Year 3 Project Report –
NSF SI2-SSI: CyberGIS Software Integration for
Sustained Geospatial Innovation

Executive Committee

Shaowen Wang, Principal Investigator
Luc Anselin, Co-Principal Investigator
Budhendra L. Bhaduri, Co-Principal Investigator
Timothy L. Nyerges, Co-Principal Investigator
Nancy R Wilkins-Diehr, Co-Principal Investigator
Michael F. Goodchild, Chair of the Science Advisory Committee
Anand Padmanabhan, Project Manager

Major Project Goals

The NSF SI2: SSI CyberGIS project has the following six major goals:

• Engage multidisciplinary communities through a participatory approach to evolving CyberGIS software requirements;
• Integrate and sustain a core set of composable, interoperable, manageable, and reusable CyberGIS software elements based on community-driven and open source strategies;
• Empower high-performance and scalable CyberGIS by exploiting spatial characteristics of data and analytical operations for achieving unprecedented capabilities for geospatial scientific discoveries;
• Enhance an online geospatial problem solving environment to allow for the contribution, sharing and learning of CyberGIS software by numerous users, which will foster the development of crosscutting education, outreach and training programs with significant broad impacts;
• Deploy and test CyberGIS software by linking with national and international cyberinfrastructure to achieve scalability to significant sizes of geospatial problems, amounts of cyberinfrastructure resources, and number of users; and
• Evaluate and improve the CyberGIS framework through domain science applications and vibrant partnerships to gain better understanding of the complexity of coupled human-natural systems.

Guided by these goals, the project in year 3 focuses on enabling representative science drivers through the establishment of the CyberGIS software environment as innovative and reliable tools and services for related science communities. Specifically the following five science areas have been advanced: 1) advanced cyberinfrastructure and geographic information science; 2) bioenergy; 3) emergency management; 4) geography and spatial sciences; and 5) geosciences.
Major Project Activities

The following activities are ordered corresponding to the five science areas above.

The NSF SI2-SSI: CyberGIS project has made substantial progress during the third project year, and has trained seven graduate students and engaged about twenty faculty, staff and postdoctoral researchers. The three interrelated pillars of the CyberGIS software environment, including the CyberGIS Gateway, CyberGIS Toolkit, and GISolve middleware, have been successfully established. The CyberGIS Gateway is focused on providing an online environment for making CyberGIS capabilities accessible to a large number of users for research and educational purposes. The CyberGIS Toolkit on the other hand is focused on loosely coupled scalable geospatial software capabilities within advanced cyberinfrastructure environments such as the NSF Extreme Science and Engineering Discovery Environment (XSEDE). GISolve bridges the CyberGIS Toolkit and Gateway by managing the complexity of interacting with heterogeneous and distributed resources and services of advanced cyberinfrastructure (including for example the NSF XSEDE and the Open Science Grid) and providing standard and friendly interfaces for the integration of scientific applications into the CyberGIS software environment. Through research and development activities on the three pillars of the CyberGIS software environment, the following major activities in five science areas were pursued.

1. Advanced cyberinfrastructure and geographic information science

Advancing the science of CyberGIS enabled the following use cases for improving scientific software and analysis.

- pRasterBlaster, a software tool contributed by the U.S. Geological Survey (USGS) for map reprojection based on parallel computing is integrated into the CyberGIS Toolkit. Our CyberGIS integration approach has made solid progress towards breaking through major computing barriers to data-intensive cartographic transformation for benefiting the USGS National Map program.
- CyberGIS Gateway has achieved interoperable access to cloud-based ArcGIS Online from Esri (a GIS industry leader and a project partner). This interoperability demonstrates the complementarity of two representative CyberGIS modalities and promises a sustainable approach to the emerging CyberGIS software ecosystem.
- A prototype of a participatory approach to community engagement on CyberGIS Gateway has been implemented by integrating the PGIST software from the University of Washington. This integration enables large-scale and high-performance collaboration within CyberGIS Gateway.
- In collaboration with Wendy K. Tam Cho at the University of Illinois, zoning analysis – a challenging problem in political science – is being tackled with a particular focus on developing a scalable computational solution to fine-scale zoning analysis on a large number of geographic units.

2. Bioenergy

In collaboration with Luis Rodriguez at the University of Illinois, a spatial decision support
system is developed to solve a set of complex optimization problems considering multiple objectives and constraints. Specifically, a supply chain optimization problem in the context of locating biomass supply and processing sites/facilities was investigated. A solution based on a simulation model together with uncertainty and sensitivity analyses was developed as part of the CyberGIS-enabled spatial decision support system.

3. Emergency management

Social media data (e.g., from Twitter) are becoming increasingly important in emergency management due to widespread use of location-aware devices. Two separate but related activities were conducted in this area: 1) Global Twitter Heartbeat project, and 2) FluMapper. For the Global Twitter Heartbeat project, we collaborated with Kalev Leetaru at the University of Illinois and the Silicon Graphics International Corp. (SGI) to achieve near-real-time analysis of massive Twitter streaming data through the use of GISolve middleware. In FluMapper, we established an online application to visualize integrated results from two spatial analyses for 1) interactive exploration of spatial distribution of flu spread risk, and 2) an aggregated mapping of movement patterns across multiple spatial and temporal scales for early detection of flu spread. These two activities demonstrate the tremendous power of the CyberGIS software environment for: 1) resolving the computing and data challenges of real-time big data in an efficient and scalable manner; and 2) supporting an interactive multi-user environment through CyberGIS Gateway for spatial analysis of massive social media data.

