a paradigm that can serve as a bridge between the original Web and the semantic Web. Internet Registry Information Service (IRIS) registers domain names while Domain Name System (DNS) publishes domain addresses with mapping of names to addresses for the original Web. Analogously, the PORTAL registers resource labels and tags while DOORS publishes resource locations and descriptions with mapping of labels to locations for the semantic Web. BioPORT is proposed as a prototype PORTAL registry specific for the problem domain

of biomedical computing. Article [27] reports on original results of the Advancing Clinico Genomic Trials (ACGT) integrated project focusing on the design and development of a European Biomedical Grid infrastructure in support of multicentric, postgenomic CTs on cancer. Postgenomic CTs use multilevel clinical and genomic data and advanced computational analysis and visualization tools to test hypothesis in trying to identify the molecular reasons for a disease and the stratification of patients in terms of treatment. Article [28] presents a new computational grid architecture based on a hybrid computing model to significantly accelerate comparative genomics applications. Article [29] investigates how grid infrastructure can facilitate high-throughput biological imaging research, and present an architecture for providing knowledge-based grid services for this field. Article [30] describes the requirements for building an automated scalable system (GADU) that can run jobs on different grids. The paper describes the resource-independent configuration of GADU using the Pegasus-based Virtual Data System that makes high-throughput computational tools interoperable on heterogeneous grid resources. The paper also highlights the features implemented to make GADU a gateway to computationally intensive bioinformatics applications on the grid.

This special section will be of great value to those interested in the development, deployment, and evaluation of grid technologies in broadly biology-related research and practice, including developers and users of life science information technology, professionals and researchers in biomedical informatics, computer scientists, health network authorities, and research network representatives.

ACKNOWLEDGMENT

Special thanks go to the reviewers who provided valuable feedbacks throughout the peer-review process. The reviewers for this BioGrid special section are listed as follows.

1) Christopher J. O. Baker
Institute for Infocomm Research, Singapore
2) Howard Bilofsky
PCBI, University of Pennsylvania, USA
3) Vincent Breton
Laboratoire de Physique Corpusculaire de Clermont-Ferrand, France
4) Nicola Cannata
Universita di Camerino, Italy
5) Vipin Chaudhary
State University of New York at Buffalo, USA
6) Susumu Date
Osaka University, Japan
7) Jauvane C. de Oliveira
National Laboratory for Scientific Computing, Brazil
8) Aiguo Du
Georgia State University, USA
9) Gilson A. Giraldi
National Laboratory for Scientific Computing, Brazil
10) Jim Hendler
Rensselaer Polytechnic Institute, USA
11) Andrew Jones
Cardiff University, U.K.
12) Fumikazu Konishi
Riken Genomic Sciences Institute, Japan
13) Hing Yan Lee
National Grid Office, Singapore
14) Yannick Legre
HealthGrid Association, France
15) Wilfred Li
San Diego Supercomputing Center, USA
16) Natalia Maltsev
Argonne National Laboratory, USA
17) Tsutomu Maruyama
Tsukuba University, Japan
18) Hideo Matsuda
Osaka University, Japan
19) Yo Matsuo
OncoTherapy Science, Inc.
20) Richard McClatchey
University of the West of England, U.K.
21) Johan Montagnat
French National Center for Scientific Research, France
22) Breanndán Ó. Nualláin
University of Amsterdam, The Netherlands
23) Silvia D. Olabarriaga
University of Amsterdam, The Netherlands
24) Motonori Ota
Tokyo Institute of Technology, Japan
25) Yi Pan
Georgia State University, USA
26) Steve Robinson
University of Ulster, Northern Ireland
27) Mathilde Romberg
University of Ulster, Northern Ireland
28) Bertil Schmidt
Nanyang Technological University, Singapore
29) Piotr Sliz
HHMI and Harvard Medical School, USA
30) Tony Solomonides
University of the West of England, U.K.
31) Martin Swain
University of Ulster, Northern Ireland
32) Tin Wee Tan
National University of Singapore, Singapore
33) Xiangguo Yan
Xian Jiaotong University, China
34) Longde Yin
University of Connecticut, USA.

We are greatly indebted to the steering committee members of the annual *International BioGrid Workshop*. The success of the workshop series has stimulated the creation of a high-impact journal special section to collate the state-of-the-art research articles from the community. Thanks also go to participants of the round-table discussions of the past two BioGrid events. Scope of this special section was based on those insightful discussions. The 2007 BioGrid steering committee members are listed as follows.

1) Jack Dongarra
University of Tennessee, Knoxville, USA
2) Ian Foster
University of Chicago and Argonne National Lab, USA
3) John Holmes
University of Pennsylvania, USA
4) Russ Miller
State University of New York at Buffalo, USA
5) Haruki Nakamura
Osaka University, Japan
6) Keith Ruskin
Yale University, USA
7) Joel Saltz
Ohio State University, USA.

CHUN-HSI HUANG, *Guest Editor*
Department of Computer Science and Engineering
University of Connecticut
Storrs, CT 06269-2155, USA

AKIHIKO KONAGAYA, *Guest Editor*
Advanced Genome Information Technology Research Group
RIKEN Genomic Sciences Center
Yokohama 230-0045, Japan

VINCENZO LANZA, *Guest Editor*
Buccheri La Ferla Hospital
Fatebenefratelli, Palermo, Italy

PETER M. A. SLOOT, *Guest Editor*
University of Amsterdam
Amsterdam 316 1098 SJ, The Netherlands

REFERENCES

[1] F. Berman, G. Fox, and T. Hey, *Grid Computing: Making the Global Infrastructure a Reality*. New York: Wiley, 2003.
[2] I. Foster and C. Kesselman, *The Grid: Blueprint for a New Computing Infrastructure*. San Francisco, CA: Morgan Kaufmann, 1999. Available: http://www-fp.mcs.anl.gov/~foster/foster-cv03-jw.pdf
[3] W. Allcock, A. Chervenak, I. Foster, L. Pearlman, V. Welch, and M. Wilde, "Globus toolkit support for distributed data-intensive science," in *Proc. Comput. High Energy Phys.*, Sep. 2001.
[4] I. Foster and C. Kesselman, "Globus: A metacomputing infrastructure toolkit," *Int. J. Supercomput. Appl.*, vol. 11, no. 2, pp. 115–128, 1997.
[5] I. Foster, "The grid: A new infrastructure for 21st century," *Phys. Today*, vol. 55, no. 2, pp. 42–47, 2002.
[6] S. Vazhkudai, S. Tuecke, and I. Foster, "Replica selection in the globus data grid," in *Proc. 1st IEEE/ACM Symp. Cluster Comput. Grid (CCGrid)*, 2001, pp. 106–113.
[7] M. Humphrey, "From Legion to Legion-G to OGSI.NET: Object-based computing for grids," in *Proc. 17th Int. Parallel Distrib. Process. Symp. (IPDPS)*, 2003, p. 207.
[8] M. Lewis, A. Ferrari, M. Humphrey, J. Karpovich, M. Morgan, A. Natarajan, A. Nguyen-Tuong, G. Wasson, and A. Grimshaw, "Support for extensibility and site autonomy in the legion grid system project," *J. Parallel Distrib. Comput.*, vol. 63, no. 5, pp. 525–538, 2003.
[9] A. Natarajan, A. Nguyen-Tuong, M. Humphrey, M. Herrick, B. Clarke, and A. Grimshaw, "The legion grid portal," *Concurrency Comput.: Practice Exp.*, vol. 14, no. 13–15, pp. 1365–1394, 2002.
[10] B. White, M. Walker, M. Humphrey, and A. Grimshaw, "LegionFS: A secure and scalable file system supporting cross-domain high-performance applications," in *Proc. ACM Supercomput.*, 2001, p. 59.
[11] J. Almond and D. Snelling, "UNICORE: Uniform access to supercomputing as an element of electronic commerce," *Future Generation Comput. Syst. (FGCS)*, vol. 15, no. 5, pp. 539–548, 1999.
[12] J. Pytlinski, L. Skorwider, V. Huber, and P. Bala, "UNICORE—An uniform platform for chemistry on the grid," *J. Comput. Methods Sci. Eng.*, vol. 2, pp. 369–376, 2002.
[13] M. Romberg, "The UNICORE grid infrastructure," *Scientic Programm.*, vol. 10, pp. 149–157, 2002.
[14] M. Romberg, "The UNICORE Architecture: Seamless access to distributed resources," in *Proc. 8th IEEE Int. Symp. High Perform. Distrib. Comput.*, 1999, pp. 287–293.
[15] R. Butler, D. Engert, I. Foster, C. Kesselman, S. Tuecke, J. Volmer, and V. Welch, "A national-scale authentication infrastructure," *IEEE Trans. Comput.*, vol. 33, no. 12, pp. 60–66, Dec. 2000.
[16] M. L. Green and R. Miller, "Molecular structure determination on a computational and data grid," in *Proc. 4th IEEE/ACM Symp. Cluter Comput. Grid—BioGrid Workshop*, 2004, pp. 320–327, (CD-ROM)
[17] H. Stockinger, A. Samar, B. Allcock, I. Foster, K. Holtman, and B. Tierney, "File and object replication in data grids," in *Proc. 10th IEEE Symp. High Perform. Distrib. Comput. (HPDC)*, 2001, pp. 76–86.
[18] *Health-Grid Whitepaper*, Health-Grid Assoc. Aubire, France, 1994.
[19] R. Butler, D. Engert, I. Foster, C. Kesselman, S. Tuecke, J. Volmer, and V. Welch, "A national-scale authentication infrastructure," *IEEE Trans. Comput.*, vol. 33, no. 12, pp. 60–66, Dec. 2000.
[20] J. M. Alonso, J. M. Ferrero Jr, V. Hernández, J. S. Germán Moltó, and B. Trénor, "A grid computing-based approach for the acceleration of simulations in cardiology," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 138–144, Mar. 2008.
[21] M. Vossberg, T. Tolxdorff, and D. Krefting, "DICOM image communication in globus-based medical grids," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 145–153, Mar. 2008.
[22] V. S. Kumar, B. Rutt, T. Kurc, U. Catalyurek, T. Pan, S. Chow, S. Lamont, M. Martone, and J. Saltz, "Large-scale biomedical image analysis in grid environments," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 154–161, Mar. 2008.
[23] D. Keator, J. Grethe, D. Marcus, B. Ozyurt, S. Gadde, S. Murphy, S. Pieper, D. Greve, R. Notestine, H. J. Bockholt, and P. Papadopoulos, "A national human neuroimaging collaboratory enabled by the biomedical informatics research network (BIRN)," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 162–172, Mar. 2008.
[24] V. Breton, N. Jacq, V. Kasam, and M. Hofmann-Apitius, "Grid added value to address malaria," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 173–181, Mar. 2008.
[25] C. Qu, F. Zimmermann, K. Kumpf, R. Kamuzinzi, V. Ledent, and R. Herzog, "Semantics enabled service discovery framework in the SIMDAT pharma grid," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 182–190, Mar. 2008.
[26] C. Taswell, "DOORS to the semantic Web and grid with a PORTAL for biomedical computing," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 191–204, Mar. 2008.
[27] M. Tsiknakis, M. Brochhausen, J. Nabrzyski, J. Pucacki, S. Sfakianakis, G. Potamias, C. Desmedt, and D. Kafetzopoulos, "A semantic grid infrastructure enabling integrated access and analysis of multilevel biomedical data in support of post-genomic clinical trials on Cancer," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 205–217, Mar. 2008.
[28] A. Singh, C. Chen, W. Liu, W. Mitchell, and B. Schmidt, "A hybrid computational grid architecture for comparative genomics," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 218–225, Mar. 2008.
[29] W. M. Ahmed, D. Lenz, J. Liu, J. P. Robinson, and A. Ghafoor, "XML-based data model and architecture for a knowledge-based grid-enabled problem-solving environment for high-throughput biological imaging," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 226–240, Mar. 2008.
[30] D. Sulakhe, A. Rodriguez, M. Wilde, I. Foster, and N. Maltsev, "Interoperability of GADU in using heterogeneous grid resources for bioinformatics applications," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 241–246, Mar. 2008.

