Scientific big data and Digital Earth

Article *in* Chinese Journal · December 2014

DOI: 10.1360/972013-1054

Dong Liang
Chinese Academy of Sciences

Chin. Sci. Bull. (2014) 59(35):5066–5073
DOI 10.1007/s11434-014-0645-3

csb.scichina.com
www.springer.com/scp

Review

Geography

Scientific big data and Digital Earth

Huadong Guo · Lizhe Wang · Fang Chen · Dong Liang

Received: 15 May 2014 / Accepted: 30 July 2014 / Published online: 1 October 2014

Abstract Big data has been a focus of research in science, technology, economics, and social studies. Many countries have already incorporated big data research into their national strategies. This paper elaborates upon the origin, connotation, and development of big data from both a spatial and temporal perspective. It proposes that scientific big data will become a new solution in scientific research as the paradigm changes from being model-driven to data-driven. This paper defines the concept of "scientific big data" and proposes strategies for solving "big data problems". Theoretical frameworks and data systems for Digital Earth are discussed with a clear conclusion that scientific big data is a prominent feature of Digital Earth. As an example, spatial cognition of the formation mechanism of China's Heihe-Tengchong Line—a geo-demographic demarcation line dividing China into two parts—is discussed within the context of big data computation and analysis for Digital Earth.

Keywords Big data · Scientific big data · Data-intensive science · Earth science · Digital Earth

Electronic supplementary material The online version of this article (doi:10.1007/s11434-014-0645-3) contains supplementary material, which is available to authorized users.

H. Guo (✉) · L. Wang · F. Chen · D. Liang
Key Laboratory of Digital Earth Sciences, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
e-mail: hdguo@radi.ac.cn

1 Introduction

Since the 1870s, new technologies and inventions have emerged in an endless stream. Since the beginning of the Second Industrial Revolution, carriers of data have practically doubled every decade. From the industrial age to the information age, a revolution in information technologies played a decisive and unprecedented role in the development of society, science, technologies, and economics. Data volume is continuously increasing, doubling every 3 years. In the last 10 years, the rapid development of computer technologies and the Internet has prompted the emergence of semi-structured and unstructured data, such as audio, video, text, and images. Social networking services, the Internet of Things, and cloud computing have been widely used, with data volume and categories experiencing rapid growth. The big data era has quietly arrived [1–4]. According to an article by the International Data Corporation (IDC) [5], all the digital data created, replicated, and consumed will double every 2 years until 2020 (Fig. 1). There were about 1.8 zettabytes (ZB) of data created, replicated, and consumed worldwide in 2011. The IDC estimates that by 2020, this figure will reach 40 ZB, with China's global share increasing from 13 % in 2012 to 21 %.

The modes of data production have gone through phases of passive system operations and active user-generated content, and have now entered the phase of an automatic perceptual system [6]. This leap is a critical factor for emerging big data. Like other research innovations, big data is also moving from the conception stage to small-scale technical practices, and eventually will become an emerging research direction. A brief description of the development process of big data is shown in Fig. 2. *Nature* took the lead in publishing a special issue on "Big Data" in

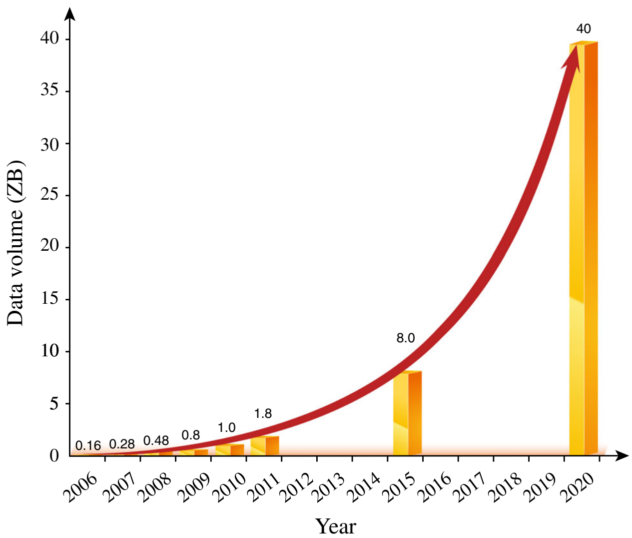

Fig. 1 Global growth trend of data volume, 2006–2020 (based on "The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east")

September 2008 [7]. It indicated that the influence of big data had reached the fields of natural science, social science, humanities, and engineering. *The Fourth Paradigm*: *Data-Intensive Scientific Discovery* [8], published by Microsoft Research in October 2009, was closely associated with big data, noting that the data-intensive scientific discovery paradigm has been established and widely acknowledged. "Data, data everywhere," published in *The Economist* in February 2010, expanded and advanced the concept of big data [9]. In February 2011, *Science* published a special issue titled "Dealing with data" [10], and the McKinsey Global Institute (MGI) released "Big data: The next frontier for innovation, competition, and productivity" [11], which claimed that big data has become a hot topic in the field of social science research. In May 2012, United Nations Global Pulse published a white paper, "Big Data for Development: Opportunities & Challenges" [12], suggesting that projects/programs of big data research promote a national strategy. In June 2012, the Gartner Group proposed the "4V" definition for big data in "The importance of 'Big Data': A definition" [13], which advocated understanding the conceptual basis of big data. "Next-generation Digital Earth" [14], published in *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, showed that humankind had entered the era of big data, and that big data would play a key role in the next generation of Digital Earth. "The Big Data for Digital Earth and Future Earth" special session was held at the 35th International Symposium on Remote Sensing of Environment in April 2013. The International Workshop on Big Data for International Scientific Programmes: Challenges and Opportunities, convened by the International Council for Science (ICSU) Committee on Data for Science and Technology (CODATA) and co-sponsored by six international organizations and the Chinese Academy of Sciences Institute of Remote Sensing and Digital Earth, took place in June 2014. The workshop aimed to shed more light on the potential role of big data in international and interdisciplinary research activities. It was noted that big data has received increasing attention and acceptance in the fields of space and Earth science.

Data are not only a resource, but also a kind of wealth. Vast and complex data will drive social, scientific, technological, and economic development. Many countries and international organizations have already promoted big data research in their national and international strategies. This is bound to bring profound changes in the future. Figure 3 shows the level of attention paid to big data in different countries. Big data research and development programs have been launched at the national or intergovernmental level by the USA, the European Union, Australia, Japan, Korea, and others. The Executive Office of the President of the USA released a report in May 2014 on big data and privacy, "Big Data: Seizing Opportunities, Preserving Values," along with an array of supporting documents, such as "Big Data and Privacy: A Technological Perspective." The reports address the conflict between big data technology and privacy, and propose six recommendations to improve and maximize the benefits of big data while minimizing its risks [15, 16]. Big data has become a manifestation of informational sovereignty; it will be the next topic of international debates and will play a significant role in border, coastal, and air defenses. Big data is unleashing interesting times of transition. It will change human life and our understanding of the world [17–19].

2 Scientific big data and its connotation

Academics in China and abroad are trying to analyze and understand the concept of big data. The current definition of big data comes from the following two different perspectives: The relative characteristic denotes those datasets that cannot be acquired, managed, or processed on common devices within an acceptable time [20], while the absolute characteristic defines big data through "4V," i.e., Volume, Variety, Veracity, and Velocity [13].

Big data research is different from traditional logical research. It uses analytical induction on a vast amount of data to statistically search, compare, cluster, and classify. It involves correlation analysis and implies that there could be certain regularity between two or more variables' values, and it aims at uncovering hidden correlated networks in datasets [21].

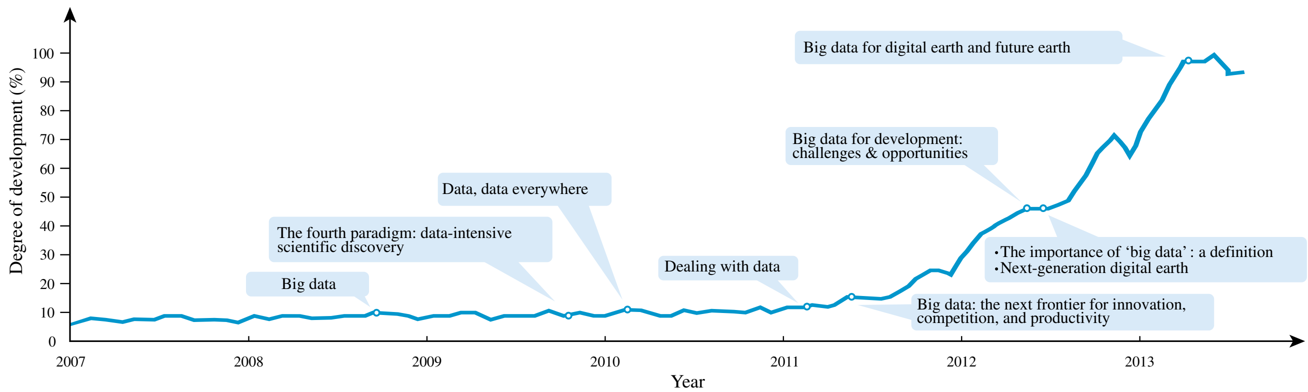

Fig. 2 Development of big data (based on Google Trends; acquired in August 2013)

Thus, it can be seen that the substantive characteristics of big-data computing comprise a paradigm shift from model-driven science to data-driven science, as well as the establishment of a data-intensive scientific approach. The scientific research methodologies have employed observation-based science from the very beginning, including the experimental science that began thousands of years ago, the theoretical science that emerged in the seventeenth century, and the computing paradigm in the twentieth century. In today's big data era, a new paradigm of data-intensive scientific discovery has emerged that is less dependent on models and *a priori* knowledge. By seeking relationships from huge amounts of data, new models, new knowledge, and new laws can be discovered and explored.

Scientific big data has a number of characteristics, including complexity, comprehensiveness, globalization, and high integration of information and communication technology. The scientific approaches are also transforming from single disciplines to multidisciplinary and interdisciplinary; from natural science to the integration of natural and social sciences; and from individual or small research groups to international scientific organizations. In addition to scientists being able to solve hard or untouchable problems through real-time dynamic monitoring and analysis of various related data, data itself can become the object and tool of research: Scientists can conceive, design, and implement their research based on the data [8].

Scientific big data has made changes in the scientific world, and research has entered into a new paradigm—a paradigm of data-intensive science. For the past few years, the National Science Foundation (NSF) has supported data-intensive scientific computation through programs. The Texas Advanced Computing Center, in partnership with Dell and Intel, has built a world-class supercomputer, Stampede. Stampede is a cornerstone of the NSF's investment in an integrated advanced cyberinfrastructure which empowers America's scientists and engineers to interactively share advanced computational resources, data, and expertise in order to further research across disciplines. Stampede is now the most powerful and capable of the 16 high-performance computing, visualization, and data analysis resources within the NSF's Extreme Digital (XD) environment. It has comprehensive processing capacity, high availability, and high performance. It has already enabled research teams to predict where and when earthquakes may strike, how much sea levels could rise, and how fast brain tumors grow (http://www.nsf.gov/news/news_summ.jsp?cntn_id=121763; http://www.nsf.gov/news/news_summ.jsp?cntn_id=127194). The California Earthquake Center used Stampede to forecast the frequency of destructive earthquakes in California, and a research team at the University of Texas at Austin used Stampede to better understand and represent the flow of ice from Antarctica into the sea using detailed numerical models [22].

Although scientific big data has become important to research, and the paradigm of data-intensive scientific discovery has been widely recognized, its associated theory, methodology, and models have yet to be employed in depth. At present, the concept and application of big data have been accepted and developed in network sciences and economic fields. But in contrast, the theoretical study and practice of scientific big data are relatively weak. This is because it has its own specific scientific connotation, called "3H", which includes the following three points:

(1) High dimension: Scientific big data represents the complicated relationship between natural and social sciences. In general, the external representations of these natural phenomena or scientific processes have high correlation and multiple data attributes. In principle, scientific big data has a high dimension [23]. As an example of time and space analysis of large-scale, complex social and economic phenomena in a geographic information system, every coordinate in space overlays all sorts of natural geographic data, spatial observation data, and socioeconomic and

cultural data. The correlation of these data is complex, coming from different sensors, with different spatiotemporal resolutions and physical significance [24].

(2) High complexity: Scientific big data mostly applies to complex nonlinear systems, and is accompanied by a complex data model. Therefore, the issue for scientific big data computation is not merely a matter of data processing and analysis; it is also a matter of joint modeling and computation with complex system modeling and data [25]. It requires integrating complex systems theory, estimation theory, and the mechanism models of corresponding disciplines to explore the solution. Modern climate science is an example [26].

(3) High uncertainty: Scientific big data, in general, comes from the natural process of perception and data acquisition. Because of the characteristics of these data sources, scientific big data generally has some error and incompletion, which results in data with high uncertainty. Scientific big data is often applied to the disciplines of natural systems, such as climate change and geo-processes. The system is represented by approximate models with concomitant high uncertainty [27]. The uncertainty of these models brings enormous challenges to computing scientific big data.

In order to overcome the challenges of big data, especially those of scientific big data, computability and solution-strategy approaches need to be developed. The first approach involves finding an approximate solution instead of an exact solution. Within the scope of acceptable precision, it is a feasible approach for reducing the complexity of solving and improving efficiency [28]. The second approach is to acquire a solution through sparse representation and dimensionality by turning a big dataset into small datasets. The original Fourier transform and wavelet transform, as well as the most popular current compressive sensing [29, 30] and dictionary learning [31], are based on this idea.

3 The theoretical framework of Digital Earth

The concept of Digital Earth was popularized in 1998 by Al Gore to describe a digital future where, in his example, a schoolgirl could interact with a computer-generated three-dimensional spinning virtual globe and access vast amounts of scientific and cultural information [32]. A popular explanation given more than 10 years ago states the following: "put the Earth into the computer". A number of Digital Earth systems have been developed for commercial, social, and scientific applications. In 2005, Google Earth made it possible for the public to explore Earth's surface freely through personal computers.

In the "Big Data Age", the Digital Earth concept has taken on a new connotation. It is a virtual Earth constructed with massive, multi-resolutioned, multi-temporal, and multi-type Earth observation and socioeconomic data as well as analysis algorithms and models [33–36]. The birth and development of big data have introduced new challenges to Digital Earth and brought it into a new generation [37–39].

Based on the above description, the basic issues of Digital Earth science include the following two aspects: (1) aggregating, representing, and analyzing multi-sourced, multivariate, heterogeneous, multi-scaled, highly spatially and temporally propertied, massive data; and (2) constructing, quantitatively analyzing, and modeling complex geoscience processes and socioeconomic phenomena.

Therefore, the basic theoretical framework of Digital Earth includes the theory of geospatial information and the theory of Earth system science.

Geospatial information theory includes Earth observation data acquisition models, data aggregation models and methods (including map cognition, sparse representation, and data fusion), data characterization theory, geospatial data analytical models and theory, and the information flow model and information field theory.

Earth system science is not intended to analyze the geoscience processes of a particular subsystem of the Earth. It is focused on the spatial and temporal analysis of complex geological processes, modeling theory in complex nonlinear systems, and decision support. Specifically, it includes the following: (1) multivariate and multiscale spatial and temporal analysis and decision support of geological processes with spatial and temporal properties; and (2) multivariate, multi-process, nonlinear, and highly coupled geoscience process modeling and systems analysis.

4 Digital Earth and big data

The science and engineering of Digital Earth fully encompass the previously mentioned "4V" characteristics of big data. Fig. S1 shows the Earth observation data flow in Digital Earth. It can be seen that in size, the data in Digital Earth has reached the exabyte (EB) level. Its main data includes images, videos, documents, and geographic information. It also involves Earth observation, scientific models, and socioeconomic and other data types [40]. Its wide range of data sources, real-time data access, and quick updates lead to a lower level of data density. In addition, the new generation of Digital Earth systems have the ability to handle massive data and quickly turn it into information to help deal with disasters and ecological issues.

The data system of Digital Earth includes geospatial data and a data platform. Geospatial data are the core part of Digital Earth. It includes various scales of geography and spatial data; multi-sensor, multi-temporal, multi-resolution Earth observation data; and various social spatial data (agriculture, resources, environment, and economy). The Digital Earth data platform, supported by national spatial data infrastructures and built on high-speed Internet, connects with multiple satellite data centers and geographical information centers to complete the whole process of spatial data acquisition, transmission, storage, processing, analysis, calculation, and distribution.

The Digital Earth data system employs the three aspects of "quantity, quality, and use". "Quantity" focuses on breakthroughs in science and technological problems brought about by the size of big data. "Quality" involves the analysis of the nature and mechanisms of the data, such as the dataset structure, system evolution, and correlation of the elements. "Use" focuses on the analysis of the method and theory of big data-driven modeling, information extraction, and knowledge discovery. The above three aspects reinforce and promote each other: Only by addressing quantity problems can we study the mechanisms of the data. Understanding the nature and mechanisms of big data can enable us to contribute to its management and better use. Research on the use of big data will lead to better understanding of the mechanisms from the viewpoint of demands, and will focus on big data processing and analysis.

More specifically, the Digital Earth data system currently uses the following methods:

(1) Quantity: Spatial data in Digital Earth are multi-sourced, heterogeneous, and massive. We should research distributed multi-source data aggregation mechanisms, such as multi-datacenter architecture [41, 42], distributed data warehouses, and virtual data aggregation [43] in order to improve data access, organization, and efficiency. We should also develop high-performance, massively parallel spatial data processing platforms to improve the performance [44, 45].

(2) Quality: Spatial data in Digital Earth are multivariate, high-dimensional, highly spatially, and temporally propertied. We need to study sparse characterization methods and models based on the analytical dictionary (such as wavelet transformation) and non-analytical dictionary (e.g., over-complete dictionary), the data flow model, the information flow model, and the information field theory in order to obtain the natural information within massive spatial data.

(3) Uses: Digital Earth uses massive data and knowledge mining. Using dynamic data-driven geoscience processes, it builds simulation models [46, 47], thus constructing complex geoscience processes and systems, reducing system complexity, inferring unknown mechanisms of geoscience processes, and supporting decisions.

5 Case study: the spatial cognizance of the Heihe-Tengchong Line

The development of Digital Earth offers opportunities for China's rapid development by providing significant decision support for sustainable economic and social development. In the context of urbanization, the study of Digital Earth can provide quantitative analysis and system simulation for significant social and economic problems, thereby providing the scientific basis for major decisions. In this paper, we take the scientific cognizance of the Heihe-Tengchong Line as an example to briefly introduce Digital Earth research methods in the context of big data.

The Heihe-Tengchong Line was first proposed by Prof. Hu Huanyong in 1935 to demarcate the population density in China between "the heavy south and the light north". Until the twenty-first century, this pattern has generally remained unchanged [48]. Since the Heihe-Tengchong Line is related to strategic issues of China's development, including resources, environment, ecology, and food security, its scientific cognizance has become increasingly significant in the age of China's urbanization.

In essence, the Heihe-Tengchong Line is a geospatial division and a geographic cognizance problem. Simultaneously, it embodies complex social and economic processes, like population, agriculture, industrial production, and transportation. The line evolved over a millennium and crossed thousands of kilometers of space; in other words, it has a long temporal sequence.

If we used traditional models or computing paradigms to research the Heihe-Tengchong Line's formation, we would have to build a giant complex nonlinear system, including mechanistic models of various processes; model interaction of various scales; and extremely complex model triggers, transmission, and feedback mechanisms. Furthermore, most of the above models would be inaccurate, incomplete, and with limitations. Therefore, these model systems are uncontrollable and unobservable, and the system behaviors are unpredictable.

What is noteworthy is that the Heihe-Tengchong Line is accompanied by long sequences of massive spatial and socioeconomic data, including geography, climate, remote sensing satellite images, population, industrial and

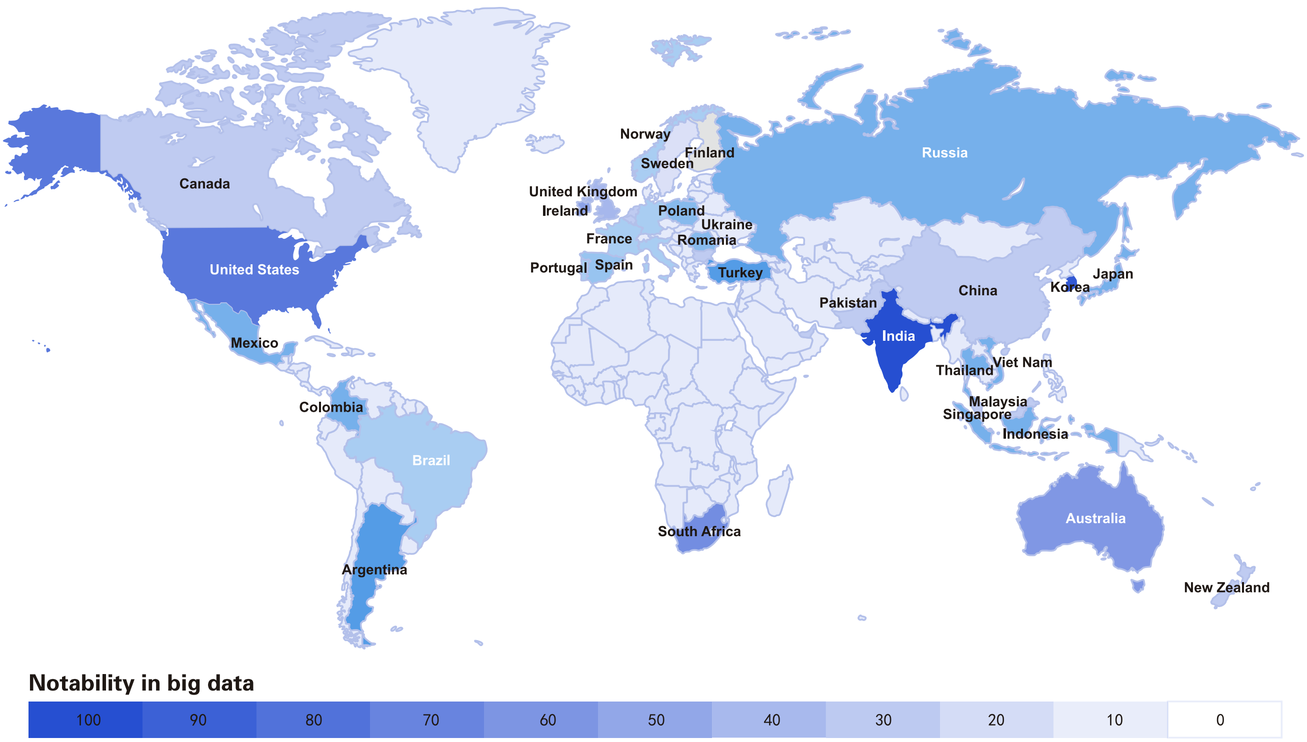

Fig. 3 Spotlight on big data in different countries (based on Google Trends; acquired in August 2013)

agricultural production, resources, and environment. These data have typical features of the scientific big data in Digital Earth, such as being high-dimensional, multivariate, multi-scaled, and having long sequences highly coupled with spatial and temporal properties. In order to research the Heihe-Tengchong Line, the Digital Earth system is the only viable data acquisition, organization, management, and analysis platform. It is not feasible to simply pile up and couple sub-processes or sub-model mechanisms to gain scientific cognizance. A reasonable approach is to first cluster massive spatial and socio-economic data on the Digital Earth platform, then model it for related data after a series of pretreatment processes, and finally mine and analyze the minimal dataset. In such a process, how one researches and develops processing theories and methods is the key to solving problems. This can involve the virtual aggregation and dynamical update of distributed multi-sourced heterogeneous spatial data; data integration, assimilation, integration, and representation of massive spatial data; various dimension reduction theories, including differential geometry and manifold analysis; and using a variety of artificial intelligence and machine learning methods to conduct massive data mining and analysis.

6 Conclusions

There is no doubt that humanity has entered the era of big data. Big data computing conforms to the universal law of technological development, including innovation, development, and maturity. According to the hype cycle shown in Fig. 4, a technology's evolution can be divided into the following five stages: Technology Trigger, Peak of Inflated Expectations, Trough of Disillusionment, Slope of Enlightenment, and Plateau of Productivity. Gartner's 2012 Hype Cycle Special Report claims that big data is one of 1,900+ technologies in the new hype cycle, and that big data technology will reach the Plateau of Productivity within 2–5 years.

Big data research is facing an imbalance between disciplines. Currently, research on big data technology and methodologies is focused on related applications of Internet and business intelligence, while the scientific and engineering disciplines which produce big data are paying less attention. Scientific big data is different from the big data of the Internet or network sciences. Further study of scientific big data will contribute to setting up a theoretical system and technical framework of big data—which would constitute the emergence of a new discipline.

