Big Earth Data Science and Engineering Postprint

Authors: Guo Huadong

Date: 2023-03-19T00:00:00+00:00

Abstract

Big data represents a strategic high ground in the era of the knowledge economy and constitutes a novel strategic resource for nations and the globe. As a crucial component of big data, Earth big data is emerging as a new frontier in Earth science, holding significant importance in advancing the profound development of Earth science and enabling major scientific discoveries. Based on an analysis of the characteristics of Earth big data, this article introduces the ongoing Strategic Priority Research Program of the Chinese Academy of Sciences, "Earth Big Data Science Project," examines the project's objectives and scientific connotations, and analyzes how Earth big data can facilitate the achievement of the United Nations Sustainable Development Goals. The Earth Big Data Science Project integrates cutting-edge scientific and technological advancements from interdisciplinary and convergent fields such as Earth science, information science, and space technology, heralding new prospects for Earth big data to serve the development of Earth science.

Full Text

Introduction

Big data represents a strategic high ground in the era of knowledge economy and constitutes a novel strategic resource for nations and the globe. As a critical component of big data, Earth big data is emerging as a new frontier in Earth science, holding significant importance for advancing deep Earth science development and enabling major scientific discoveries. This paper analyzes the characteristics of Earth big data, introduces the ongoing "Big Earth Data Science Engineering" (CASEarth) Strategic Priority Research Program of the Chinese Academy of Sciences, examines the program's objectives and scientific connotations, and analyzes how Earth big data can facilitate the achievement of the United Nations Sustainable Development Goals. The Big Earth Data Science Engineering integrates cutting-edge science and technology from interdisciplinary fields including Earth science, information science, and space technology, opening new prospects for Earth big data to serve Earth science development.

Understanding Earth Big Data

In recent years, the international community has launched a series of research initiatives related to Earth big data. The United States' "EarthCube" project seeks to adopt a holistic perspective on the Earth system and create infrastructure for managing Earth science knowledge. The European Union's "Living Earth Simulator" research program relies on open standards to provide multidimensional data and scalable services for all Earth science domains. Building upon its Data Cube project, Australia has recently launched the "Digital Earth" program. In 2017, Russia officially initiated its "Digital Earth" plan, which includes launching a series of Earth observation satellites to provide continuous data sources for the program. China has also conducted Earth observation data sharing in digital Earth and global Earth observation systems, integrating satellites, ground monitoring, and modeling systems to assess environmental conditions and make predictions to address climate change challenges and promote Earth system science development. Evidently, Earth big data development is burgeoning globally, making research in this field critically important.

Earth big data encompasses data related to the Earth's atmosphere, land, and oceans, generated through space-based Earth observation, terrestrial sensor networks, and other observation methods. These Earth-related big data are collectively termed Earth big data, which, while being an important component of big data, possesses unique characteristics [1-3]. Earth big data features large scale, diverse sources, multiple temporal phases, multi-scale dimensions, high dimensionality, high complexity, and unstructured nature, representing a new data-intensive research direction for Earth science. Earth big data is Earth-domain big data with spatial attributes, possessing general properties of big data while exhibiting strong spatiotemporal and physical correlations [14]. These characteristics can significantly advance geoscience disciplines and play crucial roles in environment, resources, disaster management, and other fields with substantial economic and social value. Earth big data brings important development opportunities for in-depth Earth science research, can lay the foundation for creating spatial Earth information science, and further promote Earth science development.

Earth big data exhibits characteristics of big data, including large volume, broad sources, multiple temporal phases, and high value, while also featuring high instantaneousness, arbitrary spatiality, and physical correlation. In terms of "massive scale," Earth big data possesses high resolution, high dynamic nature, and multi-spectral characteristics, with high data acquisition speeds and rapid update cycles. The "broad sources" stem from diverse data sources and collection methods, attributable to varied imaging principles and model changes. "Multiple temporal phases" refers to its capability for extremely short sampling intervals and high-frequency information collection. "High value" arises from its significance for ecological environment, land resources, natural disasters, and other Earth science research.

