Strengthening Open Data Infrastructure to Promote Open Science Development Postprint

Authors: Guo Huadong, Chen Hesheng, Yan Dongmei, Li Jianhui, Chen Gang, Guojin He, Ma Juncai, Liang Dong, Kong Lingyi, Guo Huadong

Date: 2023-07-09T00:00:00+00:00

Abstract

To promote economic development, social progress, and scientific and technological innovation, it is essential to vigorously promote open science, with open data becoming an important component thereof. Open data infrastructure, composed of physical entities and virtual systems, serves as the foundation for supporting open data and meeting data application needs across various domains; its development represents an objective process of information technology evolution. This article analyzes the elements of open data infrastructure, elucidates its significance and positive role in promoting the implementation of open science, examines the current state of open data infrastructure in China, and proposes development recommendations to address the existing deficiencies and challenges.

Full Text

Preamble

ChinaXiv Cooperative Journal Special Issue: Open Science Development Trends and Governance Strategies

Citation Format: Guo H D, Chen H S, Yan D M, et al. Strengthening open data infrastructure and promoting open science. *Bulletin of Chinese Academy of Sciences*, 2023, 38(6): 806-817

Title

Strengthening Open Data Infrastructure and Promoting Open Science

Authors

GUO Huadong[1,2*], CHEN Hesheng[3], YAN Dongmei[1,2], LI Jianhui[1,4], CHEN Gang[3], HE Guojin[1,2], MA Juncai[5], LIANG Dong[1,2], KONG Lingyi[1,2]

[1] International Research Center of Big Data for Sustainable Development Goals, Beijing 100094, China
[2] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
[3] Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China
[4] Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
[5] Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

Abstract

To promote economic development, social progress, and scientific and technological innovation, it is imperative to vigorously advance open science, with open data emerging as a crucial component. Open data infrastructure, composed of physical entities and virtual systems, serves as the foundation for supporting open data and meeting the application needs of diverse fields, representing an objective process in the evolution of information technology. This article analyzes the elements of open data infrastructure, elucidates its significance and positive role in facilitating the implementation of open science, examines the current state of China's open data infrastructure, and proposes development recommendations to address existing deficiencies and challenges.

Keywords: open data, open data infrastructure, open science, Big Earth Data, scientific data center

Main Body

Open science has now entered a stage of global consensus. Researchers have offered various definitions of open science from different perspectives, including activity-based, methodological, modality-based, process-based, cultural, and practical approaches [1]. Open science promotes science as a common good, encompassing the sharing of data, methods, results, and derived knowledge. As a global development trend, open science represents a paradigm shift in scientific research [2,3] that will profoundly transform the ways in which humanity conducts research and makes discoveries, holding significant importance for strengthening scientific cooperation and jointly addressing global challenges [4].

To advance global open science development, the 41st session of the UNESCO General Assembly in 2021 adopted the *Recommendation on Open Science* [5,6], aiming to make open science more transparent, accessible, equitable, and inclusive. This recommendation not only provides an international direction for open science development but also establishes channels for promoting exchange and building trust at various levels, including individuals, institutions, nations, regions, and the international community.

In recent years, the construction of large-scale scientific facilities, implementation of major scientific experiments, and widespread application of scientific sensors and sensor networks have generated multi-source, heterogeneous, and massive scientific data. Data has evolved from being merely the result of facts or observations in the research process to becoming an essential tool for research itself. Data-intensive scientific discovery has emerged as a new paradigm for scientific research in the big data era [7].

Open scientific data constitutes one of the core elements of open science [2]. Governments and institutional organizations typically formulate open data policies that define the types of data to be shared, target audiences, and sharing conditions [8]. Achieving public use, reuse, long-term preservation, and updated publication of scientific data will vigorously promote open science development, enhance the transparency, reproducibility, and collaboration of scientific and technological innovation activities, and maximize the value of science for societal development [9].

Open scientific data refers to scientific data that is publicly accessible and can be downloaded, copied, analyzed, reprocessed, and utilized for system construction and any other application purposes [10-12].

International Development and Current State of Open Scientific Data

In the United States, open scientific data has become an integral component of the nation's "information freedom and open government" initiative. The *Guidelines on the Handling of Research Data*, issued by the German Research Foundation (DFG) in 2010, explicitly states that "the sharing and reuse of scientific data hold tremendous significance for scientific research and even human society" [13]. The *Beijing Declaration on Research Data*, released by the Committee on Data for Science and Technology (CODATA) in 2019 [14], asserts that scientific data produced with public funding should be shared and reused globally to the greatest extent possible.

