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.

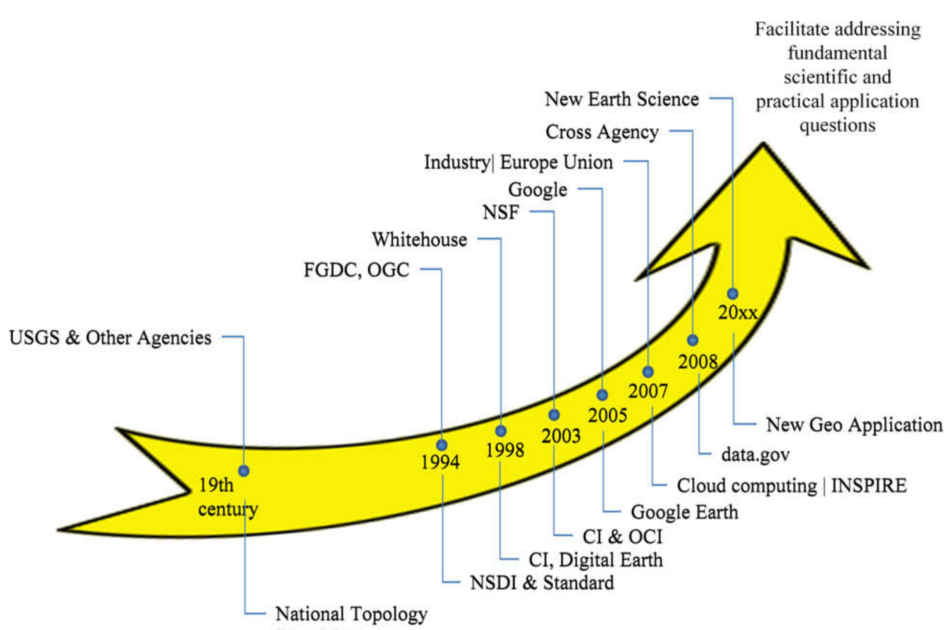

Fig. 1. History of geospatial Cyberinfrastructure.