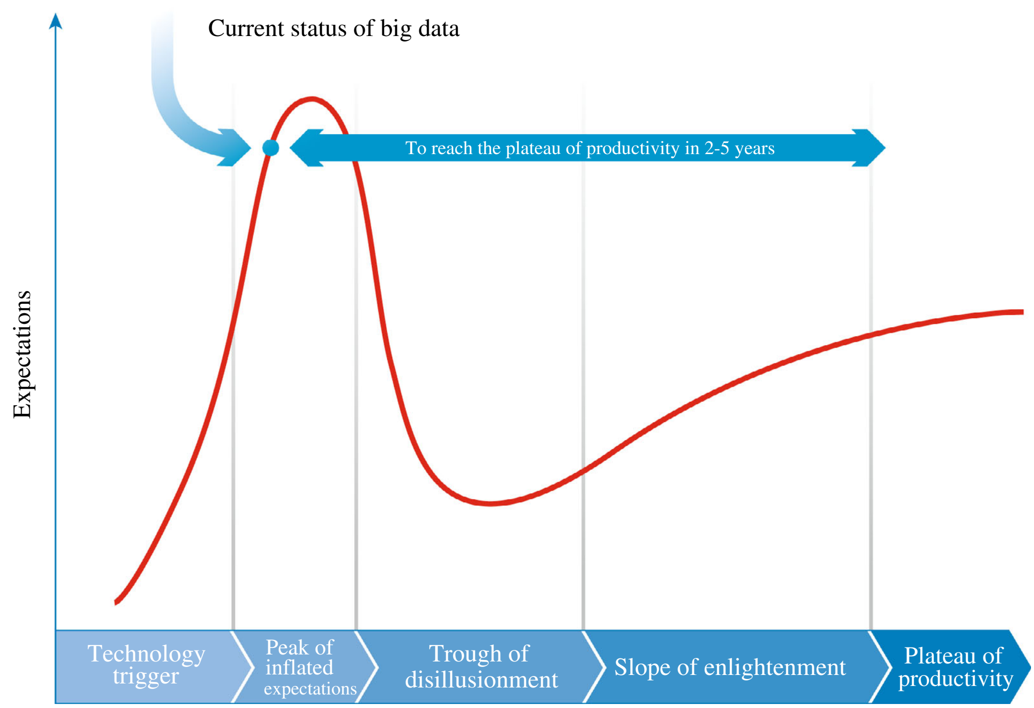

Fig. 4 Hype Cycle of big data (based on Gartner's 2012 Hype Cycle Special Report)

The development of scientific big data must be coordinated with other disciplines that possess big data sources. We need to jointly study its organization, storage, transmission, analysis, visualization methods, and the general rules for technology, while also paying close attention to the substantive characteristics of scientific big datasets, including data association, spatial alternation, and system evolution. Scientific big datasets, system theory, and methodology could then be developed based on these mechanisms.

Scientific big data is the engine for discovery and innovation. The development of scientific big data involves developing a national strategy, allocating resources for big data research, developing fundamental theories on big data, and promoting big data applications in various fields. Scientific big data covers nearly all fields in the scientific community, and it can be applied to all walks of life. It plays a role in promoting science and technology, the economy, and society. China should swiftly organize to develop scientific big data research in order to promote the development of Earth science and related technologies.

Digital Earth is a typical application of scientific big data. Its data acquisition, organization, analysis, and application reflect almost all characteristics of scientific big data. From the perspective of scientific big data, Digital Earth could be considered big Earth data, or it could be interpreted as a discipline based on the development of big Earth data [49].

Acknowledgments This work is supported by the International Cooperation and Exchange of the National Natural Science Foundation of China (41120114001).

Conflict of interest The authors declare that they have no conflict of interest.

References

1. EMC (2011) Big data: big opportunities to create business value. EMC Corporation, Hopkinton
2. Bryant RE, Katz RH, Lazowska ED (2008) Big-data computing: creating revolutionary breakthroughs in commerce, science, and society. Computing Community Consortium. http://www.cra.org/ccc/files/docs/init/Big_Data.pdf. Accessed 10 Nov 2013
3. Wigan MR, Clarke R (2013) Big data's big unintended consequences. Computer 46:46–53
4. Frankel F, Reid R (2008) Big data: distilling meaning from data. Nature 455:30
5. Gantz J, Reinsel D (2012) The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. IDC Analyze the Future, Framingham
6. Meng XF, Ci X (2013) Big data management: concepts, techniques and challenges. J Comput Res Dev 50:146–169 (in Chinese)
7. Nature (2008) Big data. Nature 455:1–136
8. Hey T, Tansley S, Tolle K (2009) The fourth paradigm: data-intensive scientific discovery. Microsoft Research, Washington
9. Cukier K (2010) Data, data everywhere: a special report on managing information. Economist. http://www.economist.com/node/15557443#sthash.QG9d5Gkr.dpbs. Accessed 22 Aug 2013
10. Science (2011) Dealing with data. Science 331:639–806
11. Manyika J, Chui M, Brown B et al (2011) Big data: the next frontier for innovation, competition, and productivity. Mckinsey