According to statistics from the Dimensions platform, a total of 9,918,741 open scientific datasets were available globally between 2012 and 2021, with China possessing 253,441 datasets, ranking second worldwide after the United States. In terms of research fields, the top three disciplines by data volume were information and computer science, information systems, and Earth science (Figure 1 [Figure 1: see original paper]). To further standardize open data, the FAIR Principles [15,16]—Findable, Accessible, Interoperable, and Reusable—were introduced in 2016. Complementing the FAIR Principles, the CARE Principles—Collective Benefit, Authority to Control, Responsibility, and Ethics—advocate for goal-oriented approaches that leverage the innovative potential of data.

While the FAIR Principles emphasize technological advancement, the CARE Principles focus more on policy transformation, with the two frameworks working in synergy [17].

Development and Current State of Open Scientific Data in China

1.2.1 Historically Rich Scientific Data Covering Various Fields According to incomplete statistics, by the end of the 20th century, China had established 5,000–6,000 scientific databases of varying scales and quality, covering all fields of science and technology. In terms of scientific data collection and accumulation, a preliminary structure had formed, with government departments as the main body and research institutes and universities complementing each other, alongside the gradual establishment of specialized data management agencies and international data cooperation and exchange channels [18].

According to the *National Scientific Data Resources Development Report (2018)*, by the end of 2017, the total volume of scientific data resources under effective management and preservation in China reached approximately 83.72 PB. Specifically, data accumulation in five key fields—life sciences and medicine, Earth and environmental sciences, physics and chemistry, Earth observation, and astronomy and space science—accounted for 26.81 PB, 24.48 PB, 16.64 PB, 9.73 PB, and 5.27 PB, respectively.

1.2.2 Scientific Data Open Sharing Gradually Incorporated into China's Policies and Regulations China has consistently emphasized the management and open sharing of scientific data. In 2006, the State Council released the *National Medium- and Long-Term Program for Scientific and Technological Development (2006–2020)*, which explicitly called for building digital technology platforms to promote scientific data sharing. China has gradually formed a data policy system centered on government agencies, industry institutions, and domain-specific data centers [19]. In 2015, the State Council issued the *Action Outline for Promoting Big Data Development*, proposing specific actions for developing scientific big data. In 2018, the General Office of the State Council published the *Measures for Scientific Data Management*, which further clarified responsibilities for scientific data management and sharing and provided clear definitions for scientific data, management protocols, responsible entities, usage methods, scope, and security and confidentiality, laying a foundation for scientific data opening and sharing [20].

1.2.3 Actively Participating in and Promoting International Scientific Data Cooperation and Sharing The International Science Council (ISC) currently has two major data organizations—the Committee on Data for Science and Technology (CODATA) and the World Data System (WDS, formerly the World Data Centers, WDC)—dedicated to organizing scientists working with data across all fields of science and technology and building a global-scale scientific data exchange system through international networks [21,22]. China joined WDS in 1988 and established nine data centers that same year in astronomy, space science, oceanography, meteorology, geology, seismology, geophysics, glaciers and frozen soils, and renewable resources and environment. Currently, the astronomy, space science, and oceanography data centers are regular members of WDS. China joined CODATA in 1984. In 2011, Chinese scholars proposed the "Hand-in-Hand Partnership Program" at CODATA to promote cooperation between international projects, facilitate the reuse of existing data resources, reduce redundant efforts, increase scientific research output, accelerate the transformation of research achievements, and bridge the digital divide [21,23].

The Chinese Academy of Sciences' Strategic Priority Program (Category A) "Big Earth Data Science Engineering" (CASEarth), launched in 2018, follows the development trend from open data to open science, empowering Earth big data with artificial intelligence to create a new model integrating data, computing, and services [24]. CASEarth promotes the integration of Earth science data, enabling multidisciplinary data correlation analysis and information fusion to drive major scientific discoveries and decision support in addressing major global sustainable development challenges [25].

Based on CASEarth, the International Research Center of Big Data for Sustainable Development Goals was officially established on September 6, 2021, aiming to provide theoretical foundations, technical methods, decision support, and think tank services for solving major sustainable development problems in China and globally [26,27]. This represents an important initiative and innovative direction for China to promote big data in service of the UN 2030 Agenda for Sustainable Development [28].

1.2.4 Actively Establishing Data Journals and Repositories to Promote Data Publishing As the concept of open data continues to deepen, China has also embarked on building scientific data journals. *China Scientific Data*, launched by the Computer Network Information Center of the Chinese Academy of Sciences (CAS) in 2016, is one of the first multidisciplinary data journals in China [29]. Reputable academic journals in fields such as biology, remote sensing science and technology, and library and information science have also successively established data paper columns [30]. *Big Earth Data*, co-sponsored by the International Society for Digital Earth and CASEarth and launched in December 2017, is the world's first geoscience journal focused on big data, aiming to build a first-class international academic exchange platform for scholars engaged in the collection, management, processing, analysis, and visualization of Earth big data [31].

China has also established data repository platforms in various fields to promote the open sharing of domain-specific data resources. Several domestic data platforms have been internationally certified as data repositories, thereby better safeguarding Chinese scientists' data sovereignty [32].

Open Data Infrastructure: Elements and Significance

Open data infrastructure serves as a shared facility supporting open data and meeting the needs of diverse scientific fields, representing an inevitable outcome of information technology evolution toward digitalization and intelligence. Data centers are the concrete manifestation of open data infrastructure, centering on data and achieving maximum value of open data through deep integration of computing, storage, network, and software resources. The elements of open data infrastructure comprise four components: data, physical, technical, and institutional. The data element is the primary component, encompassing datasets, data identifiers, and data registries. The physical element refers to the software and hardware infrastructure required for data standards, storage, management, sharing, analysis, and usage. The technical element involves core technology research and development for open data infrastructure, enabling collaborative and multidisciplinary data analysis through open computing and data processing capabilities. The institutional element provides policy guidance for the use and management of open data and open data infrastructure.

Significance and Positive Role of Open Data Infrastructure in Promoting Open Science Implementation

Open data infrastructure serves as a crucial catalyst for transforming research paradigms and a "new engine" for major scientific and technological breakthroughs [33,34]. With the development of the Internet, big data, and artificial intelligence, the pathways to scientific discovery have entered a new stage [35]. For example, the large-scale distributed data service infrastructure platform jointly developed by the Institute of High Energy Physics of CAS and the National High Energy Physics Science Data Center integrates important international high-energy physics scientific data and computing resources through ultra-high-speed networks, providing data services for scientific discoveries in particle physics, astrophysics, neutron science, photon science, and other fields. The Large High Altitude Air Shower Observatory (LHAASO) project utilizes this platform to deploy the Coryda data processing system for comprehensive collection and processing of ultra-high-energy gamma rays and other cosmic ray data. In 2022 alone, LHAASO collected 11 PB of data containing 10 trillion cosmic ray events, with data and computing resources open to the world. The annual data access and processing volume reached 448 PB, yielding major scientific achievements in frontier cosmic ray research, including the discovery of "PeV ultra-high-energy photons," "measurement of the ultra-high-energy cosmic ray light component spectrum," and "new discoveries of very-high-energy gamma-ray sources."

Current Status and Challenges of China's Open Data Infrastructure

Foundations and Advantages of China's Open Data Infrastructure

The construction requirements for open data infrastructure present diverse characteristics due to differences in target users and applications [36]. Open data infrastructure can be primarily categorized into two types: (1) large-scale scientific facilities, which are national infrastructure providing services based on common needs and serve as production units for massive data; and (2) scientific data centers, which are centralized data infrastructure supporting comprehensive research across interdisciplinary fields. The Chinese Academy of Sciences has played a significant role in the construction and systematic operation of China's open data infrastructure.

2.2.1 Large-Scale Scientific Facilities China's large-scale scientific facilities can be traced back to the major research installations built for the "Two Bombs, One Satellite" program. In the late 1980s, CAS took the lead in constructing the Beijing Electron-Positron Collider. During the Ninth and Tenth Five-Year Plan periods, China built 11 large-scale scientific facilities. After the Eleventh Five-Year Plan, development entered a rapid growth phase, with the Twelfth Five-Year Plan period seeing the completion of 22 facilities and 16 under construction. The Thirteenth and Fourteenth Five-Year Plan periods have gradually formed a construction plan for large-scale scientific facilities anchored by comprehensive national science centers [37]. Currently, China has approximately 50 large-scale scientific facilities in operation or under construction, with some facilities ranking among the global top tier [38].

Large-scale scientific facilities are divided into three categories: (1) specialized research facilities, built primarily for major scientific and technological objectives in specific disciplines; (2) public experimental facilities, which mainly support basic and applied research across multiple disciplines; and (3) public welfare scientific facilities, which provide fundamental data for national economic development, national security, and social progress [39]. The first two categories generally produce massive amounts of scientifically valuable data through experiments and observations for use by specialized fields and multiple disciplines, while the third category acquires scientific data and resources through scientific expeditions and comprehensive monitoring to support scientific research and national development.

Large-scale scientific facilities are the most important source of scientific data production. Both China and developed countries attach great importance to the development of such facilities, but there are significant differences in how their roles and functions are defined, with China placing greater emphasis on "goal-driven, problem-oriented" approaches. Under the unