

COMMUNITY ACTIVITY: WORKSHOP REPORT

CyberGIS—Toward synergistic advancement of cyberinfrastructure and GIScience: A workshop summary

Summary: This community activity report describes the outcomes of a CyberGIS workshop, held in conjunction with the UCGIS 2010 annual winter meeting and sponsored by the National Science Foundation (NSF) Office of Cyberinfrastructure. Over the one and one-half day workshop, a multidisciplinary group of experts from the international communities of cyberinfrastructure, GIScience, spatial analysis and modeling, and several other related scientific domains were brought together for a participatory meeting composed of both small- and large-group settings and to discuss the CyberGIS road map.

1 Introduction

An infrastructure often provides fundamental services to members of a society. Human infrastructures have grown from basic living services, such as road networks introduced in early days of civilization to more advanced services, such as electricity and telephony, introduced during the 20th century. Computing, information, and communication technologies have advanced toward the vision of cyberinfrastructure (CI) [3,13]. Other concepts related to CI include e-Infrastructure and e-Science [9]. Computational capabilities have consistently been improved in the past several decades and supercomputing has been playing essential roles for nurturing computational sciences across numerous domain sciences [18]. Bridging between CI and GIScience promises to advance geospatial sciences and technologies through the synthesis of data-driven and computational science approaches [27].

While Wang [22] has demonstrated a CyberGIS framework—a new GIS modality based on the synthesis of CI, GIScience, and spatial analysis—the emergence of CyberGIS has been rooted in various research areas, such as parallel and distributed GIS, spatial analysis [7, 23, 24, 33], geospatial CI [31, 32], the convergence of GIS and social media [20], and internet and web GIS [16]. In the context of CI, the NSF TeraGrid GIScience Gateway was initiated in 2006 as part of the TeraGrid Science Gateway program to provide an online

collaborative geospatial problem-solving environment for spatial research and education [27, 29]. The TeraGrid GIScience Gateway has been deployed with spatial middleware [25]—namely the GISolve Toolkit [26]—for exploiting spatial characteristics to enhance CI while advancing geospatial problem solving.

Furthermore, in the context of GIScience, Goodchild and Haining [6] envisioned the convergence of geographic information systems (GIS) and spatial data analysis through CI. Goodchild also contributed to the development of the 2006 “Report of the American Council of Learned Societies Commission on CI for the humanities and social sciences: Our cultural commonwealth.” At the University Consortium for Geographic Information Science (UCGIS) 2008 winter meeting, a federal agency and department briefing was organized to address the theme of “GIScience and CI: Making connections.” A series of planning discussions across several months followed the briefing to bring GIScience and CI communities together to explore synergistic advancement of CI and GIScience. The CyberGIS workshop was a natural outcome of these discussions facilitated by multiple leaders of the TeraGrid project and UCGIS, and made possible by support from NSF, with broad impacts on science and technology anticipated [2].

2 Workshop objectives and program

The workshop aimed to underpin fundamental issues of CyberGIS for enhancing CI while contributing to the advancement of the next-generation GIS, particularly with synergistic high-performance, distributed, and collaborative capabilities for geospatial discovery and innovation. The workshop program is summarized as follows.

The workshop combined plenary and keynote addresses with panel sessions and break-out special interest group discussions (Table 1). The keynote was given by Dr. Edward Seidel, NSF Assistant Director for Mathematical and Physical Sciences and the former Director of the NSF Office of Cyberinfrastructure. A collection of leaders in related fields delivered presentations at the workshop (see Appendix). Panel sessions were organized into several general research areas and approaches revolving around the research and development of CyberGIS. Special interest group sessions focused on specific CyberGIS topical themes.

3 Keynote and plenary addresses

3.1 Opening plenary address

The first plenary address was given by Dr. Michael Goodchild who discussed the topic “CyberGIS: What still needs to be done?” Goodchild described progress in interoperable location referencing, development of standards (e.g., web service standards from the Open Geospatial Consortium), and the explosion of citizen involvement as both producers and consumers of location- and place-based information. Goodchild posed the question “How can the ‘cyber’ aspect of CyberGIS advance these developments?” Location is a common key to many layers of information, such as land cover, flood plains, zoning, and parcel information. Spatial joins can be used to link disparate pieces of information that pertain to the same locations. However, one problem is that places can be referred to indeterminately. Data may need resampling before cross-analyses with other datasets can be conducted.

| Keynote and Plenary | Panels | Special Interest Groups and Breakout Discussions |
|---|--|--|
| <ul style="list-style-type: none"> ◇ CyberGIS: What still needs to be done? ◇ Data- and computer-driven transformation of modern science: Update on the NSF Cyberinfrastructure vision—People, sustainability, innovation ◇ Science gateways ◇ GISolve ◇ CyberGIS—Projects, services, and experiences in Australia ◇ CyberGIS—ESRI perspectives | <ul style="list-style-type: none"> ◇ Science drivers for, and applications of CyberGIS ◇ CI and GIScience, where to converge? ◇ Building effective CyberGIS ◇ Needs, strategies, and pitfalls for fostering collaboration and virtual organization | <ul style="list-style-type: none"> ◇ Computational intensity of spatial analysis and modeling ◇ Disaster response and CyberGIS ◇ Geospatial ontology and semantic web ◇ Spatial decision support systems ◇ Cyber-enabled geographic information science ◇ Geospatial middleware, clouds, and grids ◇ Data-intensive geographic information science ◇ High performance computing and CyberGIS |

Table 1: Workshop program components.

Metadata are needed for search and discover services. Interoperability between GIS servers remains a significant challenge.

“Has the case for the importance of CyberGIS been made effectively?” Goodchild asked. Clearly new discoveries and new applications are possible through the integration of disparate datasets, but compelling examples could help move the work forward. There are many examples of the increase in interest in GIS, such as the new understanding of the Battle of Gettysburg that developed based on the application of GIS [10]. “What kind of insights might we expect through the application of high performance CI?” Some examples might include insights gained from simulation-based or agent-based modeling. Goodchild concluded that CyberGIS is a powerful vision with many challenging steps before it can be fully realized, enabling location as a common key, defining GIS functionality, and finding the breakthroughs.

3.2 Keynote Address

Dr. Edward Seidel’s talk titled “Data- and computer-driven transformation of modern science: Update on the NSF CI Vision—People, sustainability, innovation” provided a view of the transformation of modern science through the development and application of CI. While illustrative examples were provided from gravitational physics, Seidel stressed that the same type of CI successful in the study of black holes can be made generically useful for nearly all sciences, such as enabling the prediction of the path of hurricanes. An additional focus important to this workshop was the increasingly interdisciplinary and collaborative approaches needed to solve complex problems. Seidel described the need for teams to form dynamically, particularly when faced with emergency scenarios. The data and software sharing necessitated by these collaborations places requirements on CI. Software, networks,

and collaborative environments, as well as data and service interoperability to enable sharing, must be in place. Social networking is becoming increasingly relevant scientifically.

The increasing complexity of CI, however, is leading to an education crisis: there is so much to learn for solving a given scientific problem through CI. NSF has started working on a foundation-wide framework for 21st century CI (CIF21)—comprehensive, balanced, and integrated [14]. This framework includes: a) high performance computing; b) data and visualization capabilities; c) software, tools, and scientific applications; d) large-scale collaborative facilities and international partners; and e) interconnected campuses and people, with an emphasis on workforce development. Participatory, community-wide task forces in these strategic areas are helping NSF develop plans. The NSF's Office of Cyberinfrastructure, sponsor of the workshop, will take a leadership role in the integrative aspects of CIF21. In summary, CI is changing science; and the NSF is responsive to this transformation. The focus of CI on people, sustainability, innovation, and integration will develop well-linked and longer-term programs that include hubs for innovation and discovery.

3.3 Science gateways

Nancy Wilkins-Diehr spoke briefly about the NSF TeraGrid Science Gateways program. She discussed the convergence of high performance computational resources and GIS. This convergence comes at a key time for both CI and GIScience communities. GIS-based analysis is faced with rapidly increasing amounts of spatial data. The reach of CI is also increasing dramatically through the use of sensors, wireless networks, and increasing availability and decreasing cost of computing and data storage. GIS can also be a conduit through which CI can have a major impact on important societal problems, such as disaster management. A review was presented of TeraGrid resources, with multiple petaflops of computing power, several petabytes of storage capacity, and a varied set of other resources.

The Gateway program began with the recognition of the explosion of digital data and the increasing sophistication of web-based applications [28]. The program has been developing the policies and the software tools to widen supercomputing access through community led web-based interfaces. Several examples of gateways were described related to geospatial research and education, including the Linked Environments for Atmospheric Discovery (LEAD) and PolarGrid (<http://polargrid.org/polargrid>). Wilkins-Diehr concluded her discussion by emphasizing the potential impacts that CI science gateways have to transform the conduct of science, and enhance public understanding of scientific work.

3.4 GISolve: An experimental and synthetic approach to CyberGIS

Dr. Shaowen Wang discussed the GISolve middleware project with a focus placed on "An experimental and synthetic approach to CyberGIS." Wang described the three-pronged emphasis on theory, experiment, and computation in science today, emphasizing that the whole is more than the sum of the parts. GISolve integrates CI environments to support large-scale and collaborative spatial analysis and modeling that can be performed by a large number of users simultaneously. GISolve has been deployed to the TeraGrid, where users can contribute or access online data, perform domain decomposition, and deploy computational tasks to underlying CI resources while providing visual representations of results once the computations are finished. The computational capabilities integrated by GISolve make computationally intensive analyses and simulations feasible, while the



visualization capabilities allow users to simultaneously examine sophisticated simulations and massive datasets [22].

The research and development of GISolve require a synergistic approach across disciplines. Spatial analysis and modeling have applications in many fields—environment, energy, health, and sustainability, to name but a few. Software tools like GISolve serve as a conduit for advancing and creating synergies between CI and GIScience.

3.5 CyberGIS perspectives from Australia

Dr. Bill Appelbe provided an international perspective by outlining his experiences with CyberGIS-related projects in Australia. In Australia, CI is funded collaboratively. Appelbe's organization, the Victorian Partnership for Advanced Computing (VPAC), is the lead for the Australian Research Collaboration Services (ARCS), responsible for the development of CI in Australia. There are several CyberGIS projects in Australia, both community (supporting and engaging users) and infrastructure (helping develop standards and tools) types. National-level community projects include IMOS (Integrated Marine Observing System, <http://www.imos.org.au>), TERN (Terrestrial Ecosystem Research Network), ABIN (Australian Biosecurity Intelligence Network, <http://www.abin.org.au>), Water Information Research and Development Alliance (<http://www.bom.gov.au/water/wirada/>), and the Cooperative Research Center for Spatial Information (<http://www.crcsi.com.au>). Infrastructure projects include the National Data Grid Raster Storage Archive, which uses a range of open source tools such as the GDAL (geospatial data abstraction library).

Appelbe also discussed the challenges for CyberGIS, with a particular focus placed on sustainability, maintainability, and extensibility. Research applications must be professionally engineered, component-based, and need to rely heavily on community inputs if they are to be used and maintained at a production level. It can be helpful to have a host institution rather than a single investigator associated with a project. CyberGIS must be scalable and able to adapt as both data and users increase. CyberGIS applications must be also able to be compatible with a variety of data formats, schemas, and data transfer mechanisms. While data standards can be a moving target, extensive use of discoverable web services can help. Curation and validation of data is critical and real time streams of data can make it more difficult. Finally, CyberGIS applications must provide open access to non-experts.

The demand for CyberGIS research and development is increasing. Appelbe mentioned Australian bushfires and floods in February 2010 as timely examples. Unfortunately, this point has been made even more clearly by the devastating floods throughout Australia in January of 2011. Water management, traffic, security, health, urban planning, and natural resource management are all additional example areas that will benefit from advancement of CyberGIS. Broad interest in CyberGIS will drive the adoption of open-source software, standards, and web services.

4 Panels

4.1 Science drivers and applications of CyberGIS

This panel was moderated by Dr. May Yuan (University of Oklahoma). The panelists included Dr. Luc Anselin (Arizona State University), Dr. David Bennett (University of Iowa),

Dr. Chris Renschler (State University of New York, University at Buffalo), Dr. Barbara Poore (US Geological Survey), and Dr. Ilya Zaslavsky (San Diego Supercomputer Center).

4.1.1 Spatial econometrics workbench

Dr. Anselin described a “spatial econometrics workbench” that is under development. Spatial econometrics is a subset of econometric methods that are concerned with spatial cross-sectional and space-time observations [1]. Variables related to location, distance, and arrangement (topology) are treated explicitly in model specification, estimation, diagnostic checking, and prediction. With a few exceptions, spatial econometric methods are still mostly absent in commercial software for statistical analysis.

The workbench includes support for distributed access to spatial data, spatial data analysis, geovisualization, spatial pattern detection, and spatial regression/process modeling. The core of the workbench consists of GeoDa, OpenGeoDa, and GeoDaSpace, maintained and developed in the GeoDa Center at the Arizona State University. This effort leverages the open source PySAL library for spatial data analysis [19]. Several challenges were noted, including interoperability of methods and models, appropriate metadata, efficient application programming interfaces, flexible scaling to larger problems, and web service performance issues. Surmounting these challenges would require fundamental research that offers new algorithms, with varied support for simulation.

4.1.2 Cyberinfrastructure-enabled agent-based modeling

Dr. Bennett presented a topic titled “CI-enabled agent-based modeling of complex adaptive spatial systems.” He discussed an agent-based model (ABM) for representing a complex adaptive spatial system [21]. Dr. Bennett posed a fundamental question: “How do we untangle the provenance of complex adaptive spatial systems?” To address this question a model would need at least: a) smart parameter sweeping algorithms that detect and explore anomalies in system dynamics; b) ways to interactively follow stored paths or “rewind” the simulation to an intermediate state and then re-launch the simulation with the same or new parameter values; c) mechanisms to capture and visualize the “stack” of precursors that might lead to desired or interesting system states; d) tools to evaluate patterns and driving processes; and e) intelligent algorithms to extract the provenance of complex system dynamics.

Building, exploring, and analyzing such complex parameters will require significant computational capabilities based on CyberGIS. To set the context, he presented a case study that models decision behavior of elk in the Yellowstone National Park [4]. The study reveals the influence of snow dynamics on elk behavior, and that the computational performance of ABM is a function of agent and environment characteristics.

4.1.3 Information Products Laboratory for Emergency Response (IPLER)

Dr. Renschler presented a talk about the Information Products Laboratory for Emergency Response (IPLER). He described the IPLER effort for the Haiti Earthquake, and stressed the importance of modular architecture for the software system producing high-resolution imagery maps.

