Journal of Global Change Data & Discovery. 2019, 3(4): 311-316
DOI:10.3974/geodp.2019.04.01
www.geodoi.ac.cn

Global Change Research Data Publishing & Repository

Geography of China in the Big Data Era

Liao, X. H.

Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Abstract: This article discusses the state of geography in China in the big data era along with the importance of geographic spatiotemporal big data in geography research. Remote sensing satellite data are the largest source of geographic big data. Since 2012, the National Remote Sensing Centre of China has used Chinese and international remote sensing satellite data to analyze the global ecological environment, issued a series of annual reports that made assessment on some global environmental problems, and helped to expand geographical research in China from the regional to the global scale. In 2018, the State Council of China proposed the Scientific Data Management Measures, which contributes to the effective management of geographic big data and promote data sharing. In the same year, the Big Geographical Data Working Committee of the Geographical Society of China was established. This committee has played an active role in addressing national needs and fostering talent in geography and geographic big data. The application of big data in contemporary geographical research will strengthen its ability to solve problems and make the research results more applicable in practice. Geographic big data also allow various theories, methods, and models in geography to be tested and refined. Thus, geographic big data are expected to play an important role in solving resource-related and environmental problems resulting from social progress and economic development.

Keywords: geographic big data; geography of China; remote sensing; global observation; ecological environment

1 Introduction

The processes of informatization, digitization, and automation have generated huge and diverse data resources on different scales throughout the world. Big data thus present both significant challenges and opportunities[1]. As of 2003, 5 extrabytes (EB) (1EB=10^9 GB) of data have been created, and the amount of generated data has exploded in the past 10 years. Globally, the amount of data generated is expected to reach nearly 40 zettabytes (ZB) (1 ZB = 10^3 EB) by 2020[2]. This trend in data growth has exceeded most previous predictions and is profoundly affecting scientific research.

Nature and *Science* have published special journal issues to discuss the opportunities and challenges related to big data research[3–4]. In March 2012, the United States government announced the launch of the Big Data Research and Development Program[5]. In May 2012, the United Nations issued a white paper entitled "Big data development, opportunities and challenges"[1]. In December 2011, China's capital markets released their first big data-

Received: 23-09-2019; **Accepted:** 10-12-2019; **Published:** 24-12-2019
Author Information: Liao, X. H. S-3620-2018, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, liaoxh@igsnrr.ac.cn
Article Citation: Liao, X. H. Geography of China in the big data era [J]. *Journal of Global Change Data* & *Discovery*, 2019, 3(4): 311−316. DOI: 10.3974/geodp.2019.04.01.

themed report entitled "The Big Data Era is Coming"[6]. In August 2015, the State Council of China issued a notice entitled "Promoting the Action for the Development of Big Data"[7]. On December 8, 2017, President Xi, J. P. stressed the implementation of the national big data strategy and accelerated the digitization of China during the Second Collective Study Conference[8]. In March 2018, the State Council of China proposed the Scientific Data Management Measures (Guobanfa [2018]17) [9]. In May 2018, President Xi, J. P. proposed the full implementation of the national big data strategy in the congratulatory letter to the China International Big Data Expo 2018[10].

Big data are proliferating in all walks of life, and the big data era has arrived. The size of a country's data and the country's ability to use the data have become important indicators of the country's national strength. The possession and control of the data are also expected to become critical both for nations and enterprises. The big data era has spread to all areas, including geography and its various subfields.

2 Characteristics of Geographic Big Data

Geography mainly concerns the spatial structures and temporal processes of natural and human elements along with the internal mechanisms and external representations of the interactions among them. Geographic science cannot be separated from the study of spatial structures, temporal processes, surface characteristics, and extreme environments. From a data perspective, scientific studies are critical to the spatial, temporal, and spectral resolutions of the data along with data availability. More detailed and accurate data will result in a higher level of cognition and a better understanding of reality. Therefore, geography has welcomed big data from their inception.

Geographic data are typical big data with the following characteristics[11–13]: (1) Large volume. The size of remote sensing data, wireless sensor network data, volunteered geographic information (VGI) data, and other geographic data are on the TB to petabyte (PB) scale or even the extrabyte (EB) scale; (2) Fast updates. An integrated network of air, space, and ground sensors means that network data are generated quickly, and VGI data can be generated and transmitted at any time and place; (3) Multimodal. Geographic big data include structured remote sensing data, ground observation network data, and semi-structured or unstructured VGI data; (4) Uncertain accuracy. A large amount of contaminated data exists, and heavy data cleaning must be conducted before using the data; (5) High value. Geographic big data can be widely used in agricultural services, urban management, resource surveys, environmental protection, disaster prevention, etc.

Eighty percent of the world's information is related to geographic location[14]. For example, data related to resource (e.g., land, minerals, and the environment) management, urban planning, transportation, water conservation, agriculture, forestry, environmental protection, emergency decision making, and so on can be generated via spatial visualization, spatial inquiry, thematic mapping, spatial analysis, and other spatial methods.

Geographic spatiotemporal data are an important component of big data. These data can be generated by various means, including remote sensing satellites (e.g., land, atmosphere, ocean, and sea ice series satellites). The number of remote sensing satellites worldwide exceeds 1,000[15]; archived data from remote sensing satellites have reached the EB level, making these data the largest source of geographic big data. Some regular observation data such as meteorological and hydrological data also exist in large quantities. There are currently thousands of observation networks around the world, including land-based meteorological networks, hydrological networks, sea-based buoy networks, and deep-sea exploration networks. VGI-based big data include mobile global position system, tagging, image, and other structured data on professional platforms along with unstructured data containing spatiotemporal information produced by social networks. The number of global internet users has exceeded four billion[16], which has resulted in a sharp increase in VGI-based big data. Other big data related to earth systems, the atmosphere, land surface, sea ice, and so on are also available. The CMIP6 project estimated that the quantity of geoscience-related big data is approximately 20–40 PB[17].

The inclusion of big data in the field of geography has produced numerous advances, including the following: (1) The combination of traditional research with in-depth data mining allows complex associations to be resolved; (2) Geographic big data are extensible, and the spatiotemporal information embedded in big data can be applied in spatial analyses through big data mining; (3) New techniques and methods that have arisen from the emergence of big data can be widely applied in geography research; (4) Transmission networking, data clouds, computing clouds have become ubiquitous; (5) The fourth scientific research paradigm (i.e., data-driven scientific discovery) is widely applied.

As an example, the most recent exploration survey of land and resources (arable land, permanent basic farmland, and woodland) in China employed several data-related methods that were not used in previous surveys[18–20]. For instance, the latest survey considered professional and social big data related to land use, basic geography, and management operations along with big data obtained by remote sensing satellites and unmanned aerial vehicles. It also employed methods such as supercomputing, deep learning, and cloud computing. Another example is traditional sampling analysis; it is difficult to study the spatial characteristics of urban populations and analyze migration without the help of big data and the associated techniques. This work can be carried out rapidly based on data available from mobile phone signals.

Geographic information system (GIS) data are essential for the application of big data in geography. GIS is a computer system for collecting, storing, managing, analyzing, displaying, and applying geographic information. GIS is a general technology for analyzing and processing massive geographic data in the real world (resources and environment)[21]. GIS has undergone several development stages, including the initial development of GIS, which occurred in the 1970s and focused on the development and management of map data, the statistics of spatial data, and map production. In the 1980s, GIS entered a stage of consolidation, which included the development of mixed-data models, comprehensive spatial analysis, and the development of professional software modules and environmental resource applications. These two eras are collectively referred to as the development stage of GIS[22–23], which is also known as the mainframe era. The 1990s were characterized by the development of geographic information science and were referred to as the PC era; in this stage, large-scale databases and networks were developed to support, geographic ontology, system modeling, and web GIS. In the 21st century, during which the internet era began, GIS entered the public service stage. This stage was characterized by the development of location service applications, computing services, grid GIS, and virtual environment applications. In the past 10 years, GIS has entered the social services stage of development, also known as the development stage of the geographic information world; this stage is characterized by automatic processing, massive storage, efficient computing, and knowledge-based services[24].

Remote sensing satellite data account for the largest proportion of geographic big data and have allowed Chinese geographical research to expand from a national scale to a global one. In previous scientific studies, developed countries mainly used primary observation data to study global problems. Now, relying on remote-sensing satellites at home and abroad, Chinese geoscientists have been able to do research on transnational and global problems. China has developed one of the most complete and expansive networks of remote sensing satellites in the world, rivaling that of the United States[25]. Chinese scientists are now able to use the basic data from foreign remote sensing satellites along with their own algorithms to produce data products and analytics with international influence. Chinese remote sensing satellites can also be used to obtain data from all over the world. For example, starting in 2012, the National Remote Sensing Centre of China has organized efforts to analyze the global ecological environment using Chinese and foreign remote sensing satellite data, resulting in a series of annual reports[26] covering topics such as ecology, vegetation, land use, agriculture, wetlands, and urbanization. These reports have had a significant impact on the activities of international scientific organizations such as the Earth Observation Organization.

3 Geographic Big Data Enhance China's Ability to Solve Practical Problems

With societal development and increasing economic growth, problems related to resources and the environment are becoming more prominent and complex. Traditional simplified models often cannot solve these complex practical problems, and solutions to these problems need to be developed and refined based on large sample sizes. In addition, many geographical theories, methods, and models require accurate boundary constraints and field data as inputs. The history of geography shows that obtaining reliable, multifactorial analysis results is related to the availability of high-quality input data. The incorporation of big data into contemporary geography will increase the ability to solve real problems[27] and ensure that the prediction results of models are consistent with reality (Figure 1).

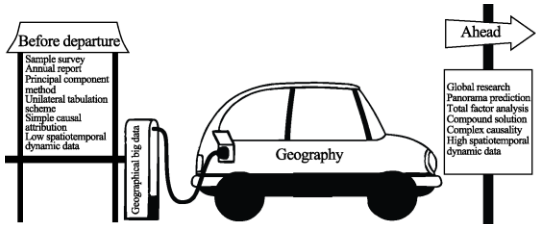

Figure 1 The injection of big data into geography promotes the further development of academic theories and methods

The advances associated with utilizing geographic big data are attributed to several factors, including the following. First, big data allow the scope of geography research to range from local sampling to global coverage (e.g., using remote sensing satellite data to estimate global vegetation biomass or crop yield). This type of research has been widely applied in the areas of tourism, natural resources, agriculture, and so on. Second, by incorporating big data, geography research can be extended from understanding the current conditions to generating high-precision historical reconstructions and highly reliable predictions of future scenarios. This type of research has been widely applied to study regional development, global change, and earth processes based on numerical models. Third, big data mining can be applied to high-resolution spatiotemporal data fitting and the analysis of spatiotemporal associations, leading to an understanding of complex relationships involving multiple factors. Finally, the large sample sets made possible by big data can be used to test and improve geographical theories, models, and methods that were formulated based on relatively small samples in the past. In this way, big data can improve the reliability of scientific research and help researchers develop solutions that are more appropriate to real-world problems.

4 Considerations Related to Geographic Big Data Application Management

The meaningful application of geographic big data cannot be separated from orderly data management, the protection of intellectual property, data security, and quality control. In August 2018, the Big Geographical Data Working Committee of the Geographical Society of China was established[28]. The demand-oriented approach proposed by the working committee is aimed at addressing the most important bottlenecks in the development of geographic big data in China, which include:

(1) Data management and data sharing

Guided by the State Council of China's Scientific Data Management Measures, it is expected to take five years of effort to provide a basic solution to the bottleneck in geographical data sharing. The national big data strategy to address this bottleneck involves setting up a platform for data publishing and sharing, constructing a national scientific data center, and encouraging society members to protect intellectual property rights.

(2) Intellectual property protection and scientific evaluation

The establishment of standards, norms, and methods for data authentication and intellectual property rights will promote the data sharing and scientific achievements, pilot and demonstration work in a period of one to two years, and popularization throughout the country in the next five years.

(3) Deficiency in global-scale research in China

The Geographical Society of China will hold a global research and development conference and will become actively involved in any relevant national science and technology activities (e.g., the remote sensing monitoring of the global ecological environment), which will help enable the publication of representative global datasets within five years. This objective is in line with the national development strategy and the United Nations sustainable development goals.

(4) Deficiency in academic papers based on scientific data

The Big Geographical Data Working Committee and the Academic Editorial Committee will work closely together to promote the publication of academic papers and original scientific data. The proportion of academic papers and data published in academic journals sponsored by the Geographical Society of China (or co-sponsored) is currrently less than 1%. However, this is expected to increase to more than 30% over the next five years as a result of the Committee's efforts.

(5) Construction of a data computing environment

The Geographical Society of China will help realize the value of geographic big data by (a) selecting and promoting practical examples of valuable computing environments for geographic big data and (b) promoting the use of big data in scientific discovery and sustainable development.

(6) Scientific application of geographic big data to sustainable development

Making big data play a positive role in the sustainable development of society is one of the important tasks of the Big Geographical Data Working Committee. The Geographical Society of China intends to explore how geographic big data can be applied to promote national and local sustainable development and ensure national and local ecological security.

(7) Fostering talent in geographic big data

The Big Geographical Data Working Committee will continue to promote the Capacity Building in 100 Universities Program on Global Change Research Data Publishing & Sharing along with the development of related textbooks and university courses with the goals of (a) introducing big data in 100 colleges and universities by 2025 and (b) making courses in geographic big data available in more than 10 universities.