Global Institute. http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx. Accessed 03 Sep 2013
12. United Nations Global Pulse (2012) Big data for development: challenges and opportunities. http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf. Accessed 05 Aug 2013
13. Douglas L (2012) The importance of big data: a definition. Gartner. https://www.gartner.com/doc/2057415?ref=SiteSearch&sthkw=The%20importance%20of%20big%20data%3A%20a%20definition&fnl=search&srcId=1-3478922254. Accessed 30 Dec 2012
14. Goodchild MF, Guo HD, Annoni A et al (2012) Next-generation digital earth. Proc Natl Acad Sci USA 109:11088–11094
15. Executive Office of the President (2014) Big data: seizing opportunities, preserving values. http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf. Accessed 01 May 2014
16. Executive Office of the President, President's Council of Advisors on Science and Technology (2014) Report to the president big data and privacy: a technological perspective. http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf. Accessed 01 May 2014
17. Office of Science and Technology Policy Executive Office of the President (2012) Obama administration unveils "big data" initiative: announces $200 million in new r&d investments. http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf. Accessed 25 Jun 2012
18. Fang JQ (2013) Network science and engineering faced with a new challenge and developing opportunity under the wave impact of big data. Chin J Nat 35:345–354 (in Chinese)
19. Bai CL (2013) Preface. In: Chinese Academy of Sciences (eds) China's e-science blue book 2013. Science Press, Beijing, i-ii
20. Mayer-Schönberger V, Cukier K (2013) Big data: a revolution that will transform how we live, work and think. Houghton Mifflin Harcourt Publishing Company, New York
21. Li GJ, Cheng XQ (2012) Research status and scientific thinking of big data. Bull Chin Acad Sci 27:647–657 (in Chinese)
22. Zhao Y (2013) NSF official on new supers, data-intensive future. e-Sci Technol Appl 4:91–93 (in Chinese)
23. Abarbanel HDI, Brown R, Sidorowich JJ et al (1993) The analysis of observed chaotic data in physical systems. Rev Mod Phys 65:1331–1392
24. Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. Wiley, New Jersey
25. Rocha LM (1999) Complex systems modeling: using metaphors from nature in simulation and scientific models. Los Alamos National Laboratory, Los Alamos
26. Barry RG, Chorley RJ (2009) Atmosphere, weather and climate. Routledge, London
27. Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. J R Stat Soc Ser B (Stat Methodol) 63:425–464
28. Han XX, Li JZ et al (2013) Efficient skyline computation on big data. IEEE Trans Knowl Data Eng 25:2521–2535
29. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52:1289–1306
30. Candès EJ, Wakin MB (2008) An introduction to compressive sampling. IEEE Signal Proc Mag 25:21–30
31. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54:4311–4322
32. Gore A (1992) Earth in the balance: ecology and the human spirit. Plume, Boston
33. Grossner KE, Goodchild MF, Clarke KC (2008) Defining a digital earth system. Trans GIS 12:145–160
34. Guo HD (2012) China's earth observing satellites for building a digital earth. Int J Digit Earth 5:185–188
35. Craglia M, Ostermann F, Spinsanti L (2012) Digital earth from vision to practice: making sense of citizen-generated content. Int J Digit Earth 5:398–416
36. Foresman T, Schade S, Georgiadou Y et al (2014) Does DE need a C? A proposal for a DE curriculum. Int J Digit Earth 7:88–92
37. Guo HD, Wang CL (2013) Digital earth: fifteen years experience and perspective. Bull Chin Acad Sci 28:59–66 (in Chinese)
38. Craglia M, Bie KD, Jackson D et al (2012) Digital earth 2020: towards the vision for the next decade. Int J Digit Earth 5:4–21
39. Ehlers M, Woodgate P, Annoni A et al (2014) Advancing digital earth: beyond the next generation. Int J Digit Earth 7:3–16
40. De Longueville B, Annoni A, Schade S et al (2010) Digital earth's nervous system for crisis events: real-time sensor web enablement of volunteered geographic information. Int J Digit Earth 3:242–259
41. Wang LZ, Tao J, Rajiv R et al (2013) G-Hadoop: MapReduce across distributed data centers for data-intensive computing. Future Gener Comput Syst 29:739–750
42. Zhang WF, Wang LZ, Liu DS et al (2013) Towards building a multi-datacenter infrastructure for massive remote sensing image processing. Concurr Comput Pract Exp 25:1798–1812
43. Chen J, Xiang LG, Gong JY (2013) Virtual globe-based integration and sharing service method of geospatial information. Sci China Earth Sci 56:1780–1790 (in Chinese)
44. Ma Y, Wang LZ, Zomaya AY et al (2013) Task-tree based large-scale mosaicking for remote sensed imageries with dynamic dag scheduling. IEEE Trans Parallel Distrib Syst. doi:10.1109/TPDS.2013.272
45. Yang CW, Xu Y, Nebert D (2013) Redefining the possibility of digital earth and geosciences with spatial cloud computing. Int J Digit Earth 6:297–312
46. Wang LZ, Chen D, Liu WY et al (2012) Parallel simulation of threat management for urban water distribution systems with MapReduce in clouds. Comput Sci Eng. doi:10.1109/MCSE.2012.89
47. Kim IH, Tsou MH (2013) Enabling digital earth simulation models using cloud computing or grid computing: two approaches supporting high-performance GIS simulation frameworks. Int J Digit Earth 6:383–403
48. Naughton B (2006) The Chinese economy: transitions and growth. The MIT Press, Cambridge
49. Guo HD (2014) Digital earth: big earth data. Int J Digit Earth 7:1–2