4. Geography and spatial sciences

- A spatial regression service is prototyped on CyberGIS Gateway by integrating related components from OpenGeoDa software and GISolve middleware.
- Researchers in health geography, public health, and spatial epidemiology use disease mapping, spatial modeling and simulation techniques to understand spatiotemporal characteristics of disease diffusion. Two complementary studies considering multiple spatial and temporal granularities: 1) a parsimonious epidemic agent-based model and 2) a computational approach to mapping environmental health variables were conducted based on the CyberGIS software environment. Through integration with high-end computing resources from the NSF XSEDE and high-throughput computing resources on the Open Science Grid, the CyberGIS software environment enables unprecedented health geographic research, for example, to consider very large spatiotemporal extents and fine-scale spatiotemporal units.

5. Geosciences

TauDEM (Terrain Analysis Using Digital Elevation Models) – a suite of Digital Elevation Model (DEM) tools for the extraction and analysis of hydrologic information from topography datasets – is being integrated with the CyberGIS software environment. Our CyberGIS integration approach has made solid progress, and in collaboration with David Tarboton at the Utah State University and the NSF XSEDE advanced user support group, significant computing barriers have been eliminated to improve the performance and scalability of TauDEM analytical modules benefiting geoscience research.
Specific Objectives Achieved

Most of the project team members have directly contributed to peer-reviewed publications and engaged in software development work. The project leaders have been active to reach out to various communities and research and education groups, and share the project progress by giving invited or keynote lectures.

Specifically, the University of Illinois at Urbana-Champaign group (including Babak Behzad, Guofeng Cao, Yizhao Gao, Hao Hu, Myunghwa Hwang, Yan Liu, Anand Padmanabhan, Eric Shook, Shaowen Wang, and Yanli Zhao) has achieved the following specific objectives:

- Established essential tools and infrastructure to facilitate software integration research and education, and coordinated the overall project activities;
-Produced a book (Springer, accepted) titled “CyberGIS for Fostering Geospatial Discovery and Innovation” edited by Shaowen Wang and Michael Goodchild;
-Established the CyberGIS speaker series at the University of Illinois at Urbana-Champaign and invited more than ten distinguished speakers from advanced cyberinfrastructure, geographic information science, and domain science communities;
-Coordinated technical work across the entire project team through weekly integration teleconferences;
-Visited and hosted visits of multiple project groups and partners for fostering work progress and developing work plans;
-Maintained project-wide integration infrastructure, including a common software repository, CyberGIS wiki, CyberGIS project web site, and a cloud computing environment for facilitating integration, testing and deployment of CyberGIS software components;
-Researched and developed a high-performance geospatial software building, testing, packaging, and deployment framework based on advanced cyberinfrastructure in order to support sustained and streamlined integration of geospatial software elements into the CyberGIS software environment;
-Significantly improved scalability and usability of CyberGIS Gateway by providing seamless access to the NSF XSEDE and the Open Science Grid through GISolve middleware;
-Collaborated with the NSF CTSC (Center for Trustworthy Scientific Cyberinfrastructure) project (http://trustedci.org/cybergis) to understand and enhance cyber-security of the CyberGIS software environment;
-Trained one postdoc and graduated five graduate students (four with Masters and one with PhD). Both the postdoctoral and PhD students have found their tenure-track faculty jobs in research universities;
-Presented at a number of conferences and departmental seminars;
-Successfully integrated a number of software components into CyberGIS Toolkit (e.g., an agent-based model, pRasterBlaster, and parallel PySAL);
-Through service-oriented integration with CyberGIS Gateway, a set of diverse applications/data are accessible to a large number of users (e.g., FluMapper, OpenGeoDA spreg, and OpenTopography); and
• Enhanced GISolve middleware to provide seamless access to a set of heterogeneous advanced cyberinfrastructure resources including the NSF XSEDE, Open Science Grid, and clouds to enable scientific problem solving.

As the science advisory committee chair, Michael Goodchild has actively attended project events, reviewed scientific findings, participated in physical and virtual meetings, and promoted the project through the presentations that he has given in a variety of settings. He also leveraged other support to contribute to the following publications that are beneficial to the project:


The research group at the GeoDa Center of the Arizona State University (including Luc Anselin, Sergio Rey, Wenwen Li, Robert Pahle and Jay Laura) has achieved the following specific objectives:

• Continued development and integration of advanced spatial analytical tools within CyberGIS Toolkit and Gateway;

• Further advanced parallelization and vectorization of algorithms contained in PySAL and GeoDa, specifically with application to map classification and spatial regression model estimation;

• Integrated several spatial analytical methods from PySAL within the Complex Systems Framework (CSF) for spatial decision support;

• Migrated spatial analytical tools from PySAL to a web service infrastructure;

• Developed a framework for metadata and provenance tracking of spatial analytical operations towards the establishment of a Semantically Enabled Spatial Analytical Workbench (SESAW); and

• Presented at a number of conferences and departmental seminars.

The group at the San Diego Supercomputer Center (SDSC) (including Nancy Wilkins-Diehr and Choonhan Youn) has achieved the following specific objectives:

• Deployed LiDAR-based topographic data access and processing services used by CyberGIS;

• Deployed OpenTopography data access and processing services onto a dedicated I/O node on the Gordon supercomputer at SDSC for CyberGIS integration research and development;
• Developed a web-service interface for raster topography (standard DEM) data access; and
• Prototyped an Open Geospatial Consortium (OGC) standard-based web-service interface to TauDEM analytical functions.