Renschler then discussed a framework for vulnerability and damage assessment supported by the National Institute for Standards and Technology (NIST). The framework



provides a resilience assessment and decision support process that incorporates traditional mitigation and response feedback loops all within a chaining of events during hazard management. That framework was then contextualized within multiple spatial scales for disaster management. The last related project discussed by Renschler was the NSF Cyber-Enabled Discovery and Innovation Type II project called "VHub: CI for volcano eruption and hazards modeling and simulation." The HUB software technology (<http://hubzero.org>) was adapted to support collaboration among involved researchers.

4.1.4 CyberGIS: Learning from users

Dr. Poore highlighted the need to recognize that CyberGIS will only be widely used if it satisfies the needs of its users. She emphasized how the CI enterprise requires expanded definitions of users and the research methods that are required to study them. CyberGIS has the opportunity to engage active users as potential citizen scientists. Doing so may reach beyond domain scientists, to educators who will use CyberGIS to teach, and to ordinary citizens, who participate in environmental decision making.

Two dominant threads for CI research have emerged from social science studies of CI: CI as large technical systems (LTS) and CI as online communities. A group of historians of science and technology have examined CI as large, mixed social and technical systems. The approach opens up previously invisible infrastructure to scrutiny to reveal how social intangibles, such as organizations, institutions, standards, laws, and markets, are as important as technologies in the success of infrastructure. Transparent, reliable infrastructures are only formed when standardized gateways between local systems develop. Path dependence can mean that early moves toward standardization limit later possibilities. Such studies complement and extend the work in GIScience that has been done in the past ten years on the rise and spread of spatial data infrastructures (SDIs).

Dr. Poore commented that the studies of CI users as domain scientists are essential for a better understanding of CyberGIS [17]. Users are no longer the passive recipients of data from national mapping agencies and the SDIs they have established. Instead they have become the producers of data and knowledge. She contended that any CyberGIS project should include qualitative social science researchers who not only study communities of domain scientists but also focus on the usability of CyberGIS for educators and citizens as well as those who wish to contribute data and information.

4.1.5 CyberGIS in hydrology

Dr. Zaslavsky focused on an ongoing activity called a hydrological information system (HIS) which is part of the activity of the Consortium of Universities for Advancement of Hydrological Systems Information (CUAHSI). The CUAHSI hydrologic information system (CUAHSI-HIS) provides web services, tools, standards, and procedures that enhance access to more and better data for hydrological analysis. HIS web services support access to national datasets such as the USGS National Water Information System (NWIS) and the EPA storage and retrieval system (STORET) in a standard way. Users anywhere with access to the Internet can register their data with CUAHSI HIS to publish data. He presented the service-oriented architecture of the HIS, discussing many of the system components.

Dr. Zaslavsky commented on critical CI issues that he has encountered through his work with HIS, which include: a) efficient management of large volumes of distributed spatiotemporal data; b) understanding and unifying data models across sub-domains of

hydrology; c) development of data exchange standards; and d) community ontology management and curation.

4.2 Cyberinfrastructure and GIScience: Where to converge?

This panel was moderated by Dr. May Yuan (University of Oklahoma) and included the following panelists: Dr. Marc Armstrong (University of Iowa), Dr. Bin Jiang (University of Gävle, Sweden), Dr. Robert Panoff (Shodor Education Foundation), and Dr. Michael Worboys (University of Maine).

4.2.1 Exascale computing, cyberinfrastructure, and geographical analysis

Dr. Armstrong described several motivating cases for projecting the influence of exascale computing on geographical problems. Problems that were previously computed heuristically might now be computed exactly. Hundreds or thousands of candidate solutions to multi-criteria problems might be computed in near-real time with tools to sort and compare results. The fluidity of two fundamental concepts in economic geography—site (the characteristics of a place) and situation (a site in relation to others)—may be much more accurately represented through exascale computing. The ability to provide continuous analysis of incoming streams of data and provide early warning detection and robust methods for false positive indications will benefit from research on exascale computing and geographical analysis. Education, however, clearly lags behind. GIScience students are not being trained with parallelism knowledge, while CI experts and computer scientists are not being trained in geographical analysis. Researchers with training across both areas will be needed to guide future developments.

One motivating example presented by Armstrong demonstrated the need for decision makers to work collaboratively to solve today's complex problems. Collaboration requires low latency information exchange. This can be a significant challenge to CI. One area of convergence might be the development and implementation of a system to address a large geographical optimization problem. These are often complex and ill-defined, multi-criteria, and multi-stakeholder problems that require interactive user responses. Such a problem would require high performance computation, high-speed, low latency networks, access to large data stores, and visualization. The needs would only be intensified if decisions rely on real-time data streams from instruments and sensors.

4.2.2 Data intensive geospatial analysis and computation

Dr. Jiang focused on the characteristics of data-intensive geospatial analysis. He contextualized his discussion within the fourth scientific paradigm: data-intensive scientific discovery [8]. His motivating scientific problem is to gain a better understanding of spatiotemporal patterns of human movement. He then described his specific research project named as "FromToMap" for both route planning and personal navigation. The research is based on a new method for deriving a fewest turns and shortest route between any given two locations using the OpenStreetMap (OSM). Significant data challenges must be addressed, including the tens or hundreds of gigabytes of memory required to hold large graphs of millions of nodes and edges representing the entire Europe road networks.

Given the challenges, Jiang argued that geospatial research should go beyond the data scale of megabytes. Spatial heterogeneity should be considered as normality as fine resolu-

tions are addressed. Such challenges require concerted efforts across the entire geospatial research and education communities on sharing data and software archival.

4.2.3 Pathways to petascale: An education perspective

Dr. Panoff discussed the undergraduate education program through the NSF's Blue Waters petascale computing program. Dr. Panoff emphasized how training has a more immediate focus—"How do I use it now?" "How do I run this code on this machine?" "How do I use the tools?"—versus education where questions include "When am I going to ever use this?" "Why would I want to, what would I learn?" "What are the tools, what can they tell me, why should I care?" Panoff sees both horizontal (collaboration among similar activities) and vertical (collaboration across education levels) integration across training and education.

Panoff distinguishes between computational "science education" and "computational science" education. The fundamentals of each include quantitative reasoning, analogical thinking, and multi-scale modeling. The Blue Waters program focuses on undergraduate education for a number of reasons: development of next generation graduate students, development of next generation K-12 teachers, and preparation of an educated workforce. The focus includes both four-year and two-year colleges. Four-year institutions include flexible interdisciplinary programs and an emphasis on computational science. Two-year colleges reach an increasing number of transitional students as well as many teacher education programs.

4.2.4 Complex and dynamic geospatial systems

Dr. Worboys described a vision of modeling complex and dynamic geospatial systems as distributed computational processes. The production of geospatial data has increasingly gone from centralized approaches to the combination of both centralization and decentralization, exemplified by the wide recognition of volunteered geographic information and sensor networks [30]. This transition enables the modeling to take micro-dynamics into account for a better holistic understanding of macro-dynamics.

To pursue the vision, Worboys identified several areas of research, namely, ontologies and data models, spatiotemporal representation and reasoning, efficient storage and retrieval of data, and effective interaction with real-time and historic geospatial information. He highlighted limited computational power as a roadblock, using a case study of studying three-dimensional spatial fields based on simulations.

4.3 Building effective CyberGIS

This panel was moderated by Dr. Budhendra Bhaduri (Oak Ridge National Laboratory), and included the following panelists: Dr. Baris Kazar (Oracle), Dr. Richard Marciano (University of North Carolina at Chapel Hill), Dr. Zhong-Ren Peng (University of Florida), Dr. Marlon Pierce (Indiana University), and Dr. Robert Raskin (National Aeronautics and Space Administration).

