Computers, Environment and Urban Systems 34 (2010) 264–277

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Geospatial Cyberinfrastructure: Past, present and future

Chaowei Yang [a,*], Robert Raskin [b], Michael Goodchild [c], Mark Gahegan [d]

[a] Joint Center for Intelligent Spatial Computing and Department of Geography and GeoInformation Science, George Mason University, Fairfax, VA 22030-4444, United States
[b] NASA Jet Propulsion Laboratory, 4800 Oak Grove Drive Pasadena, CA 91109, United States
[c] Department of Geography, University of California, Santa Barbara, 5707 Ellison Hall, Santa Barbara, CA 93106-4060, United States
[d] Rm 554, Human Sciences Building, 10 Symonds Street, Auckland, School of Environment, The University of Auckland, Private Bag 92019, Auckland, New Zealand

ARTICLE INFO

Keywords:
Cyberinfrastructure
Cloud computing
Virtual organizations
Geospatial science
Spatial computing
SDI

ABSTRACT

A Cyberinfrastructure (CI) is a combination of data resources, network protocols, computing platforms, and computational services that brings people, information, and computational tools together to perform science or other data-rich applications in this information-driven world. Most science domains adopt intrinsic geospatial principles (such as spatial constraints in phenomena evolution) for large amounts of geospatial data processing (such as geospatial analysis, feature relationship calculations, geospatial modeling, geovisualization, and geospatial decision support). Geospatial CI (GCI) refers to CI that utilizes geospatial principles and geospatial information to transform how research, development, and education are conducted within and across science domains (such as the environmental and Earth sciences). GCI is based on recent advancements in geographic information science, information technology, computer networks, sensor networks, Web computing, CI, and e-research/e-science. This paper reviews the research, development, education, and other efforts that have contributed to building GCI in terms of its *history, objectives, architecture, supporting technologies, functions, application communities, and future research directions*. Similar to how GIS transformed the procedures for geospatial sciences, GCI provides significant improvements to how the sciences that need geospatial information will advance. The evolution of GCI will produce platforms for geospatial science domains and communities to better conduct research and development and to better collect data, access data, analyze data, model and simulate phenomena, visualize data and information, and produce knowledge. To achieve these transformative objectives, *collaborative research and federated developments* are needed for the following reasons: (1) to address social heterogeneity to identify geospatial problems encountered by relevant sciences and applications, (2) to analyze data for information flows and processing needed to solve the identified problems, (3) to utilize Semantic Web to support building knowledge and semantics into future GCI tools, (4) to develop geospatial middleware to provide functional and intermediate services and support service evolution for stakeholders, (5) to advance citizen-based sciences to reflect the fact that cyberspace is open to the public and citizen participation will be essential, (6) to advance GCI to geospatial cloud computing to implement the transparent and opaque platforms required for addressing fundamental science questions and application problems, and (7) to develop a research and development agenda that addresses these needs with good federation and collaboration across GCI communities, such as government agencies, non-government organizations, industries, academia, and the public.

1. Introduction history, origin, and status

Following the invention of electronic computers in the 1940s, scientists began to transform data from paper-based copies to electronic forms, a trend that has transformed scientific research procedures by allowing for easy shipping and sharing of information among colleagues (Lerner, 2001). The invention of computer networks in the 1960s greatly simplified this sharing of electronic information, and the introduction of email, FTP, and other electronic communication protocols made computer networks a physical infrastructure that transformed how scientists, educators, government officials, and the public exchange ideas, conduct research, and share knowledge (Holzmann & Pehrson, 1994). Computer networks grew so fast that they became one of the defining features of cyberspace (Smith & Kollock, 1999) by providing important infrastructural support for our activities. The evolution of cyberspace has resulted in an increasing number of applications to support research, development, and decision making (Smith & Kollock, 1999) and improved the rate of sharing of

* Corresponding author.
E-mail address: chaowei.yang.1@gmail.com (C. Yang).

0198-9715/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2010.04.001

information from traditional mail-based time frames to Internet-based to real-time speeds associated with mobile devices (Murthy & Manimaran, 2001) such as Location Based Services (LBS, Kupper, 2005). A vast number of functions have been developed that have revolutionized how we conduct our daily work.

In 1998, the word Cyberinfrastructure (CI) was used in a White House press briefing by Richard Clarke, then United States (US) national coordinator for security, infrastructure protection, and counter-terrorism, and Jeffrey Hunker, then US director of the critical infrastructure assurance office, referring to the infrastructure underlying cyberspace. The US National Science Foundation (NSF) Computer & Information Science & Engineering (CISE) Directorate called for a review of this infrastructure and formed a blue ribbon review team that used the term CI formally for their landmark Atkins report (NSF., 2003) and established an Office of CI (OCI) to advance the research, development, and construction of CI.

As a generic information infrastructure, CI can support all scientific domains that collect, archive, share, analyze, visualize, and simulate data, information, and knowledge. Many science domains generate data and information with a geographic location reference. These georeferenced or geospatial data have inter-connections that follow geospatial principles/constraints, such those of geospatial analysis and geospatial modeling (de Smith, Longley, & Goodchild, 2007), and are distinguished from generic data by processing method requirements for providing LBS and place-based policies. A cross-cutting infrastructure that can support geospatial data processing within and across scientific domains is desirable. Geospatial CI (GCI) refers to infrastructure that supports the collection, management, and utilization of geospatial data, information, and knowledge for multiple science domains. The realization of the importance of such an infrastructure can be traced back conceptually to 1884 when the national program for topographic mapping was started, and formally to 1994 when the US Federal Geographic Data Committee (FGDC) was established to build a cross-agency National Spatial Data Infrastructure (NSDI). Since then, much progress has been made in defining standards by the Open Geospatial Consortium (OGC) and the International Organization for Standardization (ISO), implementing prototypes through various testbeds, popularizing industry products through seed funding, and building applications for this infrastructure (Fig. 1) through several initiatives. For example, in 2007, the Infrastructure for Spatial Information in the European Community (INSPIRE) directive entered into force and laid down a general framework for a Spatial Data Infrastructure (SDI) to support European Community environmental policies and activities.

These initiatives support each other with their own unique emphases: for example, the NSDI focuses on spatial data collection, sharing, and service, and its geodata.gov provides geospatial data services. Data.gov provides all publicly available US government data, with their geospatial aspects supplemented by geodata.gov. Digital Earth is a vision popularized by US Vice President Al Gore in 1998 for advancing technology to store, integrate, and utilize georeferenced data to build a virtual world for multiple applications. Grid computing is focused on distributed computers to optimize distributed computing. Cloud computing is focused on data, platforms, infrastructure, and software as services for end-users. The Global Earth Observation System of Systems (GEOSS) is an initiative to build a system of systems for global Earth observations focused initially on nine societal benefit areas.

Over time, the amount and availability of geographic information has grown exponentially, and a new dedicated GCI is needed to process and integrate geospatial information to, for example: (a) provide LBS for stakeholders, such as place-based policy makers, (b) supply geospatial analysis and modeling as services, and (c) support scientific and application problem solving across geographic regions. The Association of American Geographers (AAG) began discussing CI in 2005, and ultimately, a dedicated Specialty Group was formed to advance the geospatial aspect of CI. Relevant GCI meetings were held by the University Consortium for Geographic Information Science (UCGIS) in 2007 and at the Geographic Information Science (GIScience) Conference in 2008. This special issue of GCI by CEUS is one such effort to capture the recent and growing activities, and this paper is a review of these developments. To conduct this review, we evaluated all SCI (Science Citation Index) and EI (Engineering Index) papers relevant to CI and geospatial information. Since most of the authors are US researchers, this review emphasizes American developments, but we added other geographic regions based on our experiences.

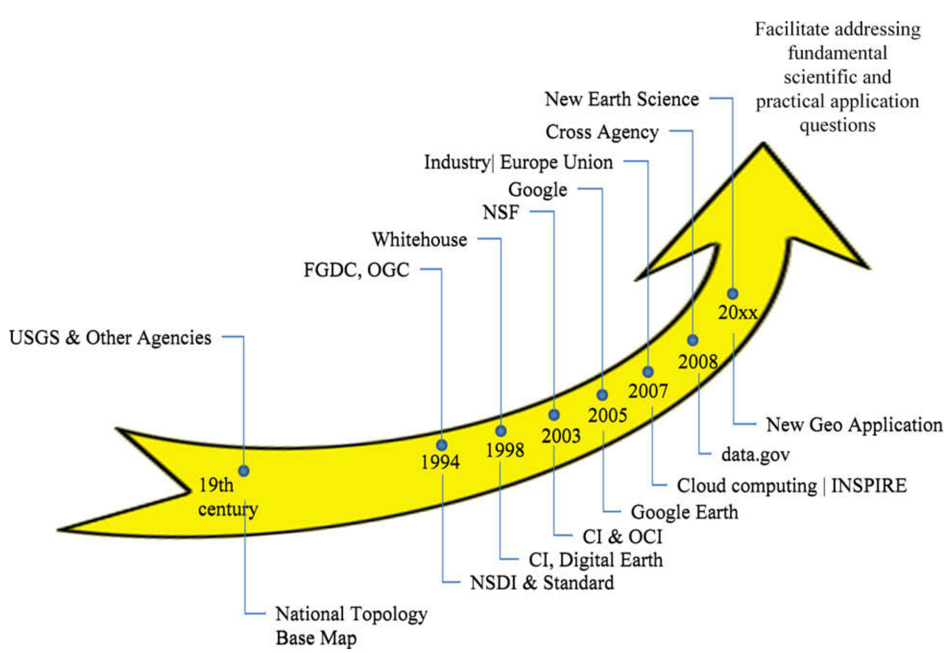

Fig. 1. History of geospatial Cyberinfrastructure.

GCI utilizes an integrated architecture that builds upon past investments to share spatial data, information, and knowledge. These systems facilitate or advance the development, research, education, popularization, and advancement of geospatial sciences and often enable functions that were not previously feasible, for example, mashups that leverage petabytes of global geospatial information served through Google Earth and Bing Maps. NSF has been a major driver in the development of CI. NSF has made significant strategic investments in CI development (both before and after the establishment of the OCI) in targeted domains, such as ecology, hydrology, and social sciences. Each of these domains requires geospatial information in their research and developments that enable research sharing and the collection of geospatial information to advance their objectives. GCI provides much needed infrastructural support for these efforts.

A predecessor to the OCI program was the NSF National Middleware Initiative (NMI), which generated many middleware products that connect software components and applications. Other relevant early NSF initiatives were the large Information Technology Research (ITR) projects and the Grid initiative. NSF. (2007) later produced a landmark CI vision document that emphasizes four strategic areas: (a) data and visualization, (b) High-Performance Computing, (c) virtual organizations, and (d) education and workforce training. All four components are fundamental to the advancement of geospatial science research. For example, the NSDI focuses on data sharing and utilization. Geospatial visualization is a longstanding research endeavor to better interpret geospatial data and information, and high-performance geospatial computing provides a driving technology to simulate and predict complex geospatial science phenomena (Yang, Li, Xie, & Zhou, 2008). Many organizations have become increasingly important for driving the advancement of geospatial science in global and climate change applications. Education and workforce training has always been at the center of geospatial sciences. Current practices in geospatial science provide crosscutting geospatial analysis and modeling support to many scientific domains including ecology, geosciences, Earth science, air quality, water studies, and atmospheric science.

Humanity in the 21st century faces great challenges to better understand why and how the Earth is changing (http://nasascience.nasa.gov) and to make better decisions at global to personal levels (Lannotta, 2007). Representative challenges include strategies to reduce energy consumption and stabilize atmospheric emissions so that the global temperature will not increase several degrees in the next century; choose a housing site that minimizes the risks of forest fire, flooding, and other natural/man-made hazards; and more generally, improve the quality of life. These practical questions require us to utilize most available geospatial data, information, and knowledge to produce scientifically sound decision supporting information and require capabilities to achieve the following (Yang & Raskin, 2009): (1) integrate real-time and historical data resources, (2) leverage both traditional and fully interoperable and open resources, (3) interpret data and information that cross-domains, regions, and cultures, and (4) capture and utilize knowledge autonomously so that the most appropriate inputs can be utilized and the best decision supporting information can be generated. To answer complex questions, we must effectively utilize facilities, instruments, and other resources to pursue fundamental and transformational questions, unravel newly revealed mysteries, and expand our understanding of the Earth and the universe (NSF., 2009). As a long-term objective, GCI will facilitate answering these daunting questions and build the capacity to leverage existing geospatial knowledge and resources, transform how we conduct research, pursue scientific and application questions, and collaborate across geographic regions, cultural differences, and domain turfs.

For a fast evolving, broad-based field such as GCI for which the geospatial dimension cuts across domains and disciplines, a review can not focus on any one aspect that might provide researchers with a clear step-by-step research agenda. Instead, we provide an overview of GCI from the perspectives of its history, current status, and future developments with a focus on aspects that are common to all relevant domains. Our objectives are to introduce GCI to the following groups: (1) scientists who could benefit from its end use capabilities, (2) information scientists who could potentially advance the leading-edge research frontiers, and (3) geospatial information scientists who desire a systematic view about GCI and its future directions. We present views of GCI from five perspectives: logical frameworks, enabling technologies, geospatial functions, domain applications, and desired future research. These five perspectives are presented in the next five sections, respectively.

2. Framework

2.1. GCI resources and logical framework

GCI includes multiple categories of resources within a flexible, scalable, and expandable framework cube (Fig. 2). This prototypical GCI cube consists of the following three dimensions:

(1) *Functions* include both generic CI functions (computing, networking, and hardware) and those that are geospatial-specific. The GCI functions include the following: (a) a middleware layer to bridge geospatial functions and resource management, monitoring, scheduling, and other system-level functions, (b) a geospatial information integration layer to integrate geospatial data, information, and knowledge flow as supported by observations, geospatial processing, and knowledge mining, and (c) geospatial functions to provide various analytical functions for end-users.
(2) The *community* represents the virtual organizations and end-user interactions within specific communities including geographic, environmental, Earth, and other science domains. This dimension also provides feedback channels for knowledge collection functions to leverage scientific community and citizen participation.
(3) *Enabling technologies* provide technological support to invent, mature, and maintain all GCI functions, such as collecting data through observations and collecting and utilizing knowledge through a semantic web.

The geospatial information integration and the geospatial functions layers distinguish GCI from other generic CIs. GCI should leverage successes derived from enabling technologies (detailed in Section 3), functional components (detailed in Section 4), and comprehensive solutions to community applications (detailed in Section 5). The Computing and Networking functions are not detailed because they are IT generic.

3. Enabling technologies

The architecture and integration of GCI benefit from numerous enabling technologies, many of which contributed to the birth of GCI.

3.1. Earth observation and sensor networks

Earth observation and sensor networks provide data collection capabilities to feed petabytes of data into a GCI on a daily basis (Freudinger, 2009). Sensor networks also utilize GCI to support the evolution from passive logging systems to an intelligent sensor networks that actively send data to servers, a capability that will

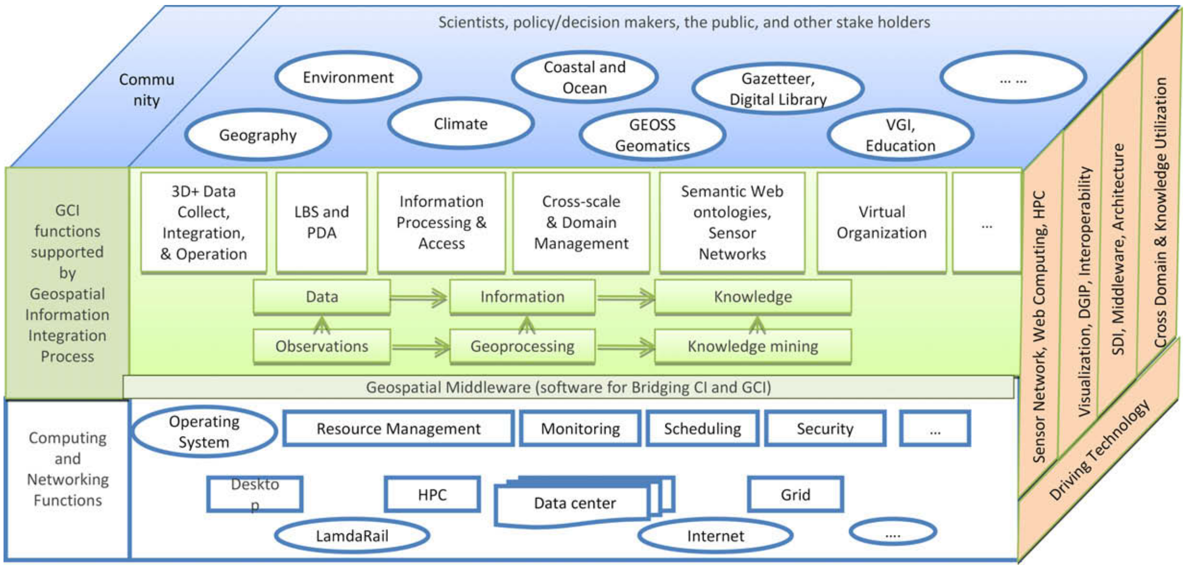

Fig. 2. GCI framework cube.

become a critical asset for Earth and environmental science studies (Hart & Martinez, 2006). Because of the direct connection of real-time sensor network to GCI, real-time decisions can be made to be effective in applications such as emergency response (Akyildiz, Su, Sankarasubramaniam, & Cayirci, 2002). This capability will provide an essential driving force for GCI in the coming decades because citizens increasingly depend on real-time information to make personal, business, and management decisions. The development of smart, self-adjustable sensor networks will further contribute to real-time decision making.

3.2. SDI

The SDI made vast amounts of geographic information that were collected by government agencies or private companies publicly available and built them into open services to be integrated into customized scientific and practical applications (Nebert, 2004). SDI developments will contribute significant lessons learned and experiences to future GCI developments, such as those for quality control and resource synchronization (Yang, Cao, & Evans, 2007).

3.3. Distributed geographic information processing (DGIP)

Distributed geographic information processing (DGIP) handles geospatial information for GCI using dispersed computing resources across platforms, such as Web computing and computer clusters (Yang & Raskin, 2009). Many geospatial processing functions need to be rethought or rewritten to fit into GCI. DGIP research will provide a guiding methodology and principles for implementing geospatial middleware that can support geospatial processing in GCI.

3.4. Web computing

Web computing, especially the recent advancements to Web 2.0 and Web 3.0, provides an important platform for GCI applications such as online data searching, mapping, and utilization and supports uniform interfaces, such as Google Maps, for the exploration of scientific data (Pierce et al., 2009). Web computing also provides innovative technical approaches (protocols, message formats, and programming tools) and groundbreaking services (such as wikis; Stein, 2008) as critical support (Fox & Pierce, 2009). Further advances in web computing toward an intelligent Semantic Web (de Longueville, 2010) will provide additional practical support to GCI.

3.5. Open and interoperable access technologies

Open and interoperable access technologies, such as XML/GML, Javascript, and AJAX, enable geospatial data to be published, accessed easily, and adapted to customized applications through mashups. The contributions from interoperability research and development over the past decade have rapidly popularized the sharing of geospatial information. Spatial Web portals (Yang, Li, Xiao, Raskin, & Bambacus, 2007) and gateways (Wilkins-Diehr, Gannon, Klimeck, Oster, & Pamidighantam, 2008) have enabled access to supercomputing and information systems that manage geophysical datasets across Grid platforms (Pierce et al., 2008). Standards organizations (e.g., the OGC) increasingly collaborate with domain disciplines (e.g., Earth science) to build cross-cutting interoperable specifications and prototypes. This recent trend will help advance sharing and interoperability within and across disciplines.

3.6. High-Performance Computing (HPC)

High-Performance Computing (HPC), including Grid computing, cluster computing, and ubiquitous computing, provides computing power for GCI users to conduct data and computing-intensive research that cannot be conducted on single computers (Xie, Yang, Zhou, & Huang, 2010; Yang, Wong, Miao, & Yang, 2010). Further research is needed on how to leverage HPC for geospatial information with new methods, approaches, and middleware.

3.7. Open-source software and middleware

Open-source software and middleware provide a glue function to integrate the components of data, processing, applications, and infrastructure. For example, the NSF NMI produced many products that are widely used in CI for supporting these functions (http://www.nsf.gov/news/special_reports/cyber/middleware.jsp). The Unidata Internet Data Distribution system provides real-time data distribution of atmospheric and meteorological datasets. Workflow

(Papadopoulos & Linke, 2009) and distributed visualization (Smarr, Brown, & de Laat, 2009) middleware are often needed for complex scientific applications. For the development and maturation of middleware, the Reuse Readiness Level (RRL, Marshall & Downs, 2008) is an important indicator of middleware readiness. Experiences with middleware technology provide important hints for how to adapt geospatial software to a GCI, but more research is required on methods to distribute, synchronize, integrate, and balance the processing within a distributed environment.

3.8. Data visualization needs in applied science

Solutions to data visualization needs in applied science made it easy to engage end-users to utilize and contribute to a GCI. Geospatial information has a geographic connection that spans multiple domains; the FGDC framework datasets can be leveraged as a framework to facilitate the sharing of data, information, and knowledge within and across domains. How to best present data and information from multiple resources and within applied science domains is a critical challenge.

3.9. Cross-domain sharing and collaborations

Cross-domain sharing and collaborations are essential to the ability of a GCI to support and leverage expertise across user communities. The past decades' efforts in multidisciplinary research and development have helped us realize the importance of cross-domain collaboration and provided lessons and experiences that can be utilized for our subsequent benefit (MacEachren & Brewer, 2004). A multi-domain perspective requires a GCI to be sufficiently expandable and flexible to support the easy plug-and-play of new functions through a service-oriented architecture (SOA) with standards-based interoperable interfaces and open-source access. Further research using social sciences to help cross domains, cultures, and jurisdictional boundaries are important for these collaborations.