The University of Washington group (including Timothy Nyerges and Mary Roderick) has achieved the following specific objectives:

• Established tools and infrastructure for CyberGIS user interface research, coordinating with the University of Illinois at Urbana-Champaign group about community engagement activities;
• Contributed integration support for participatory science aspects of the project;
• Maintained project-specific infrastructure for the Participatory Geographic Information Systems Technology contribution to the project; and
• Graduated one PhD student, and trained another PhD student. The graduated PhD student went to work for Zillow.

Though not directly sponsored by this particular project, the following publications related to the project goals were produced by project team members leveraging other support:


**Significant Results**

The following significant results are described corresponding to the five science areas outlined in major activities. The project team established the CyberGIS software environment, produced more than 20 peer-reviewed publications, and delivered more than 30 keynote and conference presentations during the third project year. A special issue on “CyberGIS: Blueprint for Integrated and Scalable Geospatial Software Ecosystems” of the flagship journal in geographic information science: *International Journal of Geographical Information Science*, edited by the project PI: Shaowen Wang, is in press and a Springer book titled “CyberGIS for Fostering Geospatial Discovery and Innovation” co-edited by Shaowen Wang and Michael Goodchild (the chair of the project science advisory committee) is in the final editing stage. The first release of
CyberGIS Toolkit, providing loosely coupled CyberGIS capabilities through integration with advanced cyberinfrastructure, will be made in September 2013. The advanced analytical, modeling, and computational capabilities of CyberGIS Toolkit is expected to benefit each of the science areas and beyond. A suite of major updates for CyberGIS Gateway with new applications integrated based on significantly improved GISolve middleware are currently under active testing and scheduled for release by the end of the current project year. Through key integration and research activities on CyberGIS Gateway, CyberGIS Toolkit and GISolve middleware, significant findings were enabled in each of the five science areas as summarized below.

1. **Advanced cyberinfrastructure and geographic information science**
   - Through systematic computational intensity analysis using parallel performance profiling tools (e.g., IPM (Integrated Performance Monitoring) and TAU (Tuning and Analysis Utilities)) for scalable high performance computing, major performance bottlenecks were identified and resolved, which led to significant improvement of scalability and performance of pRasterBlaster.
   - Our service-oriented and holistic integration approach to interoperability between two independent online environments (CyberGIS Gateway and ArcGIS Online) is generalizable to achieve interoperability among multiple CyberGIS modalities.
   - Since the zoning problem is an application of the set-partitioning problem that is known to be \(NP\)-complete and computationally challenging, a scalable parallel genetic algorithm (PGA) library was developed as a heuristic approach. To achieve computational feasibility, the PGA library implements an asynchronous migration strategy to exploit massive parallelism and high-performance computing. Computational experiments have shown that the PGA library is efficient and scalable to over 16,000 processor cores with marginal communication cost. Furthermore, this library is capable of efficiently leveraging hybrid supercomputer architecture (e.g., the Stampede supercomputer with CPU processor and Intel Xeon Phi co-processor) in a symmetric way with marginal overhead in library configuration. These advances paved the way for extreme-scale computing of spatial optimization problem solving.

2. **Bioenergy**

A spatially explicit model for biomass supply chain optimization (BioScope) developed by the research lab of Luis Rodriguez at the University of Illinois was integrated with CyberGIS Gateway to study how spatial uncertainty influences modeling results and associated spatial decisions. The integration considers spatial uncertainty propagation (i.e. how uncertainty in model inputs contributes to uncertainty in optimization results) and estimates relative influence of inputs’ variation on corresponding optimization results using a computationally intensive Monte Carlo simulation method. The results were presented in the Science of CyberGIS Session at the 109th annual meeting of the Association of American Geographers, and well received.
3. Emergency management

Through two separate but interrelated activities, we have demonstrated the CyberGIS approach to analyzing location-aware social media data and space-time trajectories embedded therein to discover dynamic spatial patterns for informing emergency management.

• Leveraging the CyberGIS software environment, the Global Twitter Heartbeat project analyzed a real-time stream of ten percent of Twitter's 500 million daily tweets as they were posted. The entire process from ingest to data analysis to map visualization was optimized to be completed within a few seconds.
• Utilizing the interactive analysis and visualization capabilities offered by CyberGIS Gateway, in FluMapper results from two complementary spatial analyses: 1) an exploratory spatial data analysis to identify areas of unusual disease risk, and 2) flow mapping to visualize movement patterns, are integrated across multiple spatiotemporal scales providing insights into unusual disease spread in near real-time.

A CyberGIS approach to handling massive social media data and near-real-time spatial analysis has been shown to be scalable, efficient, and user-friendly. By applying the CyberGIS approach, during Hurricane Sandy, we obtained a firsthand look at how the entire country was talking about, preparing and responding to the storm as it approached, made landfall, and took its course over the U.S. eastern seaboard (e.g., see this YouTube video: http://youtu.be/g3AqdIDYG0c).

4. Geography and spatial sciences

Two interrelated studies were conducted to understand the spatiotemporal characteristics of disease diffusion.

• An environmental health study based on a computational approach to disease mapping and modeling, although scientifically sound and advantageous, cannot be conducted via conventional GIS because of the approach’s critical dependence on significant computational capacity. By exploiting the powerful high-throughput computing resources on the Open Science Grid, GISolve middleware helps resolve the computational intensity challenges of the approach.
• To study the effect of spatial and temporal representations on epidemic spatial modeling, a spatially explicit agent-based model was developed. The model was part of a PhD thesis (see the project products section), which innovatively uses high-performance computing, showed significant impact of both spatial and temporal granularities used by the epidemic agent-based model on disease diffusion dynamics.