The presentations by Dr. Kazar, Dr. Marciano, and Dr. Peng focused on empirical approaches to building effective CyberGIS. Dr. Kazar addressed the challenges of handling massive high-density point cloud data, such as those acquired using LIDAR (light detection

and ranging) instrumentation. Specifically, these challenges include basic data transformation such as coordinate transformation, analysis, and visualization. Dr. Kazar described a scalable data storage, modeling, and visualization approach to these challenges. Dr. Marciano reviewed the evolution of geospatial technologies and associated applications across multiple decades based on his personal experiences. He emphasized the trend of moving toward user-centered design, and the importance of building robust CyberGIS that will be friendly accessible to massive users. Dr. Peng used multiple transportation system scenarios to describe the case for developing CyberGIS to enable interoperable transportation information and planning systems. In particular, he focused on the formalization of the semantic web for communicating geospatial information and achieving interoperability between heterogeneous transportation information systems.

Dr. Peng's presentation naturally transitioned to the broad topic of semantics in CyberGIS, addressed by Dr. Raskin from a perspective of knowledge representation and reuse. Dr. Raskin discussed the challenge of attaining shared understanding. He argued for a common vocabulary with an emphasis placed on feature descriptions and attributes. He summarized three levels of implementation strategies for semantic interoperability based on standards, languages, and comprehensive ontologies.

Dr. Pierce started his presentation by addressing key characteristics of CyberGIS, including openness (e.g., services, standards, and software), sustainability for enabling sciences, and democracy. He emphasized the importance of developing community-wide CyberGIS architecture and related it to the experience of the NSF FutureGrid project in this regard. Multiple CI-based geospatial applications were highlighted to demonstrate the effectiveness of potential holistic approaches to CyberGIS research and development.

4.4 Collaboration and virtual organization

John Towns from the National Center for Supercomputing Applications (NCSA) moderated this panel. The panelists were Dr. Timothy Nyerges (University of Washington), Ruth Pordes (Fermi National Accelerator Laboratory), Dr. Steven Prager (University of Wyoming), Dr. Daniel Sui (Ohio State University), and Dr. Elizabeth Wentz (Arizona State University).

4.4.1 Virtual organization infrastructure

Dr. Nyerges cited two of the most widely recognized definitions of a virtual organization (VO), one from the NSF ("... group of individuals whose members and resources may be dispersed geographically while the group functions as a coherent unit through the use of CI" [3]) and the second from the authors credited with originating the term ("... collection of geographically distributed, functionally and/or culturally diverse entities that are linked by electronic forms of communication and rely on lateral, dynamic relationships for coordination" [5]).

Nyerges described an example of VO infrastructure requirements that included support for complex workflows involving intense interaction, distributed participants, and substantial computing and data requirements. These needs and requirements were addressed within the context of large-scale coupled models and space-time datasets, taking as an example coastal ecosystem modeling. He ended his presentation by describing work underway by the UCGIS to foster VO infrastructure development.

4.4.2 Open science grid approach to virtual organization

Ruth Pordes first answered the question “What is a virtual organization?” by providing a definition from the Open Science Grid blueprint published in 2005: “a dynamic collection of users, resources, and services for sharing of resources” [15]. In response to the question “What is the difference between a VO and an O?” she answered that an organization holds identity information and responsibility for the individuals, who then become members of VOs, and the legal ownership of resources that are used by VOs. She then asked: “Why do virtual organizations form?” The OSG has found that VOs form as a result of well-agreed upon research or scientific goals and aims. These aims can be domain scoped (e.g., a particular Large Hadron Collider experiment), regionally scoped (e.g., the New York State Grid), or activity scoped (e.g., engagement VO for helping researchers adapt their applications to CI). The VOs form for needs that are almost always evolving, in many cases scoped by (funded) project boundaries.

Next she considered: “How are virtual organizations managed?” The concept of a VO in practice does provide a “middle-tier” for management of groups of individual users, allowing heterogeneous subsets of all users to have common policies and management. Lastly, Pordes considered: “What have been the most prominent issues faced in developing and maintaining VO?” The matching of organizational issues and constraints imposed by the technologies of the CI, with which VOs interact, is the greatest challenge. The distributed infrastructures, such as OSG, regard the VO management and membership as a primary concept when implementing and managing access control to appropriate resources. The support for groups, rather than single users, is embedded in all services and software components. Thus new VO paradigms, e.g., on-demand group creation, management, and use of CI resources and services, have proven difficult to provide and integrate into end-to-end CI.

4.4.3 Fostering collaboration and virtual organization

Dr. Prager started off by attaching global change and the developing world as a particular context to his presentation. Dr. Prager then provided a collection of characteristics of a VO including: a) organizations, individuals, or institutions who share computational resources for a common goal; b) an organized entity that does not exist in any one, central location, but instead exists solely through the Internet; c) a geographically distributed organization with common interests or goals that communicates and coordinates its work through information technologies; and d) organizations of organizations *and* individuals enabled by appropriate information technologies.

Prager addressed the question: “Why do virtual organizations form?” Both top-down and bottom-up approaches were discussed. The top-down approach is defined in terms of common interests, defined stakeholders, oftentimes with a clear mandate (legal or other). The bottom-up approach is based on emergent common interests, emergent stakeholders, and multiple competing “mandates.” Another question addressed was: “What have been the most prominent issues faced in developing/maintaining virtual organizations?” Amongst the prominent issues identified were included: conservative cultures, less open perspectives on data and resource sharing, the availability of bandwidth and computational resources, unequal distribution of capacity, and links to/between spatial data infrastructures. Lastly, he posed the question: “Are CI and VO appropriate for the developing world?” Dr. Prager believes CI and VO are appropriate for the developing world, and sees

a great potential to solve difficult research and education problems by fostering integration and interaction between the developed and developing world.

4.4.4 CyberGIS and metaverse

First coined by Neal Stephenson in his 1992 science fiction novel *Snow Crash*, “metaverse” refers to a fictional virtual world where humans, as avatars, interact with each other and software agents, in a space that uses the metaphor of the real world. Dr. Sui views the rapidly evolving metaverse as a result of several converging technologies. According to the metaverse roadmap report (<http://www.metaverseroadmap.org>), the browser for engaging this metaverse will be based upon a Web that brings together the following four technologies:

- Mirror worlds—digital representations of the atom-based physical world, such as the Google Earth, Microsoft Virtual Earth, NASA’s World Winds, ESRI’s ArcGlobe, USGS’ National Map, and the massive georeferenced GIS databases developed during past decades.
- Virtual worlds—digital representations of the imagined worlds, such as Second Life, World of War Craft, computer games, various cellular automata models, and agent-based models;
- Lifelogging—the digital capture of information about people and objects in the real or digital worlds, such as Twitter, blogs, Flickr, YouTube, and social networking sites such FaceBook or MySpace.
- Augmented reality—sensory overlays of digital information on the real and virtual worlds using head-up displays (HUDs) or other mobile/wearable devices such as cell phones or sensors.

Taking a metaverse perspective on CyberGIS will prompt researchers to think about many challenging issues for years ahead. For example, what kinds of CI are needed to support the converging operations of mirror worlds, virtual worlds, augmented reality, and lifelogging in the metaverse? In addition, the needs for developing new techniques for text mining, imagery/photo synthesis, spatial videography, web crawling, and the semantic web are evident. Today, we still do not have an interoperable approach for processing geospatial information across these four worlds in the emerging metaverse.

4.4.5 Virtual organization in the context of universities

Dr. Wentz presented a perspective on virtual organizations that connected many disciplines across a university environment. A mix of disciplines provides a rich context to clarify research and educational goals, but provides challenges with the many domain languages that are spoken to undertake the research. A transdisciplinary perspective would be helpful to address the challenges. Virtual organizations within classrooms are possible, as students’ roles play diverse disciplinary perspectives.

5 Special interest groups—Breakout sessions

5.1 Computational intensity of spatial analysis and modeling

This session was moderated by Christopher Crosby, and focused on the following questions and themes:

- What is fundamentally different about how we approach GIS problems based on CI?
- What new science questions does CyberGIS allow us to address?
- Will we gain new understanding of geospatial problems with more computing power and richer data?
- How do we adapt CI to better work with geospatial computation and scientific questions?
- What needs to be done to adapt existing GIS approaches (e.g., algorithms, data models, etc.) to take advantage of CI?

Many issues were identified, including:

- How much data do we really need, and how do we partition spatial data for scalable CI-based processing?
- What CyberGIS capabilities are needed to cope with big spatial data?
- What is the nature of processes at the resolution of data we are trying to model?
- How large simulations do we really need to run?
- How do we decompose a spatial problem into a supercomputing environment?
- What can CyberGIS accomplish that we cannot accomplish now?

While addressing the challenges of resolving computational intensity of spatial analysis and modeling, this breakout group had a strong consensus on the opportunities presented by CyberGIS. Geospatial problems are becoming increasingly CI intensive, requiring the overarching dimensions of CI (e.g., data, communication, hardware, software, etc.).

5.2 Disaster management and CyberGIS

The moderator of this discussion group was Dr. Ming-Hsiang Tsou. The group discussed open web applications that address disaster management problems. For example, the Ushahidi web site (<http://haiti.ushahidi.com/main>) supported the Haiti earthquake disaster response with messages in five categories: 1) urgent issues, 2) threats, 3) logistics, 4) responses, and 5) others. Further to this website, the group discussed how OpenStreetMap was used in the context of the Haiti disaster, <http://haiti.openstreetmap.nl/>. The group reported on discussions regarding the following five issues.

1. What are key challenges and opportunities for adopting CyberGIS in disaster management? The group considered: a) the lack of open data access infrastructure for desktop GIS applications; b) how to combine real-time data with other types of data; c) preference for processing data rather than automating processes; d) determining who controls data flows; and e) roles played by volunteers and their information. The group suggested that volunteered geographical information (VGI) play useful roles in sharing information.

2. What are basic use cases of CyberGIS? The group reached the following consensus on:
a) the potential for nurturing the culture to ready day-to-day operations for extreme events; b) a mere focus on responses is not sufficient, as preparation is also important; and c) CyberGIS cannot easily change people's behaviors, but may be valuable to encourage people to contribute information (local intelligence).
3. How can CyberGIS support communication? Several issues were addressed. First, huge amounts of data from different sensors are difficult to represent, manage, and analyze. Coordination among data providers (e.g., federal and state agencies) is a social challenge. Second, there is a lack of standardized web services. Third, greater understanding about spatial decision support for first responders is needed to guide research and development of CyberGIS.
4. What are the key factors of data access? Three factors were identified: a) spatial data certification and authorization for use and release; b) data sensitivity; and c) information at the "right" time and "right" location through the "right protocol."
5. What are the roles of dynamic and temporal data? Three aspects were addressed. First, how to handle time-sensitive information? Second, combining historical data with real-time data (e.g., from sensor networks) is useful. Third, social networks have great potential to complement traditional information management approaches.

In addition, the group posed the following questions: "Is more information better?" "Who should be informed of crisis situation, and when?" and "What information might cause 'panic'?" Finally, the group suggested that while CyberGIS research and development have investigated natural science problems to a reasonable extent, more efforts on social problems are needed for addressing challenges in disaster management.

5.3 Geospatial ontology and semantic web

The moderator for this breakout discussion was Dr. Xuan Shi. The workgroup discussion focused on approaches for constructing ontology and its link to a semantic web. A top-down approach to ontology construction was reported as a good approach for building taxonomy, while a bottom-up approach is appropriate for folksonomy development. Using a taxonomy to represent a hierarchy of meaning provides deep and robust relationships in a single inheritance approach, but oftentimes meanings are descended from multiple sources. A folksonomic approach provides a broad-based and possibly multiple inheritance approach, but is less robust because it does not resolve to a single interpretation of meaning.

A further discussion item involved ontology integration and its importance to CyberGIS. Ontology integration encounters problems since the same knowledge can be represented in different ways. Matching different semantics in different ontological frameworks is difficult, if not impossible. For example, matching ontology representations between web service modeling ontology (WSMO) and semantic markup for web services (OWL-S) remains to be a challenge because of different syntax even though both are focused on the semantics of web services. They use different logic rules for reasoning. Consequently there will be little or no semantic interoperability in the near term. Ontology integration presents a challenge requiring more research.

A further discussion addressed the semantic web and potential links to CyberGIS. Reference was made to a report titled "Rethinking the Semantic Web" [11, 12]. The report suggested that there are fundamental issues with approaches to semantic web and its ability to deal with geospatial ontology. Logic, which forms the basis of the web language for

ontology (OWL), suffers from an inability to represent exceptions to rules and the contexts in which they are valid. This causes particular problems when dealing with coordinate geometry.

5.4 Cyber-enabled geographic information science

Cyber space and technologies have been undergoing dramatic changes in the past decade or so. To name a few examples of trends: parallel computing has become the mainstream of computer architecture; distributed and service-oriented systems have become scalable and ubiquitously accessible; and social media has permeated into people's lives across the globe. The technological and related social changes within these dimensions are interdependent, and empowering and reinforcing each other. For example, the exponential growth of social media environments and users would not be possible without scalable and ubiquitous distributed systems, and increasingly powerful parallel computers.

From the perspective of GIScience, the number and diversity of applications and use cases of GIS have been increasing substantially as well. This trend will likely continue or even accelerate into the foreseeable future.

5.5 Geospatial middleware, clouds, and grids

Sixteen participants contributed to this discussion, addressing the following questions: "What do these terms mean?" "How to build geospatial elements into grids and clouds?" "What aspects of geospatial middleware are (urgently) needed?" and "What are our priorities for research, development, and education of user communities?"

Middleware glues together distributed systems and applications where clouds empower services. Grids achieve effective sharing of, and efficient access to computational resources. Geospatial is a crucial element in clouds, grids, and middleware because location and place matter to nearly every aspect of our digital environment and society. Critical spatial thinking is essential to assure middleware, clouds, and grids deliver their functions intelligently for enhancing our environment and society. One major challenge in this thinking is associated with big spatial data. New geospatial middleware, clouds, and grids need to be developed to realize the potential of big spatial data for supporting scientific innovation and discovery.

End-to-end experimental systems are important for both communities to jointly study new components and how different components will impact on the entire system. Such end-to-end systems would also be valuable to achieve sustainable and long-term research and education goals.

6 Perspectives from US agencies

Several federal agency representatives made valuable contributions to the workshop. Dr. E. Lynn Usery from the USGS moderated and also served as a panellist. Other panelists included Dr. Scott Friendschuh from the NSF Social, Behavioral, and Economic Sciences (SBE) Directorate and Dr. Maria Zemankova from the NSF Computer and Information Science and Engineering (CISE) Directorate.

6.1 Cyberinfrastructure components of the National Map

Dr. Usery discussed the National Map project that has its goal to become the nation's source for trusted, nationally consistent, integrated, and current topographic information available online. The vision for the National Map is that it becomes a national foundation for science, land and resource management, recreation, policy making, and homeland security; and founded on a seamless, continuously maintained, and nationally consistent set of base geographic data developed and maintained through partnerships.

The National Map 1.0 includes eight data layers: transportation, structures, orthoimagery, hydrography, land cover, geographic names, boundaries, and elevation. The National Map provides products and services at multiple scales and resolutions. Future versions will expand to include authoritative data sources at a wider range of scales, e-Topo maps, and ontology-driven access. This progression moves the National Map from a provider of data to that of information and knowledge.

Today, the National Map makes available 15,000 topographic maps through a viewer called Palanterra, a joint development among the National Geospatial Intelligence Agency, ESRI, and USGS. Data are acquired through GPS. Data modeling, structuring, and processing occur using CI components including ontology and semantic tools, databases, and grid and cloud computing. Geospatial data output, generation, and delivery are achieved through softcopy and printed map design, data transfer over the Web, and web 2.0 interactive tools.

Future directions through 2020 include: metadata analysis and enforcement, ontology support, semantic and tool interoperability, semantic web support, multi-resolution, multi-temporal data integration and generalization, rapid transformations on the Web (e.g., projections), data modeling for three-dimensional geospatial and temporal entity representation, knowledge discovery, distributed geospatial data collection or sensing, distributed geospatial data storage, distributed geospatial data analysis, and high-performance computing.

6.2 Funding priorities for research in GIScience

Dr. Freundsuh, Program Director for Geography and Spatial Sciences, described that the past decade has seen strong growth of fields with spatial orientations. GIScience (overlapping with computer science), spatial analysis (overlapping with mathematics), and spatial cognition and behavior (overlapping with behavioral and cognitive sciences) have joined regional science as interdisciplinary fields that are closely aligned with geography.

Work is being encouraged to advance understanding of complex systems by incorporating analyses of the interaction of simpler systems to explain observed complexity. The dynamics of complex systems is an important challenge, for example, "tipping points" where many things change dramatically at one time; and "emergent phenomena," such as phase transitions in which complex phenomena emerge despite being underdetermined by ambient conditioning factors.

Infrastructure development work is encouraged as well. This includes but is not limited to CI, instrumentation, shared databases, repositories, and consortia. NSF-wide, the Cyber-enabled Discovery and Innovation (CDI) program has three focuses: 1) "From data to knowledge" enhances human cognition and generates new knowledge from a wealth of heterogeneous digital data; 2) "Understanding complexity in natural, built, and social systems" derives fundamental insights on systems comprising multiple interacting elements;



and 3) “Building virtual organizations” enhances discovery and innovation by bringing people and resources together across institutional, geographical, and cultural boundaries.

6.3 Spatial activities across NSF

Dr. Zemankova demonstrated the extent to which spatial is an emphasis at the NSF with nearly 2000 active awards in the area. Specific examples were given including a study looking at spatiotemporal flare-ups in volcanoes, spatial or temporal fluctuations at solid-liquid interfaces, and spatial language and cognition.

The CISE Directorate funds projects in spatiotemporal databases, communications, high-performance computing, streaming data, sensor networks, images, videos, complex structures, indexing, organizing, searching, data mining, knowledge discovery, visualization, collaboration, scientific workflows, knowledge sharing, human-computer interaction, and educational technologies. One highlighted project involves geo-temporal decision-making using immersive sensor data streams; another involves visualizing knowledge domains through text mining.

Dr. Zemankova alerted participants to the importance of data science, showing a map color-coded by publications per million inhabitants. A second map showed the prevalence of mental diseases with high rates of each in nearly identical regions. We need to be careful with our tools for collecting, representing, searching, exploiting (analysing), visualizing spatiotemporal data and supporting decisions, deriving new knowledge, driving actions, and sharing results. She concluded her discussion by pointing out a related NSF workshop titled “GeoSpatial and geotemporal informatics,” organized by Peggy Agouris, addressing new challenges in spatial information extraction and modeling.

7 Summary and concluding discussion

In order to capture the momentum of stimulating discussions and crystallize the outcome of the workshop, one participant suggested that, before leaving, attendees propose the top five ideas they saw as most important to move forward for establishing CyberGIS initiatives. With around 50 attendees, this section clearly required a great deal of synthesis. Our reflections resulting from this exercise have been grouped into six primary areas: applications, data, CI access, algorithms, decision support, and education and training. Attendees observed that now was a particularly good time to be focusing on CyberGIS, with one noting that “the US federal administration has increased significantly, and across agencies, support for place-based analysis.”

7.1 Applications and science drivers

More than any other areas, participants commented on the importance of applications to drive CyberGIS. One attendee noted the importance of computation and simulation as a third pillar of science, joining theory and experimentation as described in the 2005 PITAC report titled “Computational science: Ensuring America’s competitiveness” [1]. Another recognized the importance of the formalization of GIS functionalities for interoperable implementation on CyberGIS platforms. A third commented that meaningful applications that move CyberGIS forward will “have the backing of scientific communities, lots of great

computational science and development activities ranging from algorithms to data standards." Applications will need to highlight impacts on society, but must also be well-posed and relevant to researchers.

The importance of data in applications was an oft-repeated observation. Streaming data from in situ sensors, observation networks, volunteers, and other data sources can provide inputs for near real-time spatiotemporal analysis. Aggregating data across scales is a largely unsolved challenge, noted by one participant. For example, one attendee asked what we could learn if we could see "spatiotemporal population distributions at local, regional, and national scales (perhaps even diurnal patterns), real-time population estimates or forecasting, or data assimilation for spatial modeling of geographic events and processes."

High performance analysis and visualization can help researchers gain insights and knowledge from the tremendous amount of spatial data being collected and organized. This data can also be used to validate and extend scientific models. As one attendee stated, in evaluating a driving application, one might look for applications that: a) can be understood in simple and general formulations; b) are able to be refined to a series of synthetic benchmarks of increasing complexity and fidelity; c) are reliant on both historical and current, and ongoing data acquisition; d) at terascale or above in both data and computation for high fidelity models; e) are applicable to real-world problems; and f) are validated against real-world observations.

Urban transport, freight logistics, disaster management, ecosystem modeling, crowd control and management, multi-scale forecasting of crop yields, and water resource management were raised as potential CyberGIS application areas. One attendee cautioned that challenges of this scale would need strong collaboration and funding in order to gain adoption. Some see collaborations among major GIS-related projects such as GEON (<http://www.geongrid.org>) and NEON (<http://www.neoninc.org>) as driving application areas important for the advancement of CyberGIS.

7.2 Data

Attendees agreed that the most challenging problems requiring high performance CI are those that integrate multiple data sources and applications across multiple spatiotemporal scales. To achieve this integration intelligently, fundamental advances in data modeling, categorization, and exchange standards are still needed. Metadata, standards, ontologies and ontology-aware geospatial processing, and provenance were noted as contributing to the ability to share multi-scale datasets across application areas and ensure repeatable results.

One attendee pointed out that "research in [CI] will likely tie in to the ongoing (and burgeoning) research in semantic and ontological approaches." Another recognized the need for "data exchange standards reflecting spatial data models used by scientists."

"Cybercommons" was another top five idea identified by one attendee. A cybercommons would be used to "store science workflows and science narratives (assumptions, data inputs, methods, models, parameters, and outputs) in a database framework so that science workflows and outcomes can be queried, retrieved, and compared." Such recommendations lead directly to research questions such as "What middleware, tools, and languages would be needed to create, combine and access ontologies?" "How does the cybercommons operate in a cloud or grid environment?" and "What tools can be used for provenance man-

agement to ensure that results of scientific workflows can be unambiguously interpreted and replicated?”