3.10. Knowledge capture and utilization

Knowledge capture and utilization provide for the smart discovery, integration, indexing, collection, and utilization of vast amounts of data, processing components, and other tools available in a GCI. Research on the Semantic Web (Brodaric, Fox, & McGuinness, 2009) includes research on semantic understanding (such as Embley, 2004), knowledge-based functions (Reynolds & Zhu, 2001), ontology-based dynamic query rewriting for the semantic translation of data, and other learning processes (Shimbo & Ishidab, 2003). Semantic interoperability is an important prerequisite to information integration across domains, where vocabulary differences may be common. The geospatial Semantic Web is a vision in which locations and LBS are fully understood by machines (Egenhofer, 2002). Google Earth and other visual globe technologies have provided the initial strides in this direction, which were based upon a common understanding of geospatial dimensions. An analogous understanding of features and attributes will help advance the future capabilities of GCI. Geospatial knowledge refers to a shared understanding of geospatial terms and the collective, expert intelligence it makes available (Raskin & Pan, 2005). As knowledge is inherently dynamic and expanding, a GCI must consider a constantly changing field of knowledge. The geospatial Semantic Web is a Web in which the browser, crawler, and other tools understand spatial content and can exploit this knowledge on-the-fly (Maué & Schade, 2009; Roman & Klien, 2007).

3.11. System integration architectures

Four major architectures have been widely researched and adopted for GCI system integration (Yang & Raskin, 2009).

- Multi-tier/layer organizations support and implement CI objectives through layers of functions with community-specific knowledge environments for research and education. The NSF (2003) Atkins report laid out a four-layer architecture emphasizing the following: (a) core technologies, such as HPC, incorporated into CI services, (b) CI supporting applications, (c) applications of IT to science and engineering research, and (d) science and engineering research activities. The NSF (2007) CI report further emphasizes HPC, visualization and data/information/knowledge, virtual organizations, and education and workforce training. Zhang and Tsou (2009) and Wang and Liu (2009) described layered architectures for integrating geospatial components to support GCI. These layered views of GCI are in alignment with those of the Federal Enterprise Architecture (FEA) and provide good guidance for the logic design of the components necessary to implement a GCI. Challenges remain in this architecture; for example, how to utilize a knowledge-oriented layered framework for integrating/interoperating multiple infrastructures (Jabeur, McCarthy, Xing, & Graniero, 2009).
- Mashup and plug-and-play leverage the achievements of geospatial interoperability over the past few decades. Interoperability lays out a foundation for easily integrating heterogeneous components in a plug-and-play fashion or for mashup with minor scripting of the interfaces (Bambacus et al., 2007). This plug-and-play integration has become an ideal illustration of the achievements of interoperability and provides new system integration methods that are envisioned to be critical to effective system integration in the coming decades (Nebert, 2004).
- SOA is based on the assumption that all components can be built as services and the SOA can facilitate service registration, discovery, and binding to form new functional and/or scientific applications. For example, Hey and Trefethen (2005) described how the United Kingdom's e-science program utilized SOA to compose an e-Infrastructure or CI to support multiple science domains. Bogden et al. (2006) described how an SOA was used to design the Southeastern Universities Research Association (SURA) Coastal Ocean Observing and Prediction (SCOOP) program to support coastal real-time decision making. Zhang, Tsou, Qiao, and Xu (2006) discussed the need of GCI for the development of future geospatial information services based on Grid computing, Web services, and OGC standards.
- Workflow chaining utilizes business logic in integrating applications to construct a GCI application and was popularized with OGC service proliferation within geospatial domains using BPEL, eBRIM, Kepler, and other generic- or geospatial-specific chaining engines and description languages. Minsker et al. (2006) described a workflow chaining service that integrates heterogeneous tools to support environmental engineering and hydrological science communities to (a) share data and research through interactive provenance and (b) discover, share, analyze, and evaluate research tools. Pennington et al. (2008) used workflow chaining to chain services that educate scientists to better leverage resources in their scientific research. Workflow chaining also provides an interesting architectural method for scientific experimentation (von Laszewski, Younge, Xi, Mahinthakumar, & Lizhe, 2009). An important task in the workflow chaining process is to compare the capabilities of the various available tools to compose scientifically sound applications (Deelman, Gannon, Shields, & Taylor, 2009).

These four architectures are practical and popularly used to develop GCI systems or applications. They may be mixed in a real development scheme, but each emphasizes separate aspects of the system integration: (a) multi-tier organization focuses on business logic and is suitable for requirement analysis, (b) mashup and plug-and-play emphasize practical development within the current stage by leveraging public available resources, (c) SOA emphasizes component interoperability if services are strictly conformed with standards, and (d) work flow chaining emphasize semiautomatic or automatic integration of applications. Further architectural improvements are needed and can be contributed from future GCI research pursuits (detailed in Section 6), such as how to (a) maintain and discover service performance patterns (Li, Yang, & Yang, 2010), (b) optimize a framework, and (c) build in knowledge for automating work flow chaining (Li, Yang, & Raskin, 2008).

4. GCI functions

The basic functional components within a GCI provide users access to geospatial data, information, knowledge, and processing tools.

4.1. Multi-dimensional data processing

The multidimensional nature of geospatial data and information forms an essential theoretical foundation for GCI (Li, Kim, Govindan, & Hong, 2003). For example, the inclusion of the time dimension leads to geospatial dynamics (Hornsby & Yuan, 2008). The vertical dimension is crucial in understanding the atmosphere and ocean. Other higher dimensions represent focused scientific parameters (e.g., a spectral band) that integrate domain knowledge (Yang, 2000). Research on multi-dimensional data has been ongoing for decades, but new technologies can help us to process data with better resolution in three dimensional space, time, and spectral bands and to extract information hidden in data. Therefore, efficient multi-dimensional data processing represents a challenging, driving GCI function.

4.2. Data collection and heterogeneous integration

Data collection and heterogeneous integration processes can introduce petabytes of data into a GCI. For example, a tidal creek watershed data management framework (White et al., 2009) integrates terabytes of environmental, demographic, and socioeconomic data using OGC/FGDC data sharing standards. Sensor webs, Earth observations, ground surveying, and questionnaires have produced a significant amount of geospatial information in various forms, locations, and systems and served multiple purposes (Nittel et al., 2004). Data integration is essential to the ability of a GCI to publish data (Horsburgh et al., 2009) and to feed data from sensors to users (Ledford, 2009). How to better organize the vast amounts of geospatial data across domains, resources, applications, and cultural backgrounds represents a grand challenge of GCI.

4.3. Data preservation and accessibility

Historical data are important to the understanding of dynamic processes (such as global change) and should be preserved and made accessible to users for perpetuity (Berman, 2008). For example, scientific breakthroughs or patents may be illuminated only in the context of subsequent related scientific resources (Clarkson, 2008; Jensen, Chen, & Murray, 2008). In the future, the processes associated with data, information, and knowledge transformations should be preserved as data provenance (Simmhan, Plale, & Gannon, 2008), especially for global and climate studies and long-term policy making (Morisette et al., 2009). These capabilities require a GCI to provide sustainable and long-term data archival capability and easy discovery, access, and utilization of historical datasets.

4.4. Supporting the life cycle from data to knowledge

A GCI should support the entire data life cycle, including the acquisition, verification, documentation for subsequent interpretation, integration from multiple sources, analysis, and decision support (Borgman, Wallis, Mayernik, & Pepe, 2007). For example, a GCI was utilized to integrate informatics and intelligent systems to support decision making in chemical engineering for the entire product life cycle – from individual units to enterprise-level geographically dispersed supply chains (Venkatasubramanian, 2009a, 2009b). In a collaborative environment, GCI is widely needed to manage data and serve as a tool for managers and practitioners to access, analyze, and determine data management needs (Carter & Green, 2009).

- *Metadata* are important for helping preserve understanding and building new research environments for data sharing and repurposing (Henry & Friedlander, 2008). Metadata also form the foundation for catalogs, Web portals, and other discovery services (de Longueville, 2010; Yang et al., 2007). *Unique object identifiers* provide permanent names for archived data and for their inclusion in journal article references (Digital National Framework, 2007). *Information extraction* is made possible through data mining, knowledge reasoning, and other artificial intelligence processes (e.g., Li, Yang, & Sun, 2009). Targeted parameters of importance to a science domain can be extracted to feed a model simulation or decision support system (Datcu et al., 2003), and *geospatial data mining* typically requires geospatial analysis and modeling principles in the mining process.
- *Results representation and visualization* are especially critical when using semantic technology to interpret datasets and to develop attractive end-user interfaces (Gahegan, Luo, Weaver, Pike, & Banchuen, 2009).
- *Services* should be utilized and chained in response to rapid application development, such as for emergency response (Friis-Christensen, Lucchi, Lutz, & Ostlinder, 2009).
- *Uniform access to data and information services*: Web portals (de Longueville, 2010; Maguire & Longley, 2005; Yang, Evans, et al., 2007) such as the Geospatial One Stop (GOS), provide a single entry point to data, information, and knowledge. For example, Papadopoulos and Linke (2009) utilized Web portals to enable their seamless access and utilization of geospatial resources. Web portals are also used for integrating distributed generalization and geoprocessing (Wolf & Howe, 2009).

Many achievements have been made in integrating data, information, and knowledge management and access, but supporting the automatic integration of a data life cycle requires a greater understanding of the geospatial data life cycle.

4.5. Virtual Organizations (VO)

Virtual Organizations (VO) bring together distant users to form collaborations across regions, domains, cultures, and scales (Myhill, Cogburn, Samant, Addom, & Blanck, 2008 and Baker et al., 2009). A VO is a dynamic set of individuals and/or institutions defined by a set of resource-sharing rules and conditions. The individuals/institutions share some commonality in their concerns and requirements. While they may vary in scale, scope, duration, sociology, and structure and be dispersed geographically and institutionally, they can still function as a coherent unit (Cummings, Finholt, Foster, Kesselman, & Lawrence, 2008). For example, the TeraGrid provides a comprehensive GCI for bridging domains through portals utilizing one of the world's largest computing alliances across eight campuses (Catlett, 2005). VOs have become increasingly important because of the frequent exchanges needed among geographically dispersed scientists (Gemmill, Robinson, & Scavo, 2009). The GCI community should pay attention to diversity requirements and address how to bridge geographically dispersed virtual communities that connect participants with diverse backgrounds (Myhill et al., 2008).

4.6. Semantic Web and knowledge sharing

Semantic Web and knowledge sharing is an essential ingredient to cross-domain collaborations, interdisciplinary discoveries (Berners-Lee, Hendler, & Lassila, 2001; Brodaric et al., 2009), and the life cycle from data to knowledge. For example, meaning-based data integration forms a basis for Web 3.0 heterogeneous data sharing (Lassila & Hendler, 2007), and the semantic, knowledge, or cross-cultural sharing of resources forms a basis for cross-domain studies (Lightfoot, Bachrach, Abrams, Kielman, & Weiss, 2009). A GCI should provide a common semantic framework to enable long-term semantic interoperability and shared scientific understanding.