Related technologies for Earth big data include Earth observation technology, communication technology, computing technology, and network technology. Earth big data can enhance human understanding of Earth while facing new challenges in data transmission, storage, processing, analysis, management, and sharing [5-7]. For instance, real-time acquisition of massive Earth observation data and multi-source data fusion, along with data from large-scale, long-term, low-cost sensor networks, increase storage, processing, and computational complexity. Integrated data storage environments require consideration of interaction technology and visualization between physical storage devices and unified storage platforms.

As a key technology for Earth big data, Earth observation technology has, after nearly half a century of development, provided new perspectives and methods for Earth science research, particularly playing an enormous role in macro-level cognition of Earth systems [8-12]. With the help of Earth observation technology, humans can observe Earth more conveniently and systematically, making it an important indicator for measuring a nation's scientific achievements, economic strength, and national security [13]. Currently, Earth observation data volume is growing exponentially. According to a global satellite study by the Committee on Earth Observation Satellites, over the past half-century, more than 514 Earth observation satellites have been launched for comprehensive observation of Earth systems including atmospheric, oceanic, and terrestrial systems.

Earth Big Data Objectives

The Chinese Academy of Sciences fully recognizes the importance of Earth big data and established the Strategic Priority Research Program (Category A) "Big Earth Data Science Engineering" (CASEarth) in early 2018 to systematically conduct Earth big data research. CASEarth's mission is to utilize Earth big data to drive interdisciplinary, cross-scale macro-level scientific discoveries, systematically and holistically investigate a series of major scientific questions, achieve major breakthroughs in understanding Earth system science, realize new leaps in decision support, and continuously produce significant outputs in scientific discovery, macro-level decision-making, technological innovation, and knowledge dissemination.

The overall objective of CASEarth is to build an international Big Earth Data Science Center, comprising three main components: (1) Establish Earth big data infrastructure. Break through bottlenecks in data open sharing, form a multidisciplinary integrated Earth big data and cloud service platform, and become a national big data major scientific infrastructure supporting national macro-level decision-making and major scientific discoveries. (2) Form an Earth big data discipline-driven platform. Explore new paradigms for big data-driven, multidisciplinary integrated, globally collaborative scientific discovery, demonstrating and driving major breakthroughs in Earth system science, life science, and related disciplines. (3) Construct a decision support system serving high-level government. Enable multi-issue, multi-perspective panoramic visualization analysis, simulation, and deduction capabilities, display and dynamically deduce the sustainable development process and situation of the "Belt and Road" initiative, and achieve precise evaluation and decision support for the panoramic "Beautiful China" sustainable development and national globalization strategy oriented toward a community with a shared future for mankind.

Big Earth Data Science Engineering (CASEarth)

CASEarth will demonstrate distinctive features and achieve important outputs in three aspects: (1) Realize a series of scientific discoveries. Form new methods and paradigms for big data-driven scientific discovery, reveal complex coupling interactions between different elements at global and regional scales by reproducing spatial distributions and temporal dynamics of land, ocean, atmosphere, and human society parameters, and reveal details and coupling associations of elements at different resolutions and levels. (2) Generate series of technological innovations. Build a high-precision Earth big data cloud service platform and new digital Earth system, integrating and displaying massive data and information products through precise geographical and physical associations. Establish a "nationally distributed, unified yet separate, transparent service" Earth big data and cloud service platform, forming a multidisciplinary, evolvable, service-oriented Earth big data major infrastructure. (3) Serve government decision support. Establish a big data-driven, visualized, interactive, dynamically evolvable decision support environment, achieve integrated digital reproduction of multi-source spatial information and multi-element cross-integration assessment, and provide a macro-level real-time Earth big data decision support system.

Connotation of the CASEarth Program

The CASEarth program is decomposed into the following eight research components to achieve theoretical and technological breakthroughs and innovative results. The program particularly emphasizes data sharing and encourages domestic and foreign scholars to jointly conduct research based on this platform (Figure 1 [Figure 1: see original paper]).

(1) **CASEarth Small Satellite Development**. Develop small satellites serving CASEarth, establish CASEarth satellite operation management and assessment systems, and complete CASEarth satellite data reception and product services. Research overall design technology for observation missions, high-integration and miniaturized payload technology, infrared and multi-spectral payload technology, and massive data compression, storage, and transmission technology. Through research on satellite over-

all design and payload development, satellite engineering development, satellite operation management and assessment, and satellite data reception and product services, form a complete system process from satellite requirements to data products.

(2) **Big Data and Cloud Service Platform**. Build an Earth big data cloud service platform with integrated service capabilities, providing unified computing and storage cloud services. Research standards, protocols, tools, and systems for multi-source heterogeneous massive data access, aggregation, storage management, and unified access, integrating massive multi-source scientific data resources and space-air-ground integrated Earth big data to build a distinctive Earth big data resource library. Break through technologies for unified scheduling and aggregation services of distributed computing resources, grid data computing technology, and new methods for big data computing processing and analysis mining, achieving big data-driven scientific discovery and decision support.

(3) **Digital "Belt and Road"**. Construct Earth big data integration technology and evaluation systems for the "Belt and Road" initiative, as well as multi-element scientific databases, achieving comprehensive integration of "Belt and Road" Earth big data including spatiotemporal data for 49 major categories of elements over nearly 50 years. Conduct scientific analysis of "Belt and Road" Earth big data to understand scientific patterns of environmental and resource spatial distribution, development potential, and change trends across the entire region. Establish a regional spatial assessment index system oriented toward sustainable development goals, achieve scientific monitoring of key indicators for "Belt and Road" sustainable development goals, and establish a "Belt and Road" Earth big data analysis and decision support system.

(4) **Panoramic "Beautiful China"**. Under multi-angle, multi-dimensional, multi-link, multi-factor, and multi-level perspectives, guided by Earth system science and human-land relationship theory, conduct research and development on resource-environment background distribution and pattern evolution, clean air and environmental health, ecological civilization construction, regional development and smart cities, and "Panoramic 'Beautiful China' " evaluation and decision support systems based on big data. Comprehensively display the background characteristics, clean air and environmental health, ecological civilization construction, and urban development scenarios of "Beautiful China," evaluate and predict the current status and future scenarios of "Beautiful China" construction, and provide policy recommendations for "Beautiful China" construction.

(5) **Biodiversity and Ecological Security**. Research standards supporting data integration and sharing and data integration application methods. Organically integrate biological resource data with ecological, environmental, meteorological, and national economic data to form complete data layers. Utilize analysis models and visualization technology

to achieve functional mining and utilization of biodiversity resource data, build open-source universal interfaces for biodiversity and ecological security big data processing and utilization, and establish a comprehensive big data platform centered on biodiversity and ecological security information to achieve personalized data services and decision support at different levels.

(6) **Three-Dimensional Information Ocean**. Form an ocean information resource pool of "two points and one surface," where "one surface" refers to a global-scale ocean information resource pool and data products, establishing a global ocean basic data service system. "Two points" refers to focusing on advantageous directions and regions, conducting information integration and scientific research at two strategic key points: "China's coastal waters" and the "Two Oceans and One Sea." In key regions of the "Two Oceans and One Sea," research multi-source data and change databases and structural models for South China Sea islands and reefs, a Western Pacific deep-sea biogeographic information system and multi-dimensional demonstration system for deep-sea extreme habitats, and Indian Ocean and key port area marine disaster assimilation data products and marine disaster prediction and early warning systems, ultimately achieving real-time display, dynamic simulation, and scenario analysis of ocean information.

(7) **Spatiotemporal Three-Pole Environment**. Through basic research on three-pole thematic big data sharing and integration, three-pole remote sensing comparative studies, three-pole big data analysis methods, and multi-sphere interaction models, as well as thematic research on three-pole ecological spatiotemporal dynamics and forecasting, three-pole water environment and future water security, three-pole climate change and its impact on China, Arctic shipping route monitoring and refined prediction, frozen soil changes and their service functions, and major polar engineering permafrost issues, achieve occupation of the commanding heights of Earth system science and provide decision support for polar governance and Arctic development.

(8) **Digital Earth Science Platform**. This platform serves as the comprehensive display platform for the "Big Earth Data Science Engineering" program, focusing on building Earth big data comprehensive display and decision support systems and networked information service systems. It primarily provides visual analysis support for data, services, computing, and other resources and their relationships for scientific discovery and technological innovation oriented toward multidisciplinary integration and big data-driven approaches, while also serving scientific communication and public service in resources, environment, biology, and ecology. Develop an elastic, scalable, multi-modal digital Earth science platform to ensure safe and reliable system operation, providing environments and tool support for scientific discovery and technological innovation.

CASEarth will break through bottlenecks in data open sharing, achieve comprehensive integration of dispersed data, models, and services in resources, environment, biology, ecology, and other fields, form a multidisciplinary integrated Earth big data and cloud service platform at globally advanced levels, build a big data-driven digital Earth science platform with global influence, panoramically display and dynamically deduce the progress of "Belt and Road" and "Beautiful China" sustainable development goals, and comprehensively enhance major output in scientific discovery, macro-level decision-making, technological innovation, and public knowledge dissemination services.

CASEarth Promoting UN Sustainable Development Goals

In September 2015, on the occasion of the 70th anniversary of the United Nations, heads of state and representatives gathered at UN Headquarters in New York to adopt "Transforming Our World: The 2030 Agenda for Sustainable Development." The core of this agenda is achieving global sustainable development goals, aiming for all countries and stakeholders to work together to halt Earth's degradation, consume and produce sustainably, manage Earth's environment and natural resources, enable Earth to meet the needs of present and future generations, put the world on a sustainable and resilient path, and form a world where humanity lives in harmony with nature [15].

The UN 2030 Agenda for Sustainable Development is an ambitious strategic action plan for people, planet, and prosperity, comprising 17 sustainable development goals, 169 specific targets, and 230 indicators. Currently, the United Nations, national governments, and international organizations are conducting research on indicator system construction and monitoring and assessment for the UN Sustainable Development Goals (SDGs) [16-18]. However, SDGs face many challenges in implementation, with data deficiency being the most formidable challenge for monitoring SDGs. Incomplete, inconsistent statistical systems and missing indicator systems are the main causes of data scarcity and poor quality [19,20]. The modeling of SDG monitoring and evaluation indicators is complex, and due to data availability limitations, not all selected indicators can be modeled during comprehensive evaluation [21]. Therefore, how to scientifically establish a comprehensive, interdisciplinary, multi-element interaction evaluation model library is a difficult problem.

To address these challenges and problems, the CASEarth program proposes science and technology innovation based on Earth big data to promote sustainable development goal achievement. Grounded on the Earth big data platform, it comprehensively integrates databases, model libraries, and decision method libraries in resources, environment, ecology, and biology fields, constructs a sustainable development evaluation index system and decision support platform, effectively monitors and assesses sustainability in resources, environment, and ecology, incorporates Earth big data into UN and China's sustainable development evaluation systems, and simultaneously serves China's ecological security and resource security protection efforts.

The CASEarth program's research around SDGs includes: (1) Through building an Earth big data sharing service platform, strive to become a data provider, producer, and official liaison for SDG indicator implementation; (2) Focus on SDG2, SDG6, SDG11, SDG13, SDG14, and SDG15 to conduct global, regional, national, and local-scale SDG key indicator selection, spatial assessment index system construction, and indicator monitoring and evaluation; (3) Address localization issues for sustainable development goal monitoring, especially in developing countries, and strive to promote China's application demonstration achievements as universal official SDG application demonstration cases; (4) Based on data collection and analysis, regularly monitor and assess sustainable development goal progress, forming the "Earth Big Data Supporting UN Sustainable Development Series Report"; (5) Build an open high-end think tank serving SDG target evaluation and implementation, lead relevant SDG work in China and internationally, and enhance scientific decision-making capabilities at all levels.

Earth big data provides a completely new methodology for Earth science research, is becoming a new key to understanding Earth and a new engine for Earth science research, and may bring major transformations to Earth science research. Utilizing Earth big data, combined with Earth system science models, to develop theories and methods for Earth big data knowledge discovery is a major scientific problem that should be addressed in Earth science. As a new disciplinary direction, Earth big data should be studied continuously, with emphasis on interdisciplinary research integrating Earth science, information science, and space technology, to develop the Earth big data research direction and promote Earth system science research to new heights. As a national innovation platform representing national level and capability, CASEarth will bring innovation in scientific methodology and research perspectives through Earth big data research, revolutionize sustainable development macro-level decision support, and form a new engine for Earth system science discovery.

