

The role of spatial cyberinfrastructure in geoinformation systems

Suri Nuriev^{1*}, *Igor Kartsan*^{1,2}

¹Marine Hydrophysical Institute, Russian Academy of Sciences», 299013, Sevastopol, Russia

²Reshetnev Siberian State University of Science and Technology, 648361, Krasnoyarsk, Russia

Abstract. The potential future of a spatial cyberinfrastructure is considered, combining advanced computer, information and communication technologies to expand the possibilities of computational and information scientific practice and improve the synthesis and analysis of scientific data through collaboration and sharing. Currently, spatial cyberinfrastructure represents a paradigm shift in scientific research that has facilitated easy access to computational utilities and facilitated collaboration across distances and disciplines, thereby enabling scientific breakthroughs to be achieved faster and more efficiently. Spatial cyberinfrastructure seeks to address the long-standing complex problems of processing and analyzing massive and heterogeneous spatial datasets, as well as the need and benefits of flexible and secure spatial data exchange. The focus is on spatial analysis using distributed and high performance computing, sensor networks and other advanced information technology capabilities to transform massive spatial datasets into information and knowledge. To solve large-scale and multi-scale problems, the spatial cyberinfrastructure, despite the complexity of the geographic space, allows transforming spatial analysis into powerful and affordable computing utilities to provide broad scientific breakthroughs when creating the well-known Digital Earth project. Critical analysis and recommendations are provided on issues of conflicting data that researchers do not want to exclude from various data models and geographic scales. This case study shows the value of spatial cyberinfrastructure in addressing the challenges that arise from spatial as well as multilingual, biographical and temporal ambiguities in various databases, solutions that, again, would not be possible without spatial cyberinfrastructure.

1 Introduction

Cyberinfrastructure (CI) was first coined by the National Science Foundation Blue-Ribbon Committee [1] to reflect how traditional ways of scientific research (e.g. experiments in the lab, observation in the field, processing/analysis on a single computer, and even calculations on the back of sheets of paper) are being improved and even revolutionized by the integration capabilities of high-performance computers, storage and visualization facilities for very large datasets, digitally supported sensors and tools in the environment,

* Corresponding author: surinuriev@gmail.com

virtual organizations for collaborative problem solving, and compatible packages of software services and tools [2]. The world of scientific publishing is transforming as part of the evolution of CI [3].

Examples include the discovery of abrupt transitions in Earth's climate and ecosystem dynamics, previously unknown properties of minerals at extreme temperatures and pressures deep within the Earth, simulations of the development of the early universe, development of improved ocean models, understanding of individual and group behavior and its relationship to social, economic and political structures and creation of a genetic map of connections between people [2, 4, 5]. Computation, along with theory and experiment, has become the "third pillar" of science and engineering [6]. In addition, scientific discovery requires the computational ability to synthesize and analyze very large data sets, which are integrated into the biological, physical, and social sciences and engineering, as well as the science and technology interface, where database-intensive science is defined as the "fourth paradigm" [5]. CI has evolved not only into hardware and software, but also into its own emerging area of research in the field of knowledge-intensive and digital libraries [5-9], with many countries investing in research and development of CI [10-11]. Achieving high performance computing at an affordable cost and showing significant advances in simulation tools, many challenges remain in the efficient integration of field stations containing thousands of instruments, millions of users and petabytes of data with analysis capability [5].

Spatial CI is a new term in the literature [14-16], and it is defined as a specific type of CI that synergistically integrates the capabilities of CI, geographic information systems (GIS) [17, 18] and spatial analysis [19, 20] to solve geospatial problems and decision making. By spatial or space, we mean both real, physical space (i.e. on the surface of the Earth, in the atmosphere or ocean) and virtual space (i.e. digital worlds or understanding how and where computers are connected around the world). Almost all knowledge about the world can be classified in terms of space (location, area, distance, or spatial interaction) as well as time. However, despite the fact that time is divided into globally understandable units: seconds, hours, years, and so on, spatial units and related relationships are much more complex, multidimensional (for example, x, y and z) at multiple scales and resolutions, often heterogeneous (even in a single variable representation) and always changing over time. Without a clear understanding of space, any related models, structures, and hypotheses can be misleading (especially those concerning relationships between variables). In particular, the complexity of geographic space creates significant computational and intellectual challenges in the area of distributed access to spatial data, sharing and analysis of government-funded spatial data information infrastructures [21] and the geospatial semantic web [22] (i.e. localization and integration of information without human intervention, including enabling geographic information to be searched within web pages), all of which are part of the spatial CI. However, many of these problems are already well known to those working with spatial data, and a number of non-spatial CI approaches have emerged to address these problems. Spatial CI goes beyond these existing approaches, anchoring solutions in more complex representations of the combination of mathematical and statistical models [23-26] and forging closer collaboration between computer science, information sciences and branch disciplines of geography, geology and geophysics, oceanography, ecology, engineering and environmental sciences, as well as social sciences [5, 8, 27, 28].

This interdisciplinary collaboration enables new knowledge systems that finally lead to the partial realization of the Digital Earth, [29] and is embodied in products such as Google Earth, Microsoft Bing Maps and the National Aeronautics and Space Administration (NASA) WorldWind. The flood of spatial data being collected at an accelerated rate for the foreseeable future from networks of sensors, satellites and even cell phones is still driven

by the vast needs of the aforementioned areas and cannot be well used or well understood unless it is properly managed, analyzed and shared across spatial KI. The dynamic nature of earth systems (such as waves, tides, atmospheric turbulence, and movements in the earth's crust) further complicates our efforts to accurately measure. Massive datasets are also common in the spatial analysis of human systems, including demographic and transportation systems, risk assessment, disease vectors, human mobility, and more. Spatial analysis itself (in a broad sense including spatial modeling) traditionally covers various approaches, including, but not limited to, spatial statistics [30, 31], heuristics and optimization [32, 33], and modeling for spatial problem solving and decision making [34, 35]. These methods have been widely applied in many areas [36-39], but they have been difficult to implement for large and multiscale problems that are computationally intensive and require joint input. This limitation exists despite the advances already made in dealing with the complexity of geographic space discussed earlier. However, spatial CI promises to remove this limitation and thus transform spatial analysis into powerful and accessible computational utilities to enable broad scientific breakthroughs. Spatial CI also proven invaluable in the analysis of errors that propagate due to measurements, and it facilitates the development of better models for error representation, propagation, and control in large distributed computing networks [40]. Linking CI with spatial thinking and geographic approaches offers a promising avenue for solving scientific problems and improving decision-making practices that have a significant impact on society (for example, assessing the impact of global climate change, understanding the complexity of interconnected human and natural systems, maintaining ecosystem services, conserving digital resources in the humanities and social sciences and access to them, as well as the management of transport infrastructure). It does not cover all aspects and current interests of spatial CI because the field of study is growing rapidly. However, this path is representative of current research on the long-term challenges of the complexity of spatial datasets and spatial analysis, and the need and benefits of flexible and secure spatial data exchange. This study highlights some of the discoveries and insights that could be made, and these results could not easily have happened without spatial CI.

2 Materials and research methods

It all starts with technical processing. Spatial foundations governing the interaction of various parameters and phenomena in various physical geographical studies (for example, the Earth's lithosphere, hydrosphere, atmosphere, and global models of flora and fauna) [41]. The main one is the development of architecture and algorithms for distributed processing of geographic information within the framework of spatial CI to improve understanding of ecosystem dynamics and improve prediction of the onset and extent of dust storms in the southwestern United States. As a result of the experiments, scientists were able to predict the onset of dust storms at higher resolutions (3×3 km) over longer time periods (5-10 days).

The evolution of a set of methods and software tools for the integration of multiscale, source and disciplinary oceanographic data began after several recent research expeditions to the Antarctic [40]. The initial goal was to explore several scientific hypotheses about sea ice movement and track iceberg meltwater plumes, but an important parallel effort was to create a near-real-time geospatial basis for decision making. Since they built a spatial CI to support this structure, this led to the development of a sampling design that was optimized to cover smaller areas in relation to the larger scale of the study area. This sampling strategy overcame the limitations of traditional sampling methods used in the past (i.e., using a research ship as a static platform for sampling a single parameter across stations), thereby allowing faster characterization of the ocean surface using multiple data streams

from the sea and in outer space and simultaneously on multiple spatial and temporal scales. Thus, without spatial CI, it is not possible to directly observe and characterize meltwater plumes from individual icebergs and less effectively integrate these individual results with regional and global data. The results provide insight into the impact of iceberg meltwater on the carbon flux from the ocean surface to ocean floor sediments, as well as the role of icebergs in controlling biological productivity in the Weddell Sea. These results also illustrate the importance of spatial CI in general science and highlight key architectural considerations in the development of current and future Earth observation systems, especially as oceanographers and other Earth scientists move into the era of petascale computing.