5. Geosciences

Through a series of systematic computational intensity analysis and profiling experiments using the NSF XSEDE, major performance bottlenecks were identified and a solution was recommended to the TauDEM team for resolving them. The implementation of the
recommended solution drastically improved TauDEM’s computational performance. Figure 1 shows significant performance improvement using Yellowstone DEM (19,320 x 27,814, 2GB) as input to a function (PitRemove) in TauDEM.

Figure 1. Computational intensity analysis of the PitRemove operation (a) original version of code; (b): new version with the suggested solution incorporated (Input dataset: Yellowstone DEM, 19,320 x 27,814, 2GB)

Figure 2. Spatial distribution of CyberGIS Gateway users across the globe
Key Project Outcomes

Technical achievements and peer-reviewed publications are very strong. Registered users of CyberGIS Gateway from across the globe have exceeded 750 (Figure 2). CyberGIS Toolkit has been established based on a robust integration framework and a set of rigorous strategies for building, testing, packaging, and deployment. The Toolkit includes a collection of loosely coupled CyberGIS software components integrated with advanced cyberinfrastructure, and provides the first geospatial software toolbox to take advantage of high-end cyberinfrastructure capabilities. Through major integration activities focusing on CyberGIS Toolkit, CyberGIS Gateway and GISolve middleware, the following key outcomes ordered corresponding to the five science areas were achieved.

![Figure 3. ArcGIS Online access within CyberGIS Gateway](image)

1. Advanced cyberinfrastructure and geographic information science
   - pRasterBlaster has been successfully integrated into CyberGIS Toolkit.
• A friendly user interface for seamless interoperation between ArcGIS online datasets and services and CyberGIS Gateway has been developed. Figure 3 shows the resulting user interface on the Gateway for accessing ArcGIS Online.
• The Structured Participation Toolkit (SPT) for enabling high-performance collaboration within scientific problem solving environments is being integrated seamlessly into the Gateway and provides Gateway users a platform to communicate and coordinate their activities.
• Leveraging the Parallel Genetic Algorithm (PGA) library and GISolve Open Service APIs, it is promising to solve the NP-complete zoning analysis problem at fine spatial scales. The scalable PGA library has been evaluated on four XSEDE supercomputers with desirable performance achieved in both strong and weak scaling tests.

2. Bioenergy

The BioScope model was integrated within the CyberGIS Gateway (see Figure 4) to support spatial decision-making. More than 2,500 computationally intensive model runs were conducted. These model executions, by considering spatial uncertainty and conducting sensitivity analysis, assist decision makers to comprehensively explore the input parameter space and help estimate the influence of inputs’ variation on corresponding optimization results. A graduate student working on this project participated in a summer internship program at the Oak Ridge National Laboratory (ORNL, a project partner) to potentially expand the collaborative work in this domain.

Figure 4. CyberGIS Gateway user interface for BioScope-based spatial decision-making
3. Emergency management

Two interrelated projects: The Global Twitter Heartbeat and FluMapper demonstrate the utility of location-aware social media data for emergency management.

- FluMapper user interface and visualization components were successfully integrated into CyberGIS Gateway (see Figure 5), with the computing- and data-intensive analyses leveraging advanced cyberinfrastructure via GISolve middleware.
- The Global Twitter Heartbeat project is able to create the first ever near-real-time combined population, tone and geographic analysis and heat map visualization of massive Twitter streaming data.
  - During the Supercomputing 2012 Conference (SC'12), on the conference exhibition floor, live up-to-the-moment heat maps (of the U.S and world) were created and made available to demonstrate the CyberGIS visual analytics.

These outcomes demonstrate the significant capabilities of CyberGIS to tackle computing- and data-intensive challenges while highlighting the tremendous broader impacts of this NSF project.

![Figure 5. CyberGIS Gateway user interface for the FluMapper application](image)

4. Geography and spatial sciences

- The spatial regression module of OpenGeoDA is enhanced in terms of computational scalability and can be made widely accessible simultaneously by a large number of users through integration with CyberGIS Gateway.
- The health geography and computational science research led to new fundamental understanding about spatiotemporal characteristics of disease diffusion and distribution.
A computational approach to disease mapping and modeling in an environmental health study used the CyberGIS software environment to enable researchers to solve problems that would be infeasible based on conventional GIS approaches.

A computational study of the diffusion of influenza like illness using agent-based modeling of the conterminous United States simulated tens of millions of agents across multiple spatial and temporal scales. Using high-performance CyberGIS approaches, this study was able to systematically show the effect of spatial and temporal granularities on simulation results filling a significant knowledge gap in epidemic modeling. This work is part of a finished PhD thesis.

5. Geosciences

Computing- and data-intensive TauDEM analytical functions have been successfully integrated into the CyberGIS software environment and integration work is underway to make these functions accessible to a large number of users through CyberGIS Gateway.

Education, Training and Professional Development Activities

CyberGIS Gateway has been used by hundreds of undergraduate and graduate students to learn advanced cyberinfrastructure, service-oriented computing, GIS, and spatial analysis and modeling on multiple campuses. Six courses conducted during the Summer and Fall of 2012 and the Spring of 2013 at the University of Illinois, University of North Carolina at Charlotte, and Idaho State University, including one online course, used the CyberGIS software environment. Students involved gained concrete and synergistic understanding about advanced cyberinfrastructure, CyberGIS, and related scientific problem solving.

A number of students associated with the project achieved their education and research goals and have moved on to new opportunities in academia and industry. At the University of Illinois, Eric Shook advised by Shaowen Wang, graduated with PhD in Geography and Geographic Information Science and went on to a tenure-track assistant professor position in the Department of Geography at the Kent State University. Four additional students were also advised by Shaowen Wang and graduated with Masters degree (three in Geography and Geographic Information Science, and one in Computer Science co-advised with Marc Snir). A postdoctoral researcher – Guofeng Cao – working on the project at the University of Illinois has accepted a tenure-track assistant professor position in the Department of Geosciences at the Texas Tech University. At the University of Washington, Timothy Nyerges, graduated one PhD student, and the student went on to a position in industry.

We have closely worked with our project partners to create opportunities for project members to pursue advanced training and professional development. For example, Hao Hu – a PhD student at the University of Illinois had his summer internship experience at the Oak Ridge National Laboratory (a project partner), by contributing to the development of CyberGIS applications in the domain of bioenergy.

During this project year, we continue to run weekly teleconferences for coordinating technical work across the entire project. All of the technical staff, and postdoc and graduate students are
required to participate in these teleconferences coordinated by the project manager – Anand Padmanabhan. These teleconferences have been extremely useful for technical coordination in the team, and associated documentation developed on the project wiki as a result of the teleconferences plays an important and effective role in engaging our project members, federal agency and industry partners, and related communities through the open and integrated CyberGIS software environment. Learning from each other about effectively creating CyberGIS software for scientific research and education additionally helps promote collaborative work and develop training materials for broad scientific communities.

A number of project members delivered hands-on tutorials. For example, Sergio Rey and Luc Anselin taught a tutorial on PySAL and GeoDaSpace at the 7th International Conference on Geographic Information Science, September 19-21, 2012, Columbus, Ohio, USA. Sergio Rey taught a workshop on PySAL at the Open Source Geospatial Research and Education Symposium, October 24-26, 2012, Yverdon-le-Bains, Switzerland. Two tutorials on GISolve and SimpleGrid (an open-source toolkit adapted from GISolve for learning science gateways) were successfully delivered by the University of Illinois group at the 1st International Conference on Space, Time, and CyberGIS and NSF XSEDE’12 annual conference. At the annual CyberGIS All Hands Meeting to be held in September 2013, the project team will conduct a half-day “CyberGIS in Action” hands-on and demo session to help the meeting attendees learn CyberGIS and related scientific problem solving.

**Community Engagement Activities**

Shaowen Wang chaired the organizing committee for the CyberGIS Symposium at the 109th annual meeting of the Association of American Geographers. This symposium, which was attended by more than sixty speakers and attracted several hundred participants, included 17 panel and paper sessions to facilitate the sharing of the state of the art and discussions about opportunities and challenges, cutting-edge progress, and major trends and impacts of CyberGIS. At the 7th International Conference on Geographic Information Science (GIScience 2012), a panel titled “CyberGIS and Big Data” was convened with leading experts including several members of the project: Shaowen Wang (as co-organizer), and Luc Anselin, Budhendra Bhaduri, and Dawn Wright (Esri – a project industry partner) (as panelists). Project members: Luc Anselin and Sergio Rey (Senior Personnel) participated as expert panelists, on a panel titled “Space-Time Research in GIScience” at GIScience 2012. Shaowen Wang also organized the CyberGIS Speaker Series to bring in leading thinkers and practitioners in synergistically advancing cyberinfrastructure and geographic information science to the University of Illinois at Urbana-Champaign to conduct public lectures and engage with the campus community.

Founded by the project members, the 1st International Conference on Space, Time, and CyberGIS (http://www.cigi.illinois.edu/cybergis12) held at the University of Illinois at Urbana-Champaign during August 6–9, 2012 was a great success. The conference attracted more than one hundred attendees from across the globe, and included technical training, community engagement, plenary and panel sessions, and strategic planning. The presentations at the conference have been archived for online access, and continue to receive positive feedbacks from a variety of communities.
Timothy Nyerges co-organized a workshop about sustainability learning communities and sustainability information science at the annual symposium of the University Consortium for Geographic Information Science at George Mason University, Falls Church, VA, USA, on May 20, 2013. The workshop grew out of discussions on CyberGIS for sustainability information science conducted at the 1st International Conference on Space, Time, and CyberGIS. He also chaired and co-organized the Panel Session on CyberGIS and Sustainability Sciences – Making Connections, the Association of American Geographers 109th Annual Meeting, April 9-13, 2013, Los Angeles, California, USA.

Nancy Wilkins-Diehr contributed to the following community activities related to the project:

- Steering Committee, Science Executive Education, led by Nick Berente at the University of Georgia; and
- General Chair, NSF XSEDE’13, July 22-25, 2013, San Diego, CA, USA.

Wenwen Li co-organized a Polar Cyberinfrastructure Session at the American Geophysical Union’s 45th annual Fall Meeting, at San Francisco, California, USA on December 3-7, 2012, and is co-organizing a NSF workshop on Cyberinfrastructure for Polar Science, at the University of Minnesota, MN, USA on September 9-11, 2013.

Our project members have also given numerous conference presentations and invited lectures related to the project during this project year with the following list as a selected set of examples of invited keynotes and lectures.

Invited presentation given by Luc Anselin

- “Towards a Semantically Enabled Spatial Analytical Workbench (SESAW).” Research Award Keynote Address at the 2013 Symposium of the University Consortium for Geographic Information Science, George Mason University, Fairfax, VA, May 21-23, 2013

Invited presentations given by Michael Goodchild


Invited presentations given by Sergio Rey

- “Exploratory space-time data analysis: Distributional and computational issues.” CyberGIS Speaker Series, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, March 22, 2013
- “Open source spatial data analysis: Lessons for education and research from PySAL.” Keynote Address at Open Source Geospatial Research and Education Symposium, Yverdon-les-Bains, Switzerland. October 24-26, 2012.
Invited presentations given by Shaowen Wang

- “CyberGIS for Enabling Data-Intensive Geospatial Research and Education.” Keynote Address at the 2013 CUAHSI Conference on Hydroinformatics and Modeling, Logan, Utah, USA, July 18, 2013
- “A CyberGIS Research Agenda.” Clyde Kohn Colloquium, Department of Geography, the University of Iowa, Iowa City, Iowa, USA, April 26, 2013
- “Exploring the Frontiers of CyberGIS.” Jack and Laura Dangermond Endowed Chair Colloquium, Department of Geography, University of California, Santa Barbara, Santa Barbara, California, USA, February 26, 2013

Future Activities/Plans

We plan to focus on the following set of major activities and objectives:

- Further expand research activities on the science of CyberGIS;
- Achieve new breakthroughs in the existing science areas;
- Enable new science communities who can benefit from accessing CyberGIS software capabilities;
- Further engage underrepresented and minority groups;
- Further engage related international communities and organize the second International Conference on Space, Time, and CyberGIS;
- Add new capabilities to the CyberGIS software environment based on community requirements;
- Engage related communities for developing CyberGIS education materials and strategies;
- Enhance research and education experience for undergraduate students; and
- Develop a campus-wide multi-disciplinary CyberGIS center at the University of Illinois for enabling data-rich sciences and applications.

Journal Publications


**Book**


**Book Chapters**

• Li, W., Goodchild, M.F., Church, R.L. and Zhou, B. (2012) Geospatial data mining on the web: Discovering locations of emergency service facilities. *Lecture Notes in Artificial Intelligence*, 7713: 552-563


**Thesis/Dissertations (Selected)**

• Babak Behzad. (2013) *AUTO-TUNED OPTIMIZED PARALLEL I/O FOR GISCIENCE AND SPATIAL APPLICATIONS*. University of Illinois at Urbana-Champaign

• Su Y. Han. (2012) *A CYBERGIS ENVIRONMENT FOR SPATIOTEMPORAL ANALYSIS: A CASE STUDY OF CHINA MORTALITY DATA*. University of Illinois at Urbana-Champaign
Eric Shook. (2013) *A COMPUTATIONAL APPROACH TO UNDERSTANDING SPATIAL AND TEMPORAL GRANULARITIES IN AGENT-BASED MODELING*. University of Illinois at Urbana-Champaign

**Conference Papers and Presentations (Selected)**

- Cao, G., Wang, S., and Guan, Q. (2012) *A state-space model for understanding spatial dynamics represented by areal data*. The 7th International Conference on Geographic Information Science. Columbus, Ohio, USA
- Li, W. (2013) *A hybrid indexing and ranking approach to enhance geospatial semantic search*. The 12th International Conference on GeoComputation. Wuhan, China
- Rey, S. J. (2013) *Code as text: Open source in academia*. The 12th Annual Conference on Scientific Computing with Python. Austin, TX, USA
- Rey, S. J. (2013) *Fast algorithms for a space-time concordance measure*. Western Regional Science Association. Santa Barbara, CA, USA
• Shook, E., Leetaru, K., Cao, G., Padmanabhan, A., and Wang, S. (2012). Happy or not: Generating topic-based emotional heatmaps for culturomics using cyberGIS. The 8th IEEE International Conference on eScience. Chicago, IL, USA

Technologies

• CyberGIS Toolkit provides access to scalable geospatial software capabilities within advanced cyberinfrastructure environments such as the NSF XSEDE. It is composed of a set of loosely coupled components to focus on exploiting high-performance computing resources.

• CyberGIS Gateway is an online problem-solving environment for a large number of users to access CyberGIS analytics, tools, and services. Gateway provides transparent access to advanced cyberinfrastructure resources including the NSF XSEDE, Open Science Grid, and cloud environments. Through friendly user interfaces, the Gateway makes CyberGIS capabilities accessible to a large number of users for various research and learning purposes.

• GISolve is the leading spatial middleware that integrates advanced computing and information infrastructure with geographic information system capabilities for computationally intensive and collaborative geospatial problem solving. It is openly accessible via a suite of open service Application Programming Interfaces, and has been a key enabler for the NSF CyberGIS Science Gateway.

Project Resources

GISolve Middleware
URL:  http://sandbox.cigi.illinois.edu/home/doc/gosapi/GISolveOpenServiceAPI.html
Description: Geospatial data are collected and analyzed to support scientific investigations and decision making in a wide variety of application domains (e.g., ecological and environmental sciences, transportation, public health, and business). Enormous computational resources are needed to store and manage geospatial data and derived information for such diverse purposes, and to conduct computationally intensive and collaborative geospatial analysis and problem solving. This type of analysis and problem solving represent a main area of geographic information science (GIScience), an interdisciplinary field involving geography and spatial sciences, computer science, geodesy, and information sciences to study scientific and technological issues about the development and use of geographic information systems (GIS). The purpose of GISolve is to provide user-friendly and spatially intelligent capabilities for performing computing- and data-intensive geospatial data analysis and problem solving, and help a large number of cyberGIS users directly benefit from accessing advanced cyberinfrastructure capabilities.

NSF CyberGIS project web site
URL:  http://cybergis.org
Description: Homepage of the NSF CyberGIS project, containing information about project products, publications, presentations, news and events.
The First International Conference of Space Time and CyberGIS (CyberGIS’12)
URL:  http://www.cigi.illinois.edu/cybergis12/
Description: Online materials for the 1st International Conference on Space, Time, and CyberGIS chaired by the project PI -- Shaowen Wang, including presentations and videos from keynote addresses, plenary and research talks, and panel sessions.

CyberGIS-PGIST web site
URL:  http://depts.washington.edu/pgist/cybergis
Description: Homepage of the Participatory Geographic Information Systems Technologies (PGIST) software – a CyberGIS software element.

PySAL web site
URL:  http://pysal.org
Description: Homepage of PySAL, an open source python library for spatial analytical functions, and a CyberGIS software element. PySAL provides a suite of spatial analytical methods that developers can incorporate into their own application development, and that spatial analysts may customize to further their research and is increasingly used in geographic information systems.

CyberGIS Gateway
Description: CyberGIS Gateway (https://gisolve2.cigi.illinois.edu/home/) - an online problem-solving environment for a large number of users to access CyberGIS tools and services.

CyberGIS Sandbox
Description: A development version of CyberGIS Gateway (http://sandbox.cybergis.org), uses the newest technologies, innovative spatial analysis and modeling methods, and novel applications.

CybeerGIS project wiki
Description: The project wiki (http://wiki.cybergis.org/) used to coordinate, document and disseminate project-related technical information

CyberGIS documentation
Description: Two comprehensive user guides released 1) CyberGIS Gateway user guide (https://wiki.cigi.illinois.edu/display/DOC/CyberGIS+Gateway+User+Guide); and 2) GISolve Open Service API user guide (https://wiki.cigi.illinois.edu/display/DOC/GISolve+Open+Service+API+User+Guide)

CyberGIS code repository
Description: CyberGIS SVN code repository (https://svn2.cigi.uiuc.edu:8443/), includes core software developed within the CyberGIS project (including CyberGIS Gateway, CyberGIS Toolkit, and GISolve middleware) and related software.

PySAL code repository
Description: Open source github code repository (https://github.com/pysal/pPysal) for pPySAL – a parallel python library for spatial analytical functions – developed as part of the project.
**OpenTopography code repository**
Description: An open source OpenTopography community tool repository for accessing and analyzing high-resolution, earth science-oriented, topography data ([https://sourceforge.net/projects/otforge/](https://sourceforge.net/projects/otforge/)) – a CyberGIS Gateway software element.

**pRasterBlaster code repository**
Description: Open source pRasterBlaster github code repository ([https://github.com/dmm/prasterblaster](https://github.com/dmm/prasterblaster)) contains code/library developed by USGS and the CyberGIS project team, and is designed as part of CyberGIS Toolkit to support computationally intensive map reprojection.

**Project Personnel**

<table>
<thead>
<tr>
<th>Name</th>
<th>Project Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yan Liu</td>
<td>Other Professional</td>
</tr>
<tr>
<td>Myunghwa Hwang</td>
<td>Other Professional</td>
</tr>
<tr>
<td>Babak Behzad</td>
<td>Graduate Student (research assistant)</td>
</tr>
<tr>
<td>Yizhao Gao</td>
<td>Graduate Student (research assistant)</td>
</tr>
<tr>
<td>Hao Hu</td>
<td>Graduate Student (research assistant)</td>
</tr>
<tr>
<td>Eric Shook</td>
<td>Graduate Student (research assistant)</td>
</tr>
<tr>
<td>Yanli Zhao</td>
<td>Graduate Student (research assistant)</td>
</tr>
<tr>
<td>Sergio Rey</td>
<td>Co-Investigator</td>
</tr>
<tr>
<td>Wenwen Li</td>
<td>Faculty</td>
</tr>
<tr>
<td>Robert Pahle</td>
<td>Staff Scientist (doctoral level)</td>
</tr>
<tr>
<td>Jay Laura</td>
<td>Graduate Student (research assistant)</td>
</tr>
<tr>
<td>Choonhan Youn</td>
<td>Other Professional</td>
</tr>
<tr>
<td>Mary Roderick</td>
<td>Graduate Student (research assistant)</td>
</tr>
<tr>
<td>Xuan Shi</td>
<td>Faculty</td>
</tr>
<tr>
<td>Luc Anselin</td>
<td>Co PD/PI</td>
</tr>
<tr>
<td>Budhendra L. Bhaduri</td>
<td>Co PD/PI</td>
</tr>
<tr>
<td>Michael Goodchild</td>
<td>Faculty</td>
</tr>
<tr>
<td>Timothy L. Nyerges</td>
<td>Co PD/PI</td>
</tr>
<tr>
<td>Anand Padmanabhan</td>
<td>Staff Scientist – Project Manager</td>
</tr>
<tr>
<td>Nancy R Wilkins-Diehr</td>
<td>Co PD/PI</td>
</tr>
<tr>
<td>Shaowen Wang</td>
<td>PD/PI</td>
</tr>
<tr>
<td>Michael Finn</td>
<td>Project Partner</td>
</tr>
<tr>
<td>David Mattli</td>
<td>Project Partner</td>
</tr>
<tr>
<td>E. Lynn Usery</td>
<td>Project Partner</td>
</tr>
<tr>
<td>Guofeng Cao</td>
<td>Postdoctoral (scholar, fellow or other postdoctoral position)</td>
</tr>
</tbody>
</table>
Project Partners

- Open Science Grid, https://www.opensciencegrid.org

Project Impacts

Impact on the development of the principal project discipline(s)

During the third year of this five-year project, we have made significant progress on advancing geographic information systems and science, spatial analysis and modeling, and a suite of domain science areas through novel computational science and advanced cyberinfrastructure approaches. CyberGIS Gateway, a broad approach to CyberGIS software integration research and development, represents diverse best practices of scientific and technology research along with education and training. CyberGIS Toolkit, a deep approach to CyberGIS software integration research and development, is focused on developing and leveraging innovative computational strategies needed to solve significant geospatial scientific problems by exploiting high-end cyberinfrastructure resources. The two complementary modalities – Gateway and Toolkit – play an important role in driving multi-dimensional collaborative work within the project, and between the project and various external partners and research groups while assuring the results of dynamic collaborative work will continue to enrich the CyberGIS software environment for significant benefits to interdisciplinary research and education.

Within the the current project year, our multi-institution team continues to mobilize a diverse set of groups to share technical advances and expertise, and has successfully established a process for integrating software elements into the CyberGIS software environment. This has resulted in new software capabilities that the entire project team, partners, and geospatial communities can experiment based on advanced cyberinfrastructure, and gain concrete and in-depth understanding about sustainable CyberGIS software environments driven by resolving significant scientific challenges.

Many intellectual products and publications produced by the project team members have highlighted the importance and innovative nature of CyberGIS. The concrete advances that contribute to advanced cyberinfrastructure, geographic information science, and geo and spatial disciplines have demonstrated the transformative effects of CyberGIS software research and education.

Impact on other disciplines

This project is beneficial to a broad set of scientific disciplines by creating new or enhancing existing geospatial software capabilities for tackling significant data-intensive and computational challenges. For example we have initiated investigations with public health research
communities through collaboration with a research team led by Dr. Xun Shi from Dartmouth College to enable computationally intensive disease mapping. Our data-intensive social media analysis research has involved multiple social science and public health communities. We also engaged the hydrology community through active collaboration with a team led by David Tarboton from the Utah State University. Furthermore, by leveraging broad partnerships, we have active interactions with advanced cyberinfrastructure, computer science, environmental science and engineering, political science, and statistics communities to make our software capabilities and infrastructure available to these communities while benefiting from their inputs and feedbacks. The nature of our project in terms of crosscutting advanced cyberinfrastructure and many domain science communities that the project enables has demonstrated tremendous value of advanced cyberinfrastructure-based integration-oriented geospatial software ecosystems to various science communities.

**Impact on the development of human resources**

Our project has continued to place a major emphasis on education and workforce development, which is important to the future development of both advanced cyberinfrastructure and geospatial sciences and technologies. Six graduate students and one postdoctoral researcher have finished their training enabled by the project, and moved on to their career opportunities. For example, at the University of Illinois, five graduate students leveraging this project for their thesis work and supervised by the project PI - Shaowen Wang graduated during this project year. One student – Eric Shook – received his PhD degree in Geography and Geographic Information Science and four students graduated with Masters degree (three in Geography and Geographic Information Science, and one in Computer Science). Dr. Shook has started his tenure-track assistant professor job in the Department of Geography at the Kent State University. Dr. Guofeng Cao – a postdoctoral researcher affiliated with the project – will start his tenure-track assistant professor job in the fall of 2013 in the Department of Geosciences at the Texas Tech University. Most of the students participating in the project have had co-authorship roles of peer-reviewed papers while having received extensive training on computational and CyberGIS principles by directly joining the technical work on software engineering and integration.

Multiple course modules have been developed to teach both graduate and undergraduate students the software tools and advanced cyberinfrastructure that have been produced by the project research. For example, during this project year, more than 100 students have learned how to use CyberGIS Gateway, and got impressed by the capabilities and further potential of CyberGIS. Our project members including those from our partners play an important role in creating reliable software while getting trained to work on a diverse set of advanced cyberinfrastructure and geospatial software. These members benefit from accessing a variety of software engineering practices and cutting-edge technologies, and promise to become a core workforce for sustaining CyberGIS software ecosystems.

**Impact on physical resources that form infrastructure**

Through partnership with multiple leading cyberinfrastructure projects (e.g., the NSF XSEDE and the Open Science Grid) and industry (e.g. Esri – the leading GIS software company), the CyberGIS software environment has been widely used for several science communities (e.g.,
geography and spatial sciences, hydrology, and public health) to advance scientific problem solving. CyberGIS has also been used for several undergraduate and graduate courses on multiple campuses while becoming increasingly important resources for advanced computing and information infrastructure to empower data-intensive geospatial research and education. The project has established an innovative approach to unifying a set of comprehensive infrastructure resources of multiple modalities (cloud, high-performance computing, and high-throughput computing) for systematically advancing and evolving CyberGIS software.

**Impact on institutional resources that form infrastructure**

The project has developed an open-access model for multiple institutions and multidisciplinary communities to share advanced CyberGIS infrastructure and access a rich set of CyberGIS resources.

**Impact on information resources that form infrastructure**

The project has established a suite of advanced CyberGIS resources and services for solving scientific problems and supporting education and workforce development. These resources and services are built on advanced national cyberinfrastructure, and can be widely accessed and adapted to form advanced infrastructure for data-intensive geospatial research and education.

**Impact on technology transfer**

The software developed in the project is openly accessible, promotes open-source strategies, and adopts or adapts open community standards. These strategies have positive impact on future opportunities of technology transfer.

**Impact on society beyond science and technology**

CyberGIS Gateway – representing our broad approach to advancing CyberGIS software and related scientific problem solving – can be widely accessed by general public to gain understanding about advanced cyberinfrastructure-enabled scientific problem solving through customizable and friendly user interfaces. The CyberGIS software environment helps improve the software capabilities of the USGS National Map project, which also has significant and broad societal impacts. Another example is the contribution to “The Global Twitter Heartbeat” (also see [http://www.sgi.com/go/twitter/](http://www.sgi.com/go/twitter/)), in collaboration with SGI and Kalev Leetaru of the University of Illinois, was demonstrated at the Supercomputing Conference 2012 to create the first near-real-time combined population, tone and geographic analysis and heat map visualization, shows the significant value and impact of this NSF project beyond scientific and engineering advances.