7.3 CI access

Many participants identified access to CI as a challenge. One participant saw a major shift in high-performance GIS away from client-server models towards cloud computing. Many recognized that while the needed grid, cloud-based, and high performance computational infrastructure existed, expertise in applying CI approaches to geospatial problem solving is lacking.

One participant noted that platforms such as GISolve can improve access to CI and greatly impact scientists and policy-makers. On the other hand “people need to be cultivated for CI” and additional work might be needed to ensure real uptake of CI-enabled results. The research challenges involve not only user interfaces to such platforms, but also the “provision, discovery, and integration” of underlying computational resources and services. This, in turn, drives application description and design, for example, the development of ontologies and algorithms.

Interoperability was a recurring theme in comments about applications, data, and CI access. Participants noted consistently that data and services provided by CI need to be fully documented to facilitate the integration. As stated previously, “a single core set of standardized geospatial operations needs to be identified, designed, and implemented on top of CI in order to establish a base-line level of utility from which research can be performed and replicated.” The same participant, however, cautioned about enforcing a single standard because of the rapid evolution and unpredictable manner in which standards are adopted.

7.4 Algorithms

Participants noted that advances in spatiotemporal representation are driving a need for improved algorithms. As one participant put it, “novel techniques are needed to uncover complex spatiotemporal relationships among geographic entities and systems,” for example, in making sense of increasing volumes of spatial data.

One participant identified a “culture change” needed in the geospatial communities to utilize high performance and cloud computing in GIS models, algorithms, and databases. Others stressed the need for CyberGIS to hide those complexities with well-designed interfaces for users.

Participants also considered the available computational resources today and their impact on algorithms. From tightly coupled supercomputers to grids and clouds, targeted resources often drive algorithm development. Most felt that fundamental research was still needed to redesign algorithms to expand beyond desktop computing. One participant identified the need to redesign algorithms to remove currently built-in constraints due to lack of memory, storage space, and processing power. The features of target CI platforms would also have to be considered, including the communication limitations of some grids, the memory restrictions of some graphic processing unit (GPU) implementations, and the fast connections and large shared memory available on some supercomputers.

7.5 Decision support

Many participants saw the short turnaround demanded by emergency decision support scenarios as a clear motivation for the use of high-end CI resources. But other issues were also important. One participant pointed out that while the ability to “produce, integrate, and disseminate timely and critical geographic information to decision-makers” has been demonstrated, issues of access remain—“Who can obtain how much data about what, when, in order to prevent panic, allow equal access to important resources (food, medical supplies, and so forth), and ensure that emergency personnel can successfully perform their duties.” While some methods for automating decisions are partially in place, participants noted that they are ill-suited to today’s data volumes and real-time streaming. Scalable quality assurance of data is needed, rather than top-down certification through single government agencies.

Another participant noted a lack of studies on spatial decision support systems (SDSS). While there has been much work in decision support systems, relatively little of this work has been applied to SDSS. A third participant recommended that an investigation into a “grand unified theory” of disaster planning might be worthwhile. A scientific evaluation of disaster planning might be used to determine whether in fact software and systems were the limiting factor for improved responses or whether the problem was more complex.

7.6 Education and training

Workforce development is an underlying theme throughout this report. Greater education and training efforts are needed to understand what the various components of CI provide, what GIS capabilities map well to these components, to write algorithms that take advantage of CI, and to know how to formulate new problems that might now be tackled with CyberGIS. Comments such as “people need to be cultivated for CI” and “utilizing CI will require a culture change” portend a need for significant efforts on education and training.

Several participants at the workshop emphasized the need to provide training in “basic spatial thinking and problem solving.” Students today grow up with extraordinary mapping capabilities: Google Earth, GPS, and location-based services. Because many of these technologies work so well students may assume that all geospatial problems have been solved and that data are available for all areas at all resolutions. “There is still a clear need to explain how the concepts underlying these tools can be used to answer real questions and describe (model) real phenomenon to help shape decisions.” Dr. Zemankova reported a shortage of NSF proposals in these areas. As one attendee observed, “geospatial education must begin to leverage the intrinsic spatial thinking ingrained in these students from the ubiquitous exposure to maps, data, and other geospatial services with which they are bombarded every day.” Many more students need to be educated for thinking about how spatial approaches can better address the problems of society.

Acknowledgments

We thank all of the workshop participants, and our many colleagues in the NSF TeraGrid Project and the University Consortium for Geographic Information Science for insightful discussions. We are also grateful for the assistance by Dr. Anand Padmanabhan within the CyberInfrastructure and Geospatial Information Laboratory on the development of the workshop web site: <http://www.cigi.illinois.edu/cybergis/>, and comments on early drafts

by Dr. Guofeng Cao and Yan Liu of the Lab. This material is based in part upon work supported by the National Science Foundation (NSF) under Grant Number BCS-0846655, OCI-0503697, and OCI-1047916. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

Shaowen Wang
University of Illinois at Urbana-Champaign, USA

Nancy R. Wilkins-Diehr
University of California, USA

Timothy L. Nyerges
University of Washington, USA

References

- [1] ANSELIN, L. *Spatial econometrics: Methods and models*. Kluwer Academic Publishers, 1988.
- [2] ARMSTRONG, M., NYERGES, T., WANG, S., AND WRIGHT, D. Connecting geospatial information to society through cyberinfrastructure. In *Handbook of GIS and Society Research*, T. Nyerges, H. Couclelis, and R. McMaster, Eds. Sage, London, 2011, ch. 6.
- [3] ATKINS, D., DROEGEMEIER, K., FELDMAN, S., GARCIA-MOLINA, H., KLEIN, M., MESSERSCHMITT, D., MESSINA, P., OSTRICKER, J., AND WRIGHT, M. Revolutionizing science and engineering through cyberinfrastructure: Report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure. Tech. rep., National Science Foundation, 2012. Last access on Feb. 12, 2012 at <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>.
- [4] BENNETT, D., AND TANG, W. Modelling adaptive, spatially aware, and mobile agents: Elk migration in Yellowstone. *International Journal of Geographical Information Science* 20, 9 (2006), 1039–1066. doi: 10.1080/13658810600830806.
- [5] DESANCTIS, G., AND MONGE, P. Communication processes for virtual organizations. *Journal of Computer-Mediated Communication* 3, 4 (1998), 693–703. doi:10.1111/j.1083-6101.1998.tb00083.x.
- [6] GOODCHILD, M., AND HAINING, R. GIS and spatial data analysis: Converging perspectives. *Papers in Regional Science* 83 (2004), 363–385. doi:10.1007/s10110-003-0190-y.
- [7] HEALEY, R., DOWERS, S., GITTINGS, B., AND MINETER, M. *Parallel Processing Algorithms for GIS*. Taylor and Francis, Bristol, PA, 1998.
- [8] HEY, A., TANSLEY, S., AND TOLLE, K. *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research, Redmond, Washington, 2009.
- [9] HEY, T., AND TREFETHEN, A. Cyberinfrastructure for e-science. *Science* 308, 5723 (2005), 817–821. doi:10.1126/science.1110410.