(8) Data security and scientific ethics

Data security and scientific ethics are key issues that the Geographical Society of China must emphasize. The Big Geographical Data Working Committee will prioritize data security and scientific ethics in geographic big data and work to create associated standards and guidelines.

5 Conclusion

With the advent of geographic big data, the Geographical Society of China welcomes new opportunities for development. Geographical theories, methods, and application practices at the regional and global scales will be continuously tested and refined using the large sample sets provided by big data. With the development of global-scale geographic big data, particularly the rapid proliferation of earth observation satellite data in China, important research results are being produced. Thus, geographic big data will play an important role in social progress and economic development, both globally and within China, and help solve pressing challenges related to resources and the environment.

Acknowledgement

This article is based on the report of the annual conference of the Big Data Working Committee of the Geographical Society of China on September 21, 2019. Ma, J. H. and Deng, X. M. contributed to the revision and improvement of this article.

References

[1] UN Global Pulse. Big data for development: challenges & opportunities [R/OL]. [2012-10-02]. http: //www.unglobalpulse.org/projects/BigDataforDevelopment.
[2] https://www.idc.com/.
[3] The big challenges of big data [OL]. https://www.nature.com/.
[4] Special online collection: dealing with data [OL]. https://www.sciencemag.org/site/special/data.
[5] The federal big data research and development strategic plan [OL]. https://www.nitrd.gov.
[6] https://blog.sina.com.cn/zhaogd.
[7] State Council of China. Platform for action to promote the development of big data (Guofa [2015] NO.50) [Z]. 2015-8-31.
[8] http://www.xinhuanet.com/politics/leaders/2017-12/09/c_1122084706.htm.
[9] State Council. Measures for the Administration of Scientific Data (Guobanfa [2018] NO.17) [Z]. 2018-03-17.
[10] Xi, J. P. Congratulatory letter to China International Big Data Industry Expo 2018 [R]. People's Daily, 2018-05-27 (01).
[11] Guo, H. D. Big data, big science, big discovery—a summary of the International Symposium on Big Data and Scientific Discovery [J]. *Proceedings of the Chinese Academy of Sciences*, 2014, 299(4): 500–506.
[12] Guo, H. D. Earth big data science engineering [J]. *Proceedings of the Chinese Academy of Sciences*, 2018, 33(8): 818–824.
[13] Han, P., Zhang, J. W., Wang, Y. W. Application of geographic information and location big data in map compilation [J]. *Gansu Science and Technology*, 2016, 32(1): 334–336.
[14] Cao, S. B. Overview of the remote sensing satellite market in 2017 (Part 2) [J]. *China Aerospace*, 2018(6): 73–78.
[15] https://new.qq.com/omn/20190128/20190128B00CCS.html.
[16] http://www.xinhuanet.com/expo/zt/sjdlxxdh/index.htm. 2018.
[17] https://esgf.llnl.gov/search/cmip6.
[18] Zheng, J. Y. Analysis of technical exploration based on the third national land survey [J]. *China Resources Comprehensive Utilization*, 2019, 37(9): 64–66.
[19] Zhu, W., Wang, J. C., Yuan, R. C. Talking about the application of intelligent mobile terminals in the third national land survey [J]. *Science* & *Technology and Information*, 2017(9): 155.
[20] Department of Natural Resources Investigation and Monitoring, Guangxi Natural Resources Department. How innovative application? — a picture to understand the new changes in the third national land survey [J]. *Southern Land Resources*, 2019(6): 18–19.
[21] Chen, S. P. Introduction to Geographic Information System [M]. Beijing: Science Press, 2003.
[22] Zhou, C. H. New era of geographic information system: grid geographic information system [J]. *Geographic Information World*, 2007(4): 17.
[23] Zhang, H. Y., Wang, Q. M., Zhou, C. H., *et al*. "Digital earth" and geographic information science [J]. *Geoinformatics*, 2001(4): 1–4.
[24] Zhang, S. L. The improvement of geographic information service capabilities effectively promotes social progress [J]. *Surveying and Mapping Technology* & *Equipment*, 2015, 17(1): 53–54.
[25] http://www.broadcast.hc360.com.
[26] http://www.nrscc.gov.cn/.
[27] Wang, P. Big data plugged in machine learning wings to provide new fuel for 5G [J]. *Communications World*, 2019(11): 47–48.
[28] Geographical Society of China. The Big Geographical Data Working Committee of Geographical Society of China (GSC_BigData) established [R]. *Journal of Global Change Data* & *Discovery*, 2018, 2(3): 354–356. DOI: 10.3974/geodp.2018.03.18.