Chun-Hsi Huang received the B.S. degree from the National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., the M.S. degree from the University of Southern California, Los Angeles, and the Ph.D. degree from the State University of New York, Buffalo, all in computer science, in 1989, 1994, and 2001, respectively.

He is currently an Associate Professor in the Department of Computer Science and Engineering, University of Connecticut, Storrs. His current research interests include high-performance parallel computing, life science informatics, cyber infrastructure, algorithm design and analysis, as well as experimental algorithmics.

Akihiko Konagaya received the B.S. and M.S. degrees in informatics science from Tokyo Institute of Technology, Tokyo, Japan, in 1978 and 1980, respectively.

He is the Project Director of Advanced Genome Information Technology Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama, Japan. He joined the National Electrostatics Corporation (NEC) in 1980, Japan Advanced Institute of Science and Technology in 1997, and RIKEN GSC in 2003. His current research interests include large scale bioinformatics with advanced information technologies such as ontology, mathematical simulation, and grid computing.

Vincenzo Lanza received the MD degree from the School of Medicine, University of Palermo, Italy. He has studied medicine, anesthesiology, and cardiology at the University of Palermo, Palermo, Italy. Since 1983, he has been appointed Chief of the Department of Anethesia and Intensive Care at the Buccheri La Ferla Hospital, Palermo, Italy. In 1986, he was appointed as the Chief of Anesthesia and Intensive Care, Buccheri la Ferla Hospital, Palermo, and created a computerized system to manage anesthesia and intensive care unit (ICU) activities. In addition, he is also engaged in using the Internet to connect the anesthesia network as telework to complete the patient file as well as support training anesthesiologists on duty, and is also involved in the Biogrid Project. He is the Editor-in-Chief of the *Journal of Clinical Monitoring and Computing* (Springer-Verlag).

Peter M. A. Sloot received the Ph.D. degree in biocomputing from the Dutch Cancer Institute, Amsterdam, The Netherlands, in 1988.

He studied chemistry and physics at the Dutch Cancer Institute. Since 2001, he has been a Full Professor in computational sciences at the University of Amsterdam, Amsterdam. In 1996, he received a 5 year NNV extraordinary professorship in numerical physics. His current research interests include the theory and application of complex systems through distributed mesoscopic computer simulation, biomedical systems, and is engaged in understanding how information progresses through various spatial and temporal scales. He is currently leading the European Union (EU) ViroLab Project and is also involved in four more EU projects and five National Science Foundation (NSF) Projects.