3 Results and discussion

The aim is to show that spatial CI is not only about using hardware and software or enabling the physical sciences, but also about providing knowledge that serves the needs of the social sciences and humanities, as well as multiple stakeholders and decision makers in groups of citizens from various social, economic and political backgrounds. The creation of a CI is also both a social and a scientific undertaking. Spatial CI, including the Chinese Biographical Database (the largest in the world), the Chinese Historical Geographic Information System (part of China's original Electronic Cultural Atlas Initiative) and the McGill-Harvard-Yenchin Library of Women's Writings Ming Qing database [42]. The research focuses in general on CI for the humanities data, and in particular on spatial CI that aids research into Chinese women writers, their kinship ties, their publishing locations, and their literary and social communities. Critical analysis and recommendations are provided on related issues of conflicting data that researchers do not want to exclude from different data models and geographic scales. This case study shows the value of spatial CI in addressing the difficulties that arise from spatial as well as multilingual, biographical, and temporal ambiguities in these databases, solutions that, again, would not be possible without spatial CI. Big science will continue to need small independent researchers to be the "engine of innovative research", and widespread adoption of CI will allow the two approaches to harmonize. Geographers and sociologists, as well as geoinformatics scientists are actively involved in spatial CI as users, there is a great opportunity to make spatial CI a truly user-centric enterprise [43]. Spatial CI should make room not only for scientists who will use cybernetic tools to collaborate at a distance, but also for educators who will teach with CI. This also applies to scientists who will provide data and information to CI projects on some of the most important scientific issues of the day, such as global climate change.

4 Conclusion

Citizens, along with professional scientists, can now, more and more, participate in ubiquitous cloud computing that uses a wide-ranging architecture to manage the lifecycle of virtual machines and data archives for everything from a personal address book to the largest of multidimensional, multidisciplinary scientific simulation systems. However, while public computing centers are establishing computer networks - cloud storage (Microsoft, Amazon and Google) instead, focus on the delivery of infrastructure as a service, software as a service, and so on. Huge commercial investments in clouds will make them dominant over large-scale computing hardware and software in the next decade [44-47]. Spatial CI is an important subset of the more general CI, covering both computationally intensive and interdisciplinary usage requirements such as service hosting, virtual computing environments, and virtual datasets. The specific requirements of spatial

CI are well suited to many common cloud capabilities, thus warranting further fundamental and empirical exploration. Indeed, the concept of space, which is special in CI, introduces several interesting research problems for physicists and social scientists alike. Many geographic applications are interdisciplinary and involve multiple stakeholders and decision makers who have diverse social, economic, and political conditions, thus making collaboration very important but challenging.

Spatial CI will follow the path of GIS and eventually become as transparent as GIS is becoming in the world of Google Maps and Google Earth. As the size of spatial datasets and the complexity of spatial analysis and modeling continue to increase, and the need for virtual collaboration in scientific research becomes in demand, scalable spatial CI becomes even more important and timely. The virtues of spatial CI stem from the complexity of the problems, the danger of errors that can proliferate, the deep need to develop solutions that will benefit many societal domains, the ongoing vision of access to a complete Digital Earth and the next generation of GIS - CyberGIS - with integral high-performance, distributed, and joint opportunities.

5 Acknowledgements

This study was supported by the Russian Federation State Task № FNNN-2021-0005.