4.7. HPC and associated spatial computing

HPC and associated spatial computing are essential for enabling computing-intensive and data-intensive geospatial research and applications, for example, for rapid response (Chiang, Dove, Ballard, Bostater, & Frame, 2006) and dust storm forecasting (Xie et al., 2010). Further research is needed to best leverage computing platforms for geospatial domains according to newly discovered geospatial best practices and principles.

4.8. Location based service

LBS is becoming increasingly important as witnessed by (a) the popularization of mobile devices, such as PDAs and iPhones, (b) the research and development that made LBS efficient and convenient, (c) the establishment of the Journal of LBS, and (d) the popularity of Google Earth and related virtual globe software. This need is likely to further increase as more geospatial data becomes available.

4.9. Cross-scale and domain management

Cross-scale and domain management have emerged as essential for GCI (Lightfoot et al., 2009). For example, GCI is utilized to support a broad set of functions for Chesapeake Bay studies including (Ball et al., 2008): (a) single interface access to heterogeneous datasets, (b) new tools to support multi-domain scientists, and (c) integration with other networks. Multidisciplinary research drives the integrated understanding of geospatial sciences and provides more accurate scientific studies and greater utilization than any single domain (Baker et al., 2009). However, the full implementation of cross-domain sharing and collaboration will require significant improvements in GCI for all of the functions mentioned.

5. Application domains and user communities

Geospatial principles intrinsically reside in almost all scientific domains (Yang et al., 2010). GCI functions provide domain users with real experiences and new requirements as witnessed within multiple domains.

5.1. Geographic sciences

GCI is widely used to support the sharing and utilization of geospatial data. For example, the successes of GOS, the ESRI geography network, Google Earth, and MS Virtual Earth (now Bing Maps) provide popular portals to support geographic research and international applications (Yang & Raskin, 2009). However, more research is critically needed by GCI researchers and geographers to mine the mechanisms of complex phenomena for building new GCI functions to solve complex problems.

5.2. Environmental sciences

GCI has been used extensively for environmental studies (for an example, see Minsker et al., 2006) to (a) share data and research results, (b) provide tools to discover, share, analyze, and discuss results, and (c) utilize workflow chaining services for integrating heterogeneous tools.

- GCI has been widely adopted in Water Management and Water Quality to (a) preserve and archive records, integrate data grids and GIS to provide universal access to records to support environmental policy decisions and planning for water quality and watershed management (Pezzoli, Marciano, & Robertus, 2006 and Rich, Weintraub, Ewers, Riggs, & Wilson, 2005), (b) connect heterogeneous databases and analytical functions to support a national water management system (Goodall, Jeffery, Whiteaker, Maidment, & Zaslavsky, 2008), (c) monitor the Hawaiin water and hydrological cycles by integrating wireless sensors, grid computing, and 3D geospatial data visualization tools connected through a secured single portal entry (Kido et al., 2008), and (d) support water distribution system monitoring, analysis, and control through real-time observations (Mahinthakumar et al., 2006).

However, further advancements in GCI are needed to (a) support multidisciplinary scientific research, such as the Chesapeake Bay Environmental Observatory (CBEO) (Ball et al., 2008) that integrates hundreds of parameters for short term and long-term environmental decisions, (b) support real-time decision making with real-time output, using high performance and adaptive computing (Ramaprivan, 2008), and (c) provide novel processes for understanding the composition and behavior of natural waters, the distribution of water, and more effectively coupling water supplies to societal needs (Bartrand, Weir, & Haas, 2007).

- Kahhat et al. (2008) evaluated a GCI built for waste management to track and manage the life cycle of e-waste, support future e-waste regulation and management, and analyze the impact of e-waste on US and international communities. A fully integrated waste management system is still not available to track all of the important components, such as the components of a computer. Research is needed on both the technological side for better product tracking systems and with respect to privacy concerns over what we should and should not track.
- A GCI was used in ecology and public health to (a) place information on four decades of annual mosquito population dynamics online for biomedical and public health studies (Sucaet, Van Hemert, Tucker, & Bartholomay, 2008), (b) connect terabytes to petabytes of data from relevant cancer surveillance systems and databases into a seamless thread to support public health researchers, policy makers, and administrators who were geographically dispersed (Contractor & Hesse, 2006), (c) facilitate the monitoring and analysis of the dynamics of a mosquito population (Sucaet et al., 2008), and (d) track and monitor invasive

species (Graham et al., 2008). Problems still remain though, for example, how to optimally manage terabytes of historical and real-time data to ensure maximum usability.
- Keating (2009) laid out a detailed inventory of GCI components related to Air Quality (AQ) that can be utilized to build a multi-disciplinary AQ GCI through (a) wikis, (b) network development and maintenance, (c) analytical and visualization development, and (d) outreach.

5.3. Climate sciences

GCI was utilized to integrate, archive, and distribute phenological research to support climate change analysis (Morisette et al., 2009). Within the global environment, GCI must play a much larger role in global and climate change studies by helping to integrate resources and bridge geographic regions worldwide. For example, methods are required to coordinate simulations of parameters at multiple scales to enable customized climate predictions and the retrieval of accurate historical climate records.

5.4. Coastal and ocean studies

A GCI with developed middleware is being used for model integration/analysis, data mining, cross-application integration, and to provide real-time forecasting to support coastal research and decision making (Agrawal, Ferhatosmanoglu, Niu, Bedford, & Li, 2006). A collaborative CI was used for event-driven coastal modeling to support real-time forecasting systems for the southeast US (Bogden et al., 2006). The North-East Pacific Time-series Undersea Networked Experiment (NEPTUNE) in Canada is a sensor web and virtual observatory that was created to monitor and disseminate sea floor conditions for real-time coastal applications (Clark, 2001). GCI is also used to integrate satellites, shore-based radios, and a growing fleet of smart undersea gliders to provide near real-time data for assimilative models (Cao, Yang, & Wong, 2009; Schofield, Glenn, Chant, Kohut, & McDonnell, 2007). Further research that conducts simulations or real-time forecasting for applications such as tsunami emergency prediction, response and mitigation is still needed.

5.5. Broader Earth-system sciences

GCI provides a platform for sharing data and computing resources that can bridge data producers and data consumers through portals and information exchange (Yang, Li, et al., 2007). This infrastructure assists in the collection of data, mediating cross-domain collaborations, and utilizing multiple types of knowledge to support (Kessler, Mathers, & Sobisch, 2009), for example, long-term ecology studies (Baker & Bowker, 2007), polar research (Lubin et al., 2009), the Geosciences Network (GEON, www.geogrid.org), Sharing Environmental Education Knowledge (SEEK, www.seek.state.mn.us), Earth science infrastructure in Australia (www.auscope.org.au/), and the solid Earth and environment GRID (ScenzGrid, http://www.seegrid.csiro.au/twiki/bin/view/SCENZGrid/WebHome). Within the Earth sciences, scientists face problems of how to solve the following issues: (a) integrate petabytes of data efficiently, (b) provide smart access to petabytes of data, (c) support higher resolution simulations and forecasting, and (d) visualize 3D and 4D data from both observations and model simulations.

5.6. GEOSS

OptIPlanet (Smarr et al., 2009) is utilizing GCI to solve complex global problems to control the establishment and maintenance of connections in a network to support distributed visualization. GCI can be essential for integrating global resources to support the applications of GEOSS and GEO is designing a GEOSS Common Infrastructure and implementing it by deploying a component and service registry, a clearinghouse, and portals. Problems of data and metadata provenance and distributed search performance (Yang, Cao, et al., 2007) are waiting to be addressed.

5.7. Geomatics

GCI transforms Geomatics by providing reliable communications and HPC to enable scientists, researchers, students, and practitioners to share data, tools, procedures, and expertise in geospatial information (Blais & Esche, 2008). The further introduction of mobile devices will broaden access to GCI to anywhere and anytime through LBS.

5.8. Digital libraries

GCI has been utilized to evolve digital libraries into educational resources for a cyber-teaching environment, a cyber-workbench for researchers, and an integration environment for educational research and practice (McArthur & Zia, 2008). Delserone (2009) reported utilizing GCI technologies to implement a digital library for the initiation of the e-Science and Data Services Collaborative. The further development of gazetteers and geospatial information retrieval (GIR) techniques will be essential to this application.

5.9. Education

Education has been a focus of GCI developments by providing convenient methods for educators, students, and researchers in Earth and geographic sciences to disseminate, collect, collaborate, review, and comment on data and information (Buhr, Barker, & Reeves, 2005). Various GCI technologies, such as mobile computing and modern GCI (Bugallo, Marx, Bynum, Takai, & Hover, 2009), have been leveraged to support classroom education. For example, Eschenbach et al. (2006) used GCI to support environmental education for K-12 students using CLEANER. Chang, Lim, Hedberg, The, and Theng (2005) described a portal that allows students to explore the various scenarios for sea level rise and beach erosion by integrating data through a GCI. Virtual globe technologies such as Google Earth and Microsoft Bing Maps are widely used in classrooms, but more effective methods are required to integrate heterogeneous geospatial information for the classroom and present them in an intuitive fashion that facilitates learning. Better methods for visualization, multi-dimensional data integration, and knowledge management and utilization are still required (MacEachren & Kraak, 2001).

Numerous GCI achievements have been made in transforming data-rich scientific studies over the past decade, as reviewed in previous sections. These successes range from a sub-domain project partnership for data sharing and information exchange (Yang, Li, et al., 2007) to cross-domain (Katz et al., 2009 and Winer, 2006) and cross-continent (Peters et al., 2008) collaborations. For example, TeraGrid provides a comprehensive GCI for bridging domains through portals utilizing one of the world's largest computing alliances across eight campuses (Catlett, 2005). GCI tools are used in multiple domain designs to integrate manufacturing constraints and qualitative knowledge to maximize profits and minimize the environmental impact of industrial production across geographic regions (Winer, 2006). However, most of these successes are built on simple system integrations, and many research questions remain to be addressed before we can achieve the GCI objective to facilitate answering the most daunting questions and transform how we conduct research.

6. Discussion and future strategies

To achieve the transformative function of GCI, we need to complete the following tasks: (1) research and develop a GCI-centered research platform that enables discovery in multidisciplinary science (Elmagarmid, Samuel, & Ouzzani, 2008) and knowledge sharing, and fosters more meaningful analyses of data and the visualization, modeling, and simulation of real-world phenomena, (2) change the focus from technology to humans for GCIs built to enable science and engineering discovery (Tsai et al., 2008) so that scientists can focus on ideas and innovation and engineers can focus on engineering development (Whey-Fone et al., 2008), and (3) develop new GCI technologies and tools to help answer fundamental science questions and enable complex applications that benefit human beings. However, to implement a GCI that transforms how scientists and the public pursue fundamental scientific questions and complex application problems, we must leverage the current successes and potential problem solving capabilities of GCIs with a focus on sharing and collaboration, such as geocollaboration (MacEachren & Brewer, 2004). The future of GCI poses grand challenges in many aspects including the integration of currently isolated CIs from multiple domains, the evolution of GCI from a technology-centered to a human-centered paradigm, the advancement of GCI to support multiple science domains by simulating complex phenomena in a virtual fashion, and the acceptance of GCI by a broad range of stakeholders who use geospatial data. With the advancement of technologies, more study in social sciences is urgently needed to enable efficient collaboration among team members, communities, and domains. We believe at least seven aspects need further research and development:

6.1. Studying social heterogeneity to identify geospatial problems encountered by relevant sciences and applications in the context of GCI

GCI is targeted to provide solutions to complex scientific and application problems identified through initiatives such as Digital Earth (Goodchild, 2008a) and GEOSS (Lannotta, 2007) by enabling scientists and decision makers to focus on ideas and innovations and relieving them from considering technical details. This is a very complex task that starts with clearly identifying problems that cannot be solved without a GCI; its challenges are both technological and social. The complexity of GCI itself originates with its contributions from broad technological areas, such as Grid, infrastructure, workflow, scheduling, resource discovery and allocation, security and algorithms (Chiu & Fox, 2009), business models, domain knowledge frameworks, study methods, Web 2.0, and multi-core computing (Fox & Pierce, 2009; Loudon, 2009). The identification of appropriate problems will require scientists with different backgrounds to work together. Socially, the blending of scientists across domains and geographically dispersed teams is a grand challenge as has been observed by various GCI projects, such as LEAD (Lawrence, Finholt, & Kim, 2007). Therefore, (a) the attitudes and needs of data curation in collection, representation, and dissemination also pose challenges to both knowledge sharing and social complexities for the success of a GCI (Winget, 2008), (b) how to design a GCI for diverse end-users is a challenge to cross-cultural participation (Fischer, 2009), (c) the collaboration among scientists for the next generation of GCI poses challenges due to their heterogeneity in culture, geography, and scientific domain (Olson, 2009), (d) the openness among people and domains will take time to be implemented from public debate to public policy (Wunsch-Vincent, Reynolds, & Wyckoff, 2008), and (e) how users view GCI and their willingness to contribute, take, and maintain a GCI will involve social dynamics (Ribes & Finholt, 2009).

Meyer and Schroeder (2009) argued that we need to establish a new sociology for e-Research based on GCI by developing a sociology of knowledge that is based on an understanding of how science has been historically transformed and shifted into online media. The experiences of the Open Science Grid (OSG) (Avery, 2008) broadly illustrate the breadth and scale of the efforts that a diverse, evolving collaboration must undertake to build and sustain a large-scale, multidisciplinary GCI.

6.2. Analyzing the information flow from data to information and the processing needed for solving problems

- Progress in multidimensional sciences: geospatial phenomena typically display characteristics that depend upon the selected time scale, spatial scale, and scientific domain. Therefore, a better understanding of the dynamic dimensions of scientific phenomena will help us design better adaptive computing to understand the principles driving scientific problems (Hornsby and Yuan, 2008; Maguire, Goodchild, & Rhind, 1991) and provide better technological solutions for scientific problem solving. These advancements will also provide theoretical support for developing new sensor networks for more targeted multi-dimensional data collections.
- To support place-based policy and other national and international initiatives, various SDIs should be integrated into the Global SDI (Nebert, 2004) based on a seamless integration of data. Data provenance is essential in such a process for the following tasks: (a) facilitate the sharing of geospatial datasets, (b) maintain the information processed from raw data, and (c) validate the knowledge obtained from the information. Interoperability is critical to the integration process and should be maintained at the data, information, and knowledge levels to avoid building non-sharable stove-piped data systems (Cruz, Sunna, Makar, & Bathala, 2007). A practical integration process should respect existing resources and provide metadata harvesting or on-the-fly searching to distributed catalogs based on community consensus on a digital data rights management schema.
- Quality assurance: the quality of information and resources within a GCI is not yet easy to evaluate (Giersch, Leary, Palmer, & Recker, 2008), and it remains a daunting task to manage federated virtual organizations (Gemmill et al., 2009). The uncertainty introduced in the process of a workflow within a complex decision support system is a critical issue. Advancements in quality related research within GCIs are urgently needed to better categorize, locate, select, and utilize more proper geospatial resources that address scientific and public problems.

6.3. Utilizing the Semantic Web to support building knowledge and semantics

Utilizing the Semantic Web to support building knowledge and semantics into the next generation of scientific tools will support smart processing of geospatial metadata, data, information, knowledge, and services for virtual communities and multiple scientific domains (Hendler, 2003). How to capture, represent, and integrate knowledge within and across geospatial domains are all ongoing challenges (Brodaric et al., 2009). Venkatasubramanian (2009a, 2009b) found that, in a data-rich world, we must find a way to automatically utilize knowledge acquired in the past to facilitate the automatic identification, utilization, and integration of datasets into operational systems. A semantic (ontology) based framework that is sensitive to the scale, richness, character, and heterogeneity within and across disciplines is desirable (Patterson, Faulwetter, & Shipunov, 2008). Transforming and integrating informal ontologies into formal community-accepted ontologies is a further challenge (Lumb, Freemantle, Lederman, & Aldridge, 2009). Multilingual ontologies present the challenge of matching the meanings of relevant terminologies within and across languages, such as those encountered in INSPIRE (Masser, 2007). Future research with GCIs will require us to collaborate with linguistics and translation professionals. Cross-domain ontologies will have similar benefits to multidisciplinary pursuits.

6.4. Developing geospatial middleware to provide functional and intermediate services and support service evolution to stakeholders

The GCIs envisioned cannot be built overnight and require sustainable and long-term planning, coordination, and maintenance. Therefore, an iterative process will provide intermediate services that benefit initial stakeholders and entice additional stakeholders by revising and introducing more services. The involvement of new stakeholders will also provide potential solutions to the generic challenge of maintaining a GCI after agency research or development funding is exhausted (Mackie, 2008). For example, the GOS now provides services to a wide variety of communities (Goodchild, Fu, & Rich, 2007). Google Earth and other popular mapping systems broaden the stakeholders of these GCIs from traditional geospatial sciences to the general public (Flora, 2007). The continuing evolution and success of such GCI examples and services will add more scientific research, models, and decision support analyses and provide concrete building blocks for future GCIs.

6.5. Advancing citizen sciences or public-based sciences

Advancing citizen sciences or public-based sciences will reflect the fact that cyberspace is open to the public and citizen participation will play vital roles, such as providing the following (Goodchild, 2008b): (a) Volunteer Geographic Information (VGI), (b) ratings for the quality of data, information, and services, (c) end-users to operate and test a GCI and provide feedback, and (d) data for the statistical analysis of a GCI. The full utilization of citizen contributions to control the quality of a GCI will require the active involvement of both social and relevant domain scientists (de Longueville, 2010).

6.6. Advancing GCI to geospatial cloud computing will implement transparent and opaque platforms

Advancing GCI to geospatial cloud computing will implement transparent and opaque platforms for addressing fundamental science questions and application problems through advancements in information technology and computing sciences.

- Maturation of adaptive computing: it is desirable to have a new GCI framework in which remote and in situ atmospheric sensors, data acquisition and storage systems, assimilation and prediction codes, data mining and visualization engines, and the information technology frameworks within which they operate can change configuration automatically in response to evolving weather (Droegemeier, 2009). Similar automated adaptability is desirable in other sciences, such as those that study ecological dynamics and climate change.
- The advancement of computer science and engineering is needed to (a) develop new models and methods for mass data transmission and storage, which is a challenge for all geospatial domains (Ashenfelder et al., 2009), (b) support virtual science by simulating algorithms and mechanisms within a computing environment, such as stress testing (Harris & Impelluso, 2008), biodefense (Zhang et al., 2008), and massive parallel analyses of regional earthquake activities (Zhang, Shi, & Wu, 2009), (c) transform science and engineering through collaborative, long-term, solid support from multiple domains (Freeman, Crawford, Kim, & Munoz, 2005 and NSF, 2007), and (d) ensure information security in algorithms, methods, framework integration, and solutions (Raghu & Chen, 2007; Tadiparthi & Sueyoshi, 2008).
- Advancing system architecture will support the following: (a) geospatial quality monitoring (Yang, Cao, et al., 2007), (b) system architecture optimization for better performance and results, (c) scalability and reliability for dynamic adoption (Li et al., 2010), and (d) automatic chaining of GCI components (Yue, Di, Yang, Yu, & Zhao, 2007).
- Evolution of computing to transparent and opaque cloud computing is needed to help to build end-to-end systems for end-users to inexpensively and transparently deploy computational resources to store, manipulate, and query large data sets to facilitate new science, such as environmental discovery (Govindaraju et al., 2009), and to relieve scientists and decision makers of the technical details of a GCI.

6.7. Developing a research agenda that addresses the needs of GCIs through effective federation and collaboration across communities

Developing a research agenda that addresses the needs of GCIs through effective federation and collaboration across communities, such as government agencies, non-government organizations, industries, academia, and the public is ideal. The federation/coordination and interconnection of stove-piped community and domain GCIs are needed to foster collaborations across domains, communities (de Assuncao, Buyya, & Venugopal, 2007), and continents (Peters et al., 2008). This agenda can leverage existing research facilities and GCIs including their data, analysis, metadata, and processing and analytical capabilities to maintain provenance (Simmhan et al., 2008) and promote the development of new tools to handle the integrated data, information, and knowledge. The agenda should integrate the requirements for a cross-scale, cross-domain, open, and interoperable approach to solve multidisciplinary problems, such as forecasting global environmental change at multiple spatial

Table 1
Direct connections between enabling technologies and future research needs.

	6.1 Social	6.2 Information	6.3 Semantics	6.4 Middleware	6.5 Citizen science	6.6 Cloud computing	6.7 Research agenda
3.1 Observation		×	×		×	×	×
3.2 SDI	×	×	×	×	×	×	×
3.3 DGIP		×	×	×		×	×
3.4 Web	×	×	×	×	×	×	×
3.5 Interoperability	×	×	×	×	×	×	×
3.6 HPC		×	×	×		×	×
3.7 Middleware		×	×			×	×
3.8 Representation	×	×	×		×	×	×
3.9 Cross domain	×		×		×	×	×
3.10 Knowledge	×	×	×	×	×	×	×
3.11 Architecture	×	×	×	×		×	×

Table 2
Direct connections between GCI functions and future research needs.

	6.1 Social	6.2 Information	6.3 Semantics	6.4 Middleware	6.5 Citizen science	6.6 Cloud computing	6.7 Research agenda
4.1 Data processing		×	×	×	×	×	×
4.2 Data integration	×	×	×	×	×	×	×
4.3 Preservation and accessibility	×	×	×	×	×	×	×
4.4 Data life cycle	×	×	×	×	×	×	×
4.5 VO	×		×		×	×	×
4.6 Semantics	×	×	×		×	×	×
4.7 HPC			×	×		×	×
4.8 LBS	×	×	×		×	×	×
4.9 Cross domain	×		×		×	×	×

scales (e.g., from local sites to regions and continents) and utilizing multi-scale, multi-regional weather forecasting, land cover, and soil type data for agricultural forecasting (Baker et al., 2009) as well as for ecology (Jones, 2007) and biological sciences (Yu et al., 2008). The agenda should also include cross-geographic and cross-cultural integration and evolution of social, behavioral, and economic research within multiple contexts (Lightfoot et al., 2009).

The GCIs of the future will require a new generation of research tools, techniques, and educational support (Schlager, Farooq, Fusco, Schank, & Dwyer, 2009), and a new generation of scientists and engineers who can support, learn, and collaborate with one another more effectively in cyber-enabled professional communities. This advancement is needed in a broad stream of science domains (Newman, Ellisman, & Orcutt, 2003) and will require all application domains to provide requirements, tests, and maturation environments for GCIs. The seven aspects described will benefit from and drive the enabling technologies for future GCIs (Table 1) and provide improved functions for GCI application domains (Table 2). More direct connections of an item with other items illustrate broader needs for the item.

Acknowledgements

The authors acknowledge the support of a variety of GCI projects: a NASA interoperability project NNX07AD99G, an FGDC NSDI portal project G09AC00103, and a NASA HPC project SMD-08-0768.

References

Agrawal, G., Ferhatosmanoglu, H., Niu, X., Bedford, K., Li, R. (2006). A vision for cyberinfrastructure for coastal forecasting and change analysis. Lecture Notes in Computer Science (LNCS) 4540. In *GeoSensor networks – second international conference*, GSN 2006, October 1–3, 2006, Boston, MA, United States, 2008 (pp. 151–174).

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). *IEEE Communications Magazine*, 102–114.

Ashenfelder, M., Boyko, A., Mandelbaum, J., Minor, D., McDonald, R., & Moore, R. (2009). NDIIPP models for mass data transmission and storage. *Library Trends, 57*(3), 541–555.

Avery, P. (2008). Open science grid: Building and sustaining general cyberinfrastructure using a collaborative approach. *First Monday, 12*(6), 15.

Baker, K. S., & Bowker, G. C. (2007). Information ecology: Open system environment for data, memories, and knowing. *Journal of Intelligent Information Systems: Integrating Artificial Intelligence and Database Technologies, 29*(1), 127–144.

Baker, K., Stein, J., Wharton, P., Paz, J., Kirk, W. W., & Plale, B. (2009). Cyberinfrastructure challenges to multi-regional, multi-scale weather forecasting for crop disease early warning systems. *Phytopathology, 99*(6), S7.

Ball, W. P., Brady, D. C., Brooks, M. T., Burns, R., Cuker, B. E., Di Toro, D. M., et al. (2008). Prototype system for multidisciplinary shared cyberinfrastructure: Chesapeake bay environmental observatory. *Journal of Hydrologic Engineering, 13*(10), 960–970.

Bambacus, M., Yang, P., Evans, J., Cole, M., Alameh, N., & Marley, S. (2007). ESG: An interoperable portal for prototyping applications. *URISA Journal, 19*(2), 15–21.

Bartrand, T. A., Weir, M., & Haas, C. N. (2007). Advancing the quality of drinking water: Expert workshop to formulate a research agenda. *Environmental Engineering Science, 24*(7), 863–872.

Berman, F. (2008). Got data? A guide to data preservation in the information age. *Communications of the ACM, 51*(12), 50–56.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*.

Blais, J. A. R., & Esche, H. (2008). Geomatics and the new cyber-infrastructure. *Geomatica, 62*(1), 11–22.

Bogden, P., Allen, G., Creager, G., Graves, S., Luettich, R., Ramakrishnan, L. (2006). Designing a collaborative cyberinfrastructure for event-driven coastal modeling. In *Proceedings of the 2006 ACM/IEEE conference on supercomputing*, SC'06, November 11–17, 2006, Tampa, FL, 2006 (p. 1188646).

Borgman, C. L., Wallis, J. C., Mayernik, M. S., Pepe, A. (2007). Drowning in data: Digital library architecture to support scientific use of embedded sensor networks. In *Proceedings of the 7th ACM/IEEE joint conference on digital libraries, JCDL 2007: Building and sustaining the digital environment*, June 18–23, 2007, Vancouver, BC, 2007 (p. 269–277).

Brodaric, B., Fox, P., & McGuinness, D. L. (2009). Geoscience knowledge representation in cyberinfrastructure. *Computers and Geosciences, 35*(4), 697–699.

Bugallo, M. F., Marx, M., Bynum, D., Takai, H., Hover, J. (2009). Bringing science and engineering to the classroom using mobile computing and modern cyberinfrastructure. In CSEDU 2009 – *Proceedings of the 1st international conference on computer supported education*, (1) (pp. 206–209).

Buhr, S., Barker, L., Reeves, T. C. (2005). The DLESE evaluation services group: A framework for evaluation within a digital library. In *Proceedings of the 5th ACM/IEEE joint conference on digital libraries – digital libraries: Cyberinfrastructure for research and education*, June 7–11, 2005, Denver, CO, IEEE, 2005 (p. 370).

Cao, Y., Yang, C., & Wong, D. (2009). An interoperable spatiotemporal weather radar data disseminating system. *International Journal of Remote Sensing, 30*, 1313–1326.

Carter, P. E., & Green, G. (2009). Networks of contextualized data: A framework for cyberinfrastructure data management. *Communications of the ACM, 52*(2), 105–109.

Catlett, C. E., (2005). TeraGrid: A foundation for US cyberinfrastructure. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v 3779 LNCS. In *Network and parallel computing – IFIP international conference, NPC 2005, proceedings*, November 30–December 3, 2005, Beijing, China (p. 1).

Chang, C. H., Lim, E. P., Hedberg, J. G., The, T. S., Theng, Y. L., Goh, D. H. L. (2005). Evaluating G-Portal for geography learning and teaching. In *Proceedings of the 5th ACM/IEEE joint conference on digital libraries – digital libraries: Cyberinfrastructure for research and education*, June 7–11 2005, Denver, CO (p. 21–22).

Chiang, G. T., Dove, M., Ballard, S., Bostater, C., Frame, I. (2006). A Grid Enabled Monte Carlo Hyperspectral Synthetic Image Remote Sensing Model (GRID-MCHSIM) for coastal water quality algorithm. In *Proceedings of SPIE – the international society for optical engineering, v 6360, remote sensing of the ocean, sea ice, and large water regions 2006*, September 11–13, 2006, Stockholm, Sweden (p. 636009).

Chiu, K., & Fox, G. C. (2009). Special section: Third IEEE international conference on e-science and grid computing. *Future Generation Computer Systems, 25*(4), 444–445.

Clark, H. L. (2001). New sea floor observatory networks in support of ocean science research. In *Proceedings of oceans 2001 MTS/IEEE conference*, Honolulu, Hawaii, 5–8 November 2001.

Clarkson, G. (2008). Cyberinfrastructure and patent thickets: Challenges and responses. *First Monday, 12*(6), 18.

Contractor, N., Hesse, B. W. (2006). Cyberinfrastructure for public health. In *Proceedings of the 7th annual international conference on digital government research*, May 21–24, 2006, San Diego, CA, 2006 (pp. 9–10).

Cruz, I. F., Sunna, W., Makar, N., & Bathala, S. (2007). A visual tool for ontology alignment to enable geospatial interoperability. *Journal of Visual Languages and Computing, 18*(3), 230–254.

Cummings, J., Finholt, T., Foster, I., Kesselman, C., & Lawrence, K. A. (2008). *Beyond being there: A blueprint for advancing the design, development, and evaluation of virtual organizations*. Arlington, VA: National Science Foundation. 58 p.

Datcu, M., Daschiel, H., Pelizzari, A., Quartulli, M., Galoppo, A., Colapicchioni, A., et al. (2003). Information mining in remote sensing image archives: System concepts. *IEEE Transactions on Geoscience and Remote Sensing, 41*(12), 2923–2936.

de Assuncao, M. D., Buyya, R., & Venugopal, S. (2007). InterGrid: A case for internetworking islands of Grids. *Concurrency and Computation – Practice and Experience, 20*(8), 997–1024.

de Longueville, M. B. (2010). Community-based GeoPortals: The next generation? Concepts and methods for the geospatial web 2.0. *Computers, Environment and Urban Systems, 34*(4), 299–308.

de Smith, M. J., Longley, P. A., & Goodchild, M. F. (2007). *Geospatial analysis: A comprehensive guide to principles, techniques and software tools*. Winchelsea Press.

Deelman, E., Gannon, D., Shields, M., & Taylor, I. (2009). Workflows and e-Science: An overview of workflow system features and capabilities. *Future Generation Computer Systems, 25*(5), 528–540.

Delserone, L. M. (2009). At the watershed: Preparing for research data management and stewardship at the university of minnesota libraries. *Library Trends, 57*(2), 202–210.

Digital National Framework, 2007. Unique Object Identifiers, 12 pp. <http://www.dnf.org/images/uploads/guides/DNF0011_v1-1_UniqueObjectIdentifiers.pdf>.

Droegemeier, K. K. (2009). Transforming the sensing and numerical prediction of high-impact local weather through dynamic adaptation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 367*(1890), 885–904.

Egenhofer, M. J. (2002). Toward the Semantic Geospatial Web. In *Proceedings of the 10th ACM international symposium on advances in geographic information systems*, McLean, Virginia.

Elmagarmid, A. K., Samuel, A., & Ouzzani, M. (2008). Community-cyberinfrastructure-enabled discovery in science and engineering. *Computing in Science and Engineering, 10*(5), 46–53.

Embley, D. W. (2004). Toward semantic understanding: An approach based on information extraction ontologies. In K. Schewe & H. Williams (Eds.), *Proceedings of the 15th Australasian database conference – volume 27* (Dunedin, New Zealand). ACM International Conference Proceeding Series, vol. 52. Australian Computer Society, Darlinghurst, Australia (pp. 3–12).

Eschenbach, E. J., James, H., Brus, C., Carlson, P., Giammar, D., Grauer, B., Hotaling, L., Oguntimein, G., Safferman, S., Wentling, T. (2006). Transforming K-12 education via the collaborative large-scale engineering analysis network for environmental research (CLEANER) project. In *Proceeding of ASEE annual conference and exposition, conference proceedings, 2006 ASEE annual conference and exposition*, June 18–21, 2006, Chicago, IL. 13 p.

Fischer, G. (2009). End-user development and meta-design: Foundations for cultures of participation. *LNCS, 5435*(3–14).

Flora, B. (2007). Google earth impact: Saving science dollars and illuminating geoscience. *ECONTENT, 30*(3), 12–13.

Fox, G., & Pierce, M. (2009). Grids challenged by a Web 2.0 and multicore sandwich. *Concurrency Computation – Practice and Experience, 21*(3), 265–280.

Freeman, P. A., Crawford, D. L., Kim, S., & Munoz, J. L. (2005). Cyberinfrastructure for science and engineering: Promises and challenges. *Proceedings of the IEEE, 93*(3), 682–691.

Freudinger, L.C. (2009). Cyberinfrastructure for Airborne Sensor Webs. *NASA Dryden Flight Research Center Report: DFRC-927*, 5p.

Friis-Christensen, A., Lucchi, R., Lutz, M., & Ostlinder, N. (2009). Service chaining architectures for applications implementing distributed geographic information processing. *IJGIS, 23*(5), 561–580.

Gahegan, M., Luo, J., Weaver, S. D., Pike, W., & Banchuen, T. (2009). Connecting GEON: Making sense of the myriad resources, researchers and concepts that comprise a geoscience cyberinfrastructure. *Computers and Geosciences, 35*(4), 836–854.