- [10] KNOWLES, A., ROUSB, W., ABSBERE, C., FARRELL, L., FEINBERG, A., HUMBER, T., KUZZY, G., AND WIRENE, C. What could lee see at gettyburg? In *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*, A. Knowles and A. Hillier, Eds. Esri Press, 2008, pp. 235–266.
- [11] MCCOOL, R. Rethinking the semantic web, Part I. *IEEE Internet Computing* 9 9, 6 (2005), 86–87. doi:10.1109/MIC.2005.133.
- [12] MCCOOL, R. Rethinking the semantic web, Part II. *IEEE Internet Computing* 10, 1 (2006), 93–96. doi:10.1109/MIC.2006.18.
- [13] NATIONAL, SCIENCE, AND TECHNOLOGY COUNCIL (NSTC). Social, behavioral, and economic research in the federal context. Last access on Feb. 12, 2012 at http://www.nsf.gov/sbe/prospectus_v10_3_17_09.pdf, 2009.
- [14] NATIONAL SCIENCE FOUNDATION (NSF). Cyberinfrastructure vision for 21st century discovery. Last access on Feb. 12, 2012 at <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>, 2007.
- [15] OPEN SCIENCE GRID CONSORTIUM. Open science grid virtual organization. Last access on Feb. 12, 2012 at <http://osg-docdb.opensciencegrid.org/0000/000018/008/OSG%20Blueprintv.13.pdf>, 2005.
- [16] PENG, Z., AND TSOU, M. *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. John Wiley & Sons, Hoboken, New Jersey, 2003.
- [17] POORE, B. Users as essential contributors to spatial cyberinfrastructure. *Proceedings of the National Academy of Sciences* 108, 14 (2011), 5510–5515. doi:10.1073/pnas.0907677108.
- [18] REED, D. Computational science: Ensuring america’s competitiveness, report to the president by the president’s information technology advisory committee. Tech. rep., The National Coordination Office for Information Technology Research and Development (NCO/IT R&D), 2005. Last access on Feb. 12, 2012 at http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf.
- [19] REY, S., AND ANSELIN, L. PySAL: A python library of spatial analytical methods. *Review of Regional Studies* 37, 1 (2007), 5–27.
- [20] SUI, D., AND GOODCHILD, M. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science* 25, 11 (2011), 1737–1748. doi:10.1080/13658816.2011.604636.
- [21] TANG, W., WANG, S., BENNETT, D., AND LIU, Y. Agent-based modeling within a cyberinfrastructure environment: a service-oriented computing approach. *International Journal of Geographical Information Science* 25, 9 (2011), 1323–1346. doi:10.1080/13658816.2011.585342.
- [22] WANG, S. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers* 100, 3 (2010), 535–557. doi:10.1080/00045601003791243.

- [23] WANG, S., AND ARMSTRONG, M. A quadtree approach to domain decomposition for spatial interpolation in grid computing environments. *Parallel Computing* 29, 10 (2003), 1481–1504. doi:10.1016/j.parco.2003.04.003.
- [24] WANG, S., AND ARMSTRONG, M. A theoretical approach to the use of cyberinfrastructure in geographical analysis. *International Journal of Geographical Information Science* 23, 2 (2009), 169–193. doi:10.1080/13658810801918509.
- [25] WANG, S., ARMSTRONG, M., AND BENNETT, D. Conceptual basics of middleware design to support grid computing of geographic information. In *Proc. Second International Conference on Geographic Information Science* (Boulder, CO, 2002), pp. 197–200.
- [26] WANG, S., ARMSTRONG, M., NI, J., AND LIU, Y. GISolve: A grid-based problem solving environment for computationally intensive geographic information analysis. In *Proc. 14th International Symposium on High Performance Distributed Computing (HPDC), Challenges of Large Applications in Distributed Environments (CLADE) Workshop* (2005), IEEE Press, pp. 3–12. doi:10.1109/CLADE.2005.1520892.
- [27] WANG, S., AND LIU, Y. TeraGrid GIScience gateway: Bridging cyberinfrastructure and GIScience. *International Journal of Geographical Information Science* 23, 5 (2009), 631–656. doi:10.1080/13658810902754977.
- [28] WILKINS-DIEHR, N. Science gateways—common community interfaces to grid resources. *Concurr. Comput. : Pract. Exper* 19, 6 (2007), 743–749. doi:10.1002/cpe.v19:6.
- [29] WILKINS-DIEHR, N., GANNON, D., KLIMECK, G., OSTER, S., AND PAMIDIGHANTAM, S. TeraGrid science gateways and their impact on science. *Computer* 41, 11 (2008), 32–41. doi:10.1109/MC.2008.470.
- [30] WORBOYS, M., AND DUCKHAM, M. Monitoring qualitative spatiotemporal change for geosensor networks. *International Journal of Geographical Information Science* 20, 10 (2006), 1087–1108. doi:10.1080/13658810600852180.
- [31] WRIGHT, D., AND WANG, S. The emergence of spatial cyberinfrastructure. *Proceedings of the National Academy of Sciences* 108, 14 (2011), 5488–5491. doi:10.1073/pnas.110305110.
- [32] YANG, C., RASKIN, R., GOODCHILD, M., AND GAHEGAN, M. Geospatial cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems* 34, 4 (2010), 264–277. doi:10.1016/j.compenvurbsys.2010.04.001.
- [33] YANG, C., WU, H., HUANG, Q., LI, Z., AND LI, J. Using spatial principles to optimize distributed computing for enabling the physical science discoveries. *Proceedings of the National Academy of Sciences* 108, 14 (2011), 5498–5503. doi:10.1073/pnas.0909315108.

Appendix: Workshop participants

Bill Appelbe, *Victorian Partnership for Advanced Computing, Australia*; Pankaj Agarwal, *Duke University*; Luc Anselin, *Arizona State University*; Marc Armstrong, *University of Iowa*; David Bennett, *University of Iowa*; Budhendra Bhaduri, *Oak Ridge National Laboratory*; Chris Crosby,

San Diego Supercomputer Center; Cheryl Eavey, NSF Directorate for Social, Behavioral, and Economic Sciences (SBE); Geoffrey Fox, Indiana University; Scott Freundsuh, NSF SBE; Daniel Goldberg, University of Southern California; Michael Goodchild, University of California at Santa Barbara; Qingfeng Guan, University of Nebraska-Lincoln; Diansheng Guo, University of South Carolina; Bin Jiang, University of Gävle, Sweden; Hassan Karimi, University of Pittsburgh; Baris Kazar, Oracle; Steve Kopp, ESRI; Scott Lathrop, Argonne National Laboratory/National Center for Supercomputing Applications; Naicong Li, University of Redlands; Richard Marciano, University of North Carolina at Chapel Hill; Timothy Nyerger, University of Washington; Zhong-Ren Peng, University of Florida; Robert Panoff, Shodor Education Foundation; James Pick, University of Redlands; Barbara Poore, US Geological Survey; Ruth Pordes, Fermi National Accelerator Laboratory; Steven D. Prager, University of Wyoming; Youliang Qiu, University of Florida; Rob Raskin, National Aeronautics and Space Administration; Chris Renschler, State University of New York Buffalo; Edward Seidel, NSF Directorate for Mathematical and Physical Sciences; Xuan Shi, Georgia Institute of Technology; Daniel Sui, Ohio State University; John Towns, National Center for Supercomputing Applications; Ming-Hsiang Tsou, San Diego State University; E. Lynn Usery, US Geological Survey; Shaowen Wang, National Center for Supercomputing Applications/University of Illinois at Urbana Champaign; Elizabeth A. Wentz, Arizona State University; Nancy Wiegand, University of Wisconsin-Madison; Nancy Wilkins-Diehr, San Diego Supercomputer Center; James W. Wilson, James Madison University; Mike Worboys, University of Maine; Ningchuan Xiao, Ohio State University; Chaowei Yang, George Mason University; Xinyue Ye, Bowling Green State University; May Yuan, University of Oklahoma; Ilya Zaslavsky, San Diego Supercomputer Center; Maria Zemankova, NSF Directorate for Computer and Information Science and Engineering.