References

1. Atkins D.E., Kelvin K. Droegemeier, Stuart I. Feldman, Hector Garcia-Molina, Michael L. Klein, David G. Messerschmitt, Paul Messina, Jeremiah P. Ostriker, Margaret H. Wright, Revolutionizing science and engineering through cyber-infrastructure: Report of the national science foundation Blue-Ribbon advisory panel on cyberinfrastructure, National science foundation publication NSF0728 (National science foundation, Washington, DC), 2003.
2. Crawford D., Leland Jameson, Margaret, Leinen, José Muñoz, Stephen Meacham, Michael Plesniak, Cyberinfrastructure vision for 21st century discovery, National science foundation publication CISE051203 (National science foundation, Washington, DC), 2007.
3. Renear A.H., Palmer C.L., Strategic reading, ontologies, and the future of scientific publishing, 2009, *Science* 325:828–832, DOI: 10.1126/science.1157784.
4. Buetow K.H., Cyberinfrastructure: empowering a “third way” in biomedical research, 2005, *Science* 308:821–824, DOI: 10.1126/science.1112120.
5. Hey T., Tansley S., Tolle K., The fourth paradigm: Data-intensive scientific discovery (Microsoft research, Redmond, WA), 2009.
6. Kartsan I.N., Efremova S.V., Khrapunova V.V., Tolstopiatov M.I. Choice of optimal multiversion software for a small satellite ground-based control and command complex. *IOP Conference Series: Materials Science and Engineering*. 2018. V. 450(2). 022015. DOI:10.1088/1757-899X/450/2/022015.
7. Kartsan I.N., Efremova S.V. Distributed control of university small spacecraft. Actual problems of aviation and cosmonautics. *Proceedings of the V International Scientific-Practical Conference dedicated to the Day of Cosmonautics*. Krasnoyarsk. 2019. P. 47 - 48. (in Russian).
8. Benioff M.R., Edward D. Lazowska, Computational science: ensuring america’s competitiveness, report to the President by the President’s information technology

- advisory committee (National coordination office for information technology research and development, Arlington, VA), 2005.
9. Wang S., Zhu X-G., Coupling cyberinfrastructure and geographic information systems to empower ecological and environmental research, 2008, *Bioscience* 58: 94–95.
 10. Hey T., Trefethen A.E., Cyberinfrastructure for e-Science, 2005, *Science* 308:817–821, DOI: 10.1126/science.1110410.
 11. Foster I., Service-oriented science, 2005, *Science* 308: 814–817, DOI: 10.1126/science.1110411.
 12. Bohannon J., Distributed computing. Grassroots supercomputing, 2005, *Science* 308:810–813, DOI: 10.1126/science.308.5723.810.
 13. Clery D., Voss D., All for one and one for all, 2005, *Science* 308:809, DOI:10.1126.
 14. Lazowska E.D., Patterson D.A., An endless frontier postponed, 2005, *Science* 308:757, DOI:10.1126/science.1113963.
 15. Craglia M., Michael Goodchild, et.al., Ed Parsons., Next-generation digital earth: A position paper from the Vespucci Initiative for the advancement of geographic information science. *Int J Spatial Data Infrastructures Res* 3:146–167, 2008.
 16. Blais J.A.R., Esche H., Geomatics and the new cyberinfrastructure. *Geomatica* 62:11–22, 2008.
 17. Wang S., Liu Y., TeraGrid GIScience gateway: bridging cyberinfrastructure and GIScience. *Int J Geogr Inf Sci* 23:631–656, 2009, DOI:10.1080/13658810902754977.
 18. Zhang T., Tsou M-H., Developing a grid-enabled spatial web portal for internet GIServices and geospatial cyberinfrastructure. *Int J Geogr Inf Sci* 23: 605–630, 2009, DOI:10.1080/13658810802698571.
 19. Maguire D.J., Goodchild M.F., Rhind D.W., eds., *Geographical information systems: principles and applications* (Wiley, New York), 1991.
 20. Gewin V., Mapping opportunities. *Nature* 427: 376–377, 2004.
 21. Taaffe E.J., The spatial view in context. *Ann Assoc Am Geogr* 64:1–16, 1974, DOI:10.1111/J.1467-8306.1974.TB00950.X.
 22. Tyapkin V.N., Kartsan I.N., Dmitriev D.D., Efremova S.V. Algorithms for adaptive processing of signals in a flat phased antenna array. *International Siberian Conference on Control and Communications (SIBCON)*. 2017. Proceedings 2017. 7998452. DOI:10.1109/SIBCON.2017.7998452.
 23. Miller H.J., Tobler's First Law and spatial analysis. *Ann Assoc Am Geogr* 94:284–289, 2004, DOI:10.1111/j.1467-8306.2004.09402005.x.
 24. Onsrud H., ed., *Research and theory in advanced spatial data infrastructure concepts* (ESRI Press, Redlands, CA), 2007.
 25. Kartsan I.N. Ground control system for small spacecraft // *Bulletin of Siberian State Aerospace University named after Academician M.F. Reshetnev*. 2009. V. 3 (24). P. 89 - 92. (in Russian).
 26. Kartsan I.N., Goncharov A.E., Zelenkov P.V., Kovalev I.V., Fateev Y.L., Tyapkin V.N., Dmitriev D.D. Applying filtering for determining the angular orientation of spinning objects during interference. *IOP Conference Series: Materials Science and Engineering*. 2016. 012020. DOI:10.1088/1757-899X/155/1/012020.
 27. Egenhofer M., Toward the geospatial semantic web. *Advances in geographic information systems international symposium*, eds Makki Y., Pissinou N. (Association for Computing Machinery, McLean, VA), 2002, pp 1–4, DOI:10.1145/585147.585148.

28. Anselin L., Florax R., Rey S., eds., *Advances in spatial econometrics: methodology, tools and applications* (Springer, Berlin), 2004, DOI:10.1007/978-3-662-05617-2.
29. Wang S., Armstrong M., A theoretical approach to the use of cyberinfrastructure in geographical analysis, *Int J Geogr Inf Sci* 23:169–193, 2009, DOI:10.1080/13658810801918509.
30. Wang S., A cyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis, *Ann Assoc Am Geogr* 100:535–557, 2010, DOI:10.1080/00045601003791243.
31. Penninga F., Van Oosterom P.J.M., A simplicial complex-based DBMS approach to 3D topographic data modelling. *Int J Geogr Inf Sci* 22:751–779, 2008, DOI:10.1080/13658810701673535.
32. Baker K.S., Chandler C.L., Enabling long-term oceanographic research: changing data practices, information management strategies and informatics. *Deep-Sea Res II* 55(18–19):2132–2142, 2008, DOI:10.1016/j.dsr2.2008.05.009.
33. Yang C., Raskin R., Goodchild M., Gahegan M., Geospatial cyberinfrastructure: Past, present and future. *Comput Environ Urban Syst* 34:264–277, 2010, DOI:10.1016/j.compenvurbsys.2010.04.001.
34. Gore A., The digital Earth: understanding our planet in the 21st Century. *Photogramm Eng Remote Sensing* 65:528, 1999, DOI:10.1080/00050348.1998.10558728.
35. Averyanov V.S., Kartsan I.N., Efremova S.V. Methods of automated detection of anomalies and nonlinear transitions by autonomous unmanned aerial vehicles. *Journal of Physics: Conference Series*. 2021. 42001. DOI: 10.1088/1742-6596/1889/4/042001
36. Pirkhavka A.P., Legkiy N.M., Kartsan I.N., Zhukov A.O. Analysis of the operational readiness of telecommunication networks. *Journal of Physics: Conference Series*. 2021. 32053. DOI: 10.1088/1742-6596/2094/3/032053
37. Anselin L., What is special about spatial data? Alternative perspectives on spatial data analysis, technical report 89-4 (National center for geographic information and analysis, Santa Barbara, CA), 1989.
38. Anselin L., Local indicators of spatial association-LISA. *Geogr Anal* 27:93–115, 1995.
39. Kartsan I.N., Shumakova N.A., Zhukov A.O., Platonov O.A., Efremova S.V. Safety system for space vehicles using integrated devices located outside of the satellite platform. *JOP Conference Series: Metrological Support of Innovative Technologies*. 2020. 42022. DOI: 10.1088/1742-6596/1515/4/042022
40. Turton I., Openshaw S., High-performance computing and geography: developments, issues, and case studies. *Environ Plan A* 30:1839–1856, 1998, DOI:10.1068/a301839.
41. Krzanowski R., Raper J., *Spatial evolutionary modelling* (Oxford University Press, Oxford), 2001.
42. Batty M., *Cities and complexity: understanding cities through cellular automata, agent-Based models, and fractals* (MIT Press, Cambridge, MA), 2005.
43. Jankowski P., Nyerges T., GIS-supported collaborative decision making: Results of an experiment. *Ann Assoc Am Geogr* 91:48–70, 2001, DOI:10.1111/0004-5608.00233.
44. Zhukov A.O., et.al., Adaptation of receiving channels to the spectrum of the received signal. *JOP Conference Series: Metrological Support of Innovative Technologies*. 2020. 032070. DOI: 10.1088/1742-6596/1515/3/032070
45. Goodchild M.F., Janelle D.G., eds., *Spatially integrated social science* (Oxford University Press, Oxford), 2004.

46. Miller H.J., Wentz E.A., Representation and spatial analysis in geographic information systems. *Ann Assoc Am Geogr* 93:574–594, 2003, DOI:10.1111/1467-8306.9303004.
47. Wright D.J., Spatial data infrastructures for coastal environments. Remote sensing and geospatial technologies for coastal ecosystem assessment and management. Lecture notes in geoinformation and cartography, ed Yang X. (Springer, Berlin), pp 91–112, 2009, DOI: 10.1007/978-3-540-88183-4_5.
48. Tesfatsion L., Agent-based computational economics: Growing economies from the bottom up. *Artif Life* 8:55–82, 2002, DOI:10.1162/106454602753694765.
49. Helly J.J., Kaufman R.S., Vernet M., Stephenson G.R., Spatial characterization of the meltwater field from icebergs in the Weddell Sea. *Proc Natl Acad Sci USA* 108:5492–5497, 2011, DOI:10.1073/pnas.0909306108.
50. Dembickij N.L., et.al., Distributed control network for complex radio engineering systems on continuum computers. *IOP Conference Series: Metrological Support of Innovative Technologies*. 2020. 42076. DOI: 10.1088/1742-6596/1515/4/042076
51. Yang C., Wu H., Huang Q., Li Z., Li J., Using spatial principles to optimize distributed computing for enabling the physical science discoveries. *Proc Natl Acad Sci USA* 108:5498–5503, 2011, DOI:10.1073/pnas.090931510.
52. Efremova S.V., Kartsan I.N., Zhukov A.O. An ordered ranking multi-attributive model for decision-making systems with attributes of control systems software. *IOP Conference Series: Materials Science and Engineering*. 2021. 12068. DOI: 10.1088/1757-899X/1047/1/012068
53. Kulikov A., Timoshenko A., Zhukov A., Kartsan I. Control system for multi-agent groups of heterogeneous sensors. *CEUR Workshop Proceedings*. 2021. V. 3091. P. 87. DOI: 10.47813/dnit-mip5/2022-3091-87-92
54. Semenenko M.G., Kniazeva I.V., Beckel L.S., Rutskiy V.N., Tsarev R.Yu., Yamskikh T.N., Kartsan I.N. How to use neural network and web technologies in modeling complex technical systems. *IOP Conference Series: Materials Science and Engineering*. 2019. 32095. DOI: 10.1088/1757-899X/537/3/032095
55. Kudymov V.I., et.al., Formation of optimal composition of the modules of single-function multiversion software for automated control system of the satellite communication system. *IOP Conference Series: Materials Science and Engineering*. 2018. 052009. DOI: 10.1088/1757-899X/450/5/052009
56. Sieber R.E., Wellen C.C., Jin Y., Spatial cyberinfrastructures, ontologies, and the humanities. *Proc Natl Acad Sci USA* 108:5504–5509, 2011, DOI:10.1073/pnas.0911052108.
57. Poore B.S., Users as essential contributors to spatial cyberinfrastructures. *Proc Natl Acad Sci USA* 108:5510–5515, 2011, DOI:10.1073/pnas.0907677108.
58. Fox A., Computer science. Cloud computing—what’s in it for me as a scientist? *Science* 331:406–407, 2011, DOI:10.1126/science.1198981.
59. Gao X., Ma Y., Pierce M., Lowe M., Fox G., Building a distributed block storage system for cloud infrastructure. *Proceedings of the 2nd IEEE international conference on cloud computing technology and science (CloudCom)*, eds Fox G., Zhao G., Qui J., Hughes A. (Institute of Electrical and Electronics Engineers, Indianapolis, IN), 2010, pp 1–9, DOI:10.1109/CloudCom.2010.72.
60. Preparata F.P., Shamos M.I., Computational geometry: an introduction (Springer, Berlin), 1990.