Gemmill, J., Robinson, J. P., & Scavo, T. (2009). Cross-domain authorization for federated virtual organizations using the my Vocs collaboration environment. *Concurrency Computation – Practice and Experience, 21*(4), 509–532.

Giersch, S., Leary, H., Palmer, B., Recker, M. (2008). Developing a review process for online resources. In *Joint Conference on Digital Libraries (JCDL 2008)*, June 15–19, 2008, Austin, TX, 457.

Goodall, J. L., Jeffery, S., Whiteaker, T. L., Maidment, D. R., & Zaslavsky, I. (2008). A first approach to web services for the National Water Information System. *Environmental Modelling and Software, 23*(4), 404–411.

Goodchild, M. F. (2008a). The use cases of digital earth. *International Journal of Digital Earth, 1*(1), 31–42.

Goodchild, M. F. (2008b). Commentary: Whither VGI? *GeoJournal, 72*, 239–244.

Goodchild, M. F., Fu, P. D., & Rich, P. (2007). Sharing geographic information: An assessment of the geospatial one-stop. *Annals of the Association of American Geographers, 97*(2), 250–266.

Govindaraju, R. S., Engel, B., Ebert, D., Fossum, B., Huber, M., Jafvert, C., et al. (2009). Vision of cyberinfrastructure for end-to-end environmental explorations (C4E4). *Journal of Hydrologic Engineering, 14*(1), 53–64.

Graham, J., Simpson, A., Crall, A., Jarnevich, C., Newman, G., & Stohlgren, T. J. (2008). Vision of a cyberinfrastructure for nonnative, invasive species management. *Bioscience, 58*(3), 263–268.

Harris, R., & Impelluso, T. (2008). Virtual stress testing machine and the cyber-infrastructure. *Engineering with Computers, 24*(2), 107–117.

Hart, J. K., & Martinez, K. (2006). Environmental sensor networks: A revolution in the earth system science? *Earth–Science Reviews, 78*(3–4), 177–191.

Hendler, J. (2003). Enhanced: Science and the semantic web. *Science, 299*(5606), 520–521.

Henry, C. J., Friedlander, A. (2008). Cataloging hidden special collections and archives: Building a new research environment. In *Joint conference on digital libraries (JCDL 2008)*, June 15–19, 2008, Austin, TX (p. 439).

Hey, T., & Trefethen, A. E. (2005). Cyberinfrastructure for e-Science. *Science, 308*(5723), 817–821.

Holzmann, G. J., & Pehrson, B. (1994). *The early history of data networks (perspectives)*. Wiley-IEEE Computer Society Press. 304 pp.

Horsburgh, J. S., Tarboton, D. G., Piasecki, M., Maidment, D. R., Zaslavsky, I., Valentine, D., et al. (2009). An integrated system for publishing environmental observations data. *Environmental Modelling and Software, 24*(8), 879–888.

Hornsby, S. K., & Yuan, M. (Eds.). (2008). *Understanding dynamics of geographic domains*, New York, NY: CRC Press, 240p.

Jabeur, N., McCarthy, J. D., Xing, X., & Graniero, P. A. (2009). A knowledge-oriented meta-framework for integrating sensor network infrastructures. *Computers and Geosciences, 35*(4), 809–819.

Jensen, K., Chen, J., & Murray, F. (2008). A simple method to improve life sciences patent searches using the cyberinfrastructure at the National Institutes of Health. *First Monday, 12*(6), 7.

Jones, M. B. (2007). Meta-information systems and ontologies – A special feature from ISEI'06. *Ecological Informatics, 2*(3), 193–194.

Kahhat, R., Kim, J., Xu, M., Allenby, B., Williams, E., & Zhang, P. (2008). Exploring e-waste management systems in the United States. *Resources, Conservation and Recycling, 52*(7), 955–964.

Katz, D. S., Allen, G., Cortez, R., Cruz-Neira, C., Gottumukkala, R., Greenwood, Z. D., et al. (2009). Louisiana: A model for advancing regional e-research through cyberinfrastructure. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences, 367*(1897), 2459–2469.

Keating, T. (2009). Cyberinfrastructure for Air Quality Management, EPA internal document, 10.

Kessler, H., Mathers, S., & Sobisch, H. G. (2009). The capture and dissemination of integrated 3D geospatial knowledge at the British Geological Survey using GSI3D software and methodology. *Computers and Geosciences, 35*(6), 1311–1321.

Kido, M. H., Mundt, C. W., Montgomery, K. N., Asquith, A., Goodale, D. W., & Kaneshiro, K. Y. (2008). Integration of wireless sensor networks into cyberinfrastructure for monitoring Hawaiian "mountain-to-sea" environments. *Environmental Management, 42*(4), 658–666.

Kupper, A. (2005). *Location-based Services: Fundamentals and Operation*. Wiley. 386 p.

Lannotta, B. (2007). GEOSS: A global view. *Aerospace America, 45*(8), 36–40.

Lassila, O., & Hendler, J. (2007). Embracing "Web 3.0". *Internet Computing, 11*(3), 90–93.

Lawrence, K. A., Finholt, T. A., Kim, I. (2007). Warm fronts and high pressure systems: Overcoming geographic dispersion in a meteorological cyberinfrastructure project. In *Proceedings of the 40th annual hawaii international conference on system sciences*, 3–6 January 2007, Waikoloa, HI, USA, 10 pp.

Ledford, H. (2009). Cyberinfrastructure: Feed me data. *Nature, 459*(7250), 1047–1049.

Lerner, F. A. (2001). *The story of libraries: From the invention of writing to the computer age* (2nd Ed.). Continuum International Publishing Group. 246 p.

Li, X., Kim, Y. J., Govindan, R., Hong, W. (2003). Multi-dimensional range queries in sensor networks. In *Proceedings of the 1st international conference on Embedded networked sensor systems*, November 5, 2003, Los Angeles, California, USA (pp. 63–75).

Li, W., Yang, C., and Raskin, R. (2008). A semantic enhanced model for searching in spatial web portals. In: *Proceeding of AAAI/SSKI Spring Workshop*, Palo Alto, CA (pp. 47–50).

Li, W., Yang, C., & Yang, C. (2010). An active crawler for discovering geospatial web services and their distribution pattern – A case study of OGC Web Map Service. *International Journal of Geographic Information Science*. doi:10.1080/13658810903514172.

Li, W., Yang, C., & Sun, D. (2009). Mining the correlations and rules of geophysical parameters' contribution to tropical cyclones. *Computers & Geosciences, 35*(2), 309–316.

Lightfoot, D. W., Bachrach, C., Abrams, D., Kielman, J., & Weiss, M. (2009). *Social, behavioral and economic research in the federal context*. National Science and Technology Council. 65 p.

Loudon, T. V. (2009). Four interacting aspects of a geological survey knowledge system. *Computers and Geosciences, 35*(4), 700–705.

Lubin, D., Ayres, G., Hart, S., Lubin, D., Ayres, G., & Hart, S. (2009). Remote sensing of polar regions lessons and resources for the international polar year. *Bulletin of the American Meteorological Society, 90*(6), 825.

Lumb, L. I., Freemantle, J. R., Lederman, J. I., & Aldridge, K. D. (2009). Annotation modeling with formal ontologies: Implications for informal ontologies. *Computers and Geosciences, 35*(4), 855–861.

MacEachren, A. M., & Brewer, I. (2004). Developing a conceptual framework for visually-enabled geocollaboration. *International Journal of Geographical Information Science, 18*(1), 1–34.

MacEachren, A. M., & Kraak, M.-J. (2001). Research challenges in geovisualization. *Cartography and Geographic Information Science, 28*(1), 3–12.

Maguire, D. J., Goodchild, M. F., & Rhind, D. W. (Eds.). (1991). *Geographical information systems: Principles and applications* (pp. 1056). New York: Wiley.

Mackie, C. J. (2008). Cyberinfrastructure, institutions, and sustainability. *First Monday, 12*(6), 15.

Maguire, D. J., & Longley, P. A. (2005). The emergence of geoportals and their role in spatial data infrastructures. *Computer, Environment, and Urban Systems, 29*(1), 3–14.

Mahinthakumar, K., von Laszewski, G., Ranjithan, R., Brill, D., Uber, J., Harrison, K., Sreepathi, S., Zechman, E. (2006). An adaptive cyberinfrastructure for threat management in urban water distribution systems. In *Proceedings of computational science-ICCS 2006. Part III (lecture notes in computer science, vol. 3993)*, May 2006, Reading (pp. 401–408).

Marshall, J. J., Downs, R. R. (2008). Reuse readiness levels as a measure of software reusability. In *2008 IEEE international geoscience and remote sensing symposium – proceedings*, vol. 3(1) (pp. III1414–III1417).

Masser, I. (2007). *Building european spatial data infrastructures*. ESRI Press (p. 91).

Maué, P., & Schade, S. (2009). Data integration in the geospatial Semantic Web. *Journal of Cases on Information Technology, 11*(4), 100–122.

McArthur, D. J., Zia, L. L. (2008). From NSDL 1.0 to NSDL 2.0: Towards a comprehensive cyberinfrastructure for teaching and learning. In *Joint conference on digital libraries (JCDL 2008)*, June 15–19, 2008, Austin, TX (pp. 66–69).

Meyer, E. T., & Schroeder, R. (2009). Untangling the Web of E-Research: Towards a sociology of online knowledge. *Journal of Informatics, 3*(3), 246–260.

Minsker, B., Myers, J., Marikos, M., Wentling, T., Downey, S., Liu, Y., Bajcsy, P., Kooper, R., Marini, L., Contractor, N., Green, H. D., Futrelle, J. (2006). NCSA environmental cyberinfrastructure demonstration project: Creating cyber environments for environmental engineering and hydrological science communities. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, SC'06, 2006, November 11–17, 2006, Tampa, FL, United States (p. 1188622).

Morisette, J. T., Richardson, A. D., Knapp, A. K., Fisher, J. I., Graham, E. A., Abatzoglou, J., et al. (2009). Tracking the rhythm of the seasons in the face of global change: Phenological research in the 21st century. *Frontiers in Ecology and the Environment, 7*(5), 253–260.

Murthy, C. S., Manimaran, G. (2001). *Resource management in real-time systems and networks*. The MIT Press, 464 p.

Myhill, W. N., Cogburn, D. L., Samant, D., Addom, B. K., & Blanck, P. (2008). Developing accessible cyberinfrastructure-enabled knowledge communities in the national disability community: Theory, practice, and policy. *Assistive Technology, 20*(3), 157–174.

Nebert, D. (2004). GSDI Cook Book, Version 2.0. 171 p.

Newman, H. B., Ellisman, M. H., & Orcutt, J. A. (2003). Data-intensive e-science frontier research. *Communications of the ACM, 46*(11), 68–77.

Nittel, S., Stefanidis, A., Cruz, I., Egenhofer, M., Goldin, D., Howard, A., et al. (2004). Report from the first workshop on geo sensor networks. *ACM SIGMOD Record, 33*(1), 141–144.