References

1. Guo H D, Chen R S, Xu Z W, et al. Big data in natural sciences and humanities—A summary of the 6th Roundtable Meeting on Frontier Exploration between China and Germany. *Bulletin of Chinese Academy of Sciences*, 2016, 31(6): 707-716.
2. Guo H D. Big data, big science, big discovery—A summary of the International Workshop on Big Data and Scientific Discovery. *Bulletin of Chinese Academy of Sciences*, 2014, 29(4): 500-506.
3. Guo H D, Wang L, Liang D. Big earth data from space: A new engine for earth science. *Chinese Science Bulletin*, 2016, 61(7): 650-660.
4. Guo H D, Liu Z, Jiang H, et al. Big earth data: A new challenge and opportunity for digital earth's development. *International Journal of Digital Earth*, 2017, 10(1): 1-12.

5. Vatsavai R R, Ganguly A, Chandola V, et al. Spatio temporal data mining in the era of big spatial data: algorithms and applications. *ACM Sigspatial International Workshop on Analytics for Big Geospatial Data*, 2012.
6. Reinsel D, Gantz J, Rydning J. Data age 2025: The evolution of data to life-critical don't focus on big data. Framingham: IDC Analyze the Future, 2017.
7. Guo H D. Steps to the digital Silk Road. *Nature*, 2018, 554: 25-27.
8. He G, Wang L, Ma Y, et al. Processing of earth observation big data: Challenges and countermeasures. *Chinese Science Bulletin*, 2015, 60(5-6): 470-478.
9. Ramapriyan H, Brennan J, Waler J, et al. Managing Big Data: NASA Tackles Complex Data Challenges. *Earth Imaging Journal*. [2013-10-18]. http://eijournal.com/print/articles/managing-bigdata.
10. Pekel J, Cottam A, Gorelick N, et al. High-resolution mapping of global surface water and its long-term changes. *Nature*, 2016, 540: 418-422.
11. Hansen M C, Potapov P V, Moore R, et al. High-resolution global maps of 21st-century forest cover change. *Science*, 2013, 342: 850-853.
12. Guo H D. Big Earth data: A new frontier in Earth and information sciences. *Big Earth Data*, 2017, 1: 4-20.
13. United Nations. *Transforming our World: The 2030 Agenda for Sustainable Development*. New York: United Nations, 2015. https://sustainabledevelopment.un.org/post2015/transformingourworld/publication.
14. Anderson K, Ryan B, Sonntag W, et al. Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-spatial Information Science*, 2017, 20: 77-96.
15. World Health Organization. Global Burden of Disease. [2018-06-30]. http://www.who.int/topics/global_{{{burden}}*{{of}}*}{disease}/en/.
16. Getting started with the Sustainable Development Goals: A guide for stakeholders. [2018-06-30]. http://unsdsn.org/wp-content/uploads/2015/12/151211-getting-started-guide-FINAL-PDF-.pdf.
17. United Nations Statistics Division. *The Sustainable Development Goals Report 2018*. [2018-06-30]. https://unstats.un.org/sdgs/report/2018/.
18. Mark E, Bengtsson M, Akenji L. An optimistic analysis of the means of implementation for Sustainable Development Goals: Thinking about goals as means. *Sustainability*, 2016, 8(9): 962-977.

Author Information

Guo Huadong is a professor at the Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences (CAS), an academician of CAS, foreign member of the Russian Academy of Sciences, foreign member of the Finnish Society of Sciences and Letters, and fellow of The World Academy of Sciences for the advancement of science in developing countries (TWAS). He currently serves as President of the International Society for Digital Earth (ISDE), Member of the UN 10-Member Group to Support the Technology Facilitation Mechanism, Chairman of the International Committee on Remote Sensing of Environment (ICORSE), Director of the International Centre on Space Technologies for Natural and Cultural Heritage (HIST) under the auspices of UNESCO, Chair of the Science Committee of the Digital Belt and Road Program (DBAR), and Editor-in-Chief of the *International Journal of Digital Earth* and *Big Earth Data*. He specializes in remote sensing information science, radar Earth observation, and digital Earth. He is currently the principal investigator of the CAS Strategic Priority Research Program (Category A) "Big Earth Data Science Engineering." E-mail: hdguo@radi.ac.cn

Note: Figure translations are in progress. See original paper for figures.