NSF. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the national science foundation blue-ribbon advisory panel on cyberinfrastructure*. NSF Panel reports. 84 pp.

NSF. (2007). *Cyberinfrastructure vision for 21st century discovery*. 64 pp.

NSF. (2009). *NSF-supported research infrastructure: Enabling discovery, innovation and learning*, NSF, NSF-09-13, 148 p.

Olson, G. M. (2009). The next generation of science collaboratories. In *2009 International Symposium on Collaborative Technologies and Systems (CTS 2009)* (pp. xv–xvi).

Papadopoulos, A. I., & Linke, P. (2009). A decision support grid for integrated molecular solvent design and chemical process selection. *Computers and Chemical Engineering, 33*(1), 72–87.

Patterson, D. J., Faulwetter, S., & Shipunov, A. (2008). Principles for a names-based cyberinfrastructure to serve all of biology. *Zootaxa, 1950*, 153–163.

Pennington, D., Michener, D., William, K., Katz, S., Downey, L., & Schildhauer, M. (2008). Transforming scientists through technical education a view from the trenches. *Computing in Science and Engineering, 10*(5), 28–33.

Peters, D. P. C., Groffman, P. M., Nadelhoffer, K. J., Grimm, N. B., Coffins, S. L., Michener, W. K., et al. (2008). Living in an increasingly connected world: A framework for continental-scale environmental science. *Frontiers in Ecology and the Environment, 6*(5), 229–237.

Pezzoli, K., Marciano, R., Robertus, J. (2006). Regionalizing integrated watershed management: A strategic vision. In *ACM international conference proceeding series – proceedings of the 7th annual international conference on digital government research*, May 21–24, 2006, San Diego, CA (pp. 444–445).

Pierce, M. E., Fox, G. C., Aktas, M. S., Aydin, G., Gadgil, H., Qi, Z. G., et al. (2008). The QuakeSim project: Web services for managing geophysical data and applications. *Pure and Applied Geophysics, 165*(3–4), 635–651.

Pierce, M. E., Fox, G. C., Choi, J. Y., Guo, Z., Gao, X., & Ma, Y. (2009). Using Web 2.0 for scientific applications and scientific communities. *Concurrency Computation – Practice and Experience, 21*(5), 583–603.

Raghu, T. S., & Chen, H. (2007). Cyberinfrastructure for homeland security: Advances in information sharing, data mining, and collaboration systems. *Decision Support Systems, 43*(5), 1321–1323.

Ramaprivan, H. K. (2008). NASA's Earth Science Data Systems: A 'Bit of History' and Observations. National Science Foundation workshop, May 2008, 25 p.

Raskin, R., & Pan, M. (2005). representation in the semantic web for Earth and environmental terminology (SWEET). *Computers and Geosciences, 31*(9), 1119–1125.

Reynolds, R. G., & Zhu, S. (2001). Knowledge-based function optimization using fuzzy cultural algorithms with evolutionary programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 31*(1), 1–18.

Ribes, D., & Finholt, T. A. (2009). The long now of technology infrastructure: Articulating tensions in development. *Journal of the Association for Information Systems, 10*(5), 375–398.

Rich, P. M., Weintraub, L. H. Z., Ewers, M. E., Riggs, T. L., Wilson, C. J. (2005). Decision support for water planning: The ZeroNet water-energy initiative. In *Proceedings of the 2005 World Water and Environmental Resources Congress*, May 15–19, 2005, Anchorage, AK (p. 468).

Roman, D., Klien, E. (2007). SWING – A semantic framework for geospatial services. In A. Scharl & K. Tochtermann (Eds.), *The geospatial web: How geobrowsers, social software and the web 2.0 are shaping the network society* (pp. 229–234).

Schlager, M. S., Farooq, U., Fusco, J., Schank, P., & Dwyer, N. (2009). Analyzing online teacher networks cyber networks require cyber research tools. *Journal of Teacher Education, 60*(1), 86–100.

Schofield, O., Glenn, S., Chant, R., Kohut, J., McDonnell, J. (2007). Implementing a new paradigm in ocean observing: A view from the coolroom, Oceans Conference Record (IEEE). In *Proceeding of oceans 2007 MTS/IEEE conference*, September 29–October 4, 2007, Vancouver, BC (p. 4449209).

Shimbo, M., & Ishidab, T. (2003). Controlling the learning process of real-time heuristic search. *Artificial Intelligence, 146*(1), 1–41.

Simmhan, Y. L., Plale, B., & Gannon, D. (2008). Query capabilities of the Karma provenance framework. *Concurrency and Computation – Practice and Experience, 20*(5), 441–451.

Smarr, L., Brown, M., & de Laat, C. (2009). Special section: OptIPlanet – The OptIPuter global collaboratory. *Future Generation Computer Systems, 25*(2), 109–113.

Smith, M.A., Kollock, P. (1999). Communities in Cyberspace. Routledge Press (p. 336).

Stein, L. D. (2008). Wiki features and commenting – Towards a cyberinfrastructure for the biological sciences: Progress, visions and challenges. *Nature Reviews Genetics, 9*(9), 678–688.

Sucaet, Y., Van Hemert, J., Tucker, B., & Bartholomay, L. (2008). A web-based relational database for monitoring and analyzing mosquito population dynamics. *Journal of Medical Entomology, 45*(4), 775–784.

Tadiparthi, G. R., & Sueyoshi, T. (2008). A novel steganographic algorithm using animations as cover. *Decision Support Systems, 45*(4), 937–948.

Tsai, W. F., Huang, W. C., Lin, F. P., Hung, B., Wang, Y. T., Shiau, S., et al. (2008). The human-centered cyber infrasctructure for scientific and engineering grid applications. *Journal of the Chinese Institute of Engineers, 31*(7), 1127–1139.

Venkatasubramanian, V. (2009a). Cyberinfrastructure, informatics, and intelligent systems: Commentary on the new thematic area for Computers and Chemical engineering. *Computers and Chemical Engineering, 33*(1), 2–3.

Venkatasubramanian, V. (2009b). Drowning in data: Informatics and modeling challenges in a Data-Rich Networked World. *Aiche Journal, 55*(1), 2–8.

von Laszewski, G., Younge, A., Xi, H., Mahinthakumar, K., Lizhe, W. (2009). Experiment and workflow management using Cyberaide Shell. In *2009 9th IEEE/ACM international symposium on cluster computing and the grid (CCGrid 2009)* (pp. 568–573).

Wang, S. W., & Liu, Y. (2009). TeraGrid GIScience Gateway: Bridging cyberinfrastructure and GIScience. *International Journal of Geographic Information Science, 23*(5), 631–656.

Whey-Fone, T., Weicheng, H., Fang-Pang, L., Hung, B., Yao-Tsung, W., Shiau, S., et al. (2008). The human-centered cyberinfrastructure for scientific and engineering grid applications, vol. 31(7) (pp. 1127–1139).

White, D. L., Wolf, D., Porter, D. E., Sanger, D. M., Riekerk, G. H. M., DiDonato, G., et al. (2009). Development of a data management framework in support of southeastern tidal creek research. *Environmental Monitoring and Assessment, 150*(1–4), 323–331.

Wilkins-Diehr, N., Gannon, D., Klimeck, G., Oster, S., & Pamidighantam, S. (2008). TeraGrid science gateways and their impact on science. *Computer, 41*(11), 32–42.

Winer, E. H. (2006). Multidisciplinary design optimization. *Aerospace America, 44*(12), 41.

Winget, M. (2008). The working scientist and the realities of data curation: A qualitative study addressing attitudes and needs. In *Joint Conference on Digital Libraries (JCDL 2008)*, 474 p.

Wolf, E. B., Howe, K. (2009). Web-client based distributed generalization and geoprocessing. In: *Proceedings of the International Conference on Advanced Geographic Information Systems and Web Services*, GEOWS 2009 (pp. 123–128).

Wunsch-Vincent, S., Reynolds, T., & Wyckoff, A. (2008). Implementing openness: An international institutional. *First Monday, 12*(6), 15.

Xie, J., Yang, C., Zhou, B., & Huang, Q. (2010). High-performance computing for the simulation of dust storms. *Computers Environment and Urban Systems, 34*(4), 278–290.

Yang, C. (2000). Theory, Techniques and implementation methods of WebGIS. Unpublished Ph.D. dissertation. Beijing: Peking University (in Chinese).

Yang, C., Cao, Y., & Evans, J. (2007). WMS performance and client design principles. *The journal of GeoInformation Science and Remote Sensing, 44*(4), 320–333.

Yang, P., Evans, J., Cole, M., Alameh, N., Marley, S., & Bambacus, M. (2007). The emerging concepts and applications of the spatial web portal. *PE&RS, 73*(6), 691–698.

Yang, P., Li, W., Xiao, D., Raskin, R., Bambacus, M. (2007). Earth information exchange - Sharing earth science information through interoperable approach and cyberinfrastructure. In *Proceedings of SPIE - The International Society for Optical Engineering, v 6753, n PART 2, Geoinformatics 2007: Geospatial Information Science*, May 25–27, 2007, Nanjing, China (p. 675321).

Yang, C., Li, W., Xie, J., & Zhou, B. (2008). Distributed geospatial information processing: Sharing earth science information to support Digital Earth. *International Journal of Digital Earth, 1*(3), 259–278.

Yang, C., & Raskin, R. (2009). Introduction to distributed geographic information processing research. *International Journal of Geographic Information Science, 23*(5), 553–560.

Yang, C., Wong, D., Miao, Q., & Yang, R. (2010). *Advanced geoInformation science*. CRC Press. 608p.

Yu, E. T., Hawkins, A., Kuntz, I. D., Rahn, L. A., Rothfuss, A., Sale, K., et al. (2008). The Collaboratory for MS3D: A new cyberinfrastructure for the structural elucidation of biological macromolecules and their assemblies using mass spectrometry-based approaches. *Journal of Proteome Research, 7*(11), 4848–4857.

Yue, P., Di, L., Yang, W., Yu, G., & Zhao, P. (2007). Semantics-based automatic composition of geospatial Web service chains. *Computers and Geosciences, 33*(5), 649–665.

Zhang, C., Crasta, O., Cammer, S., Will, R., Kenyo, R., Sullivan, D., et al. (2008). An emerging cyberinfrastructure for biodefense pathogen and pathogen-host data. *Nucleic Acids Research, 36*, D884–D891.

Zhang, T., Tsou, M.H., Qiao, Q.H., Xu, L. (2006). Building an intelligent geospatial cyberinfrastructure: An analytical problem solving approach. In *Proceedings of SPIE - The International Society for Optical Engineering, v 6420, Geoinformatics 2006: Geospatial Information Science*, October 28–29 2006, Wuhan, China (p. 64200A).

Zhang, H., Shi, Y. L., & Wu, Z. L. (2009). A framework of massively parallel analysis of regional earthquake activities. *ACTA Geologica Sinica – English Edition, 83*(4), 786–800.

Zhang, T., & Tsou, M. H. (2009). Developing a grid-enabled spatial Web portal for Internet GIServices and geospatial cyberinfrastructure. *International Journal of Geographic Information Science, 23*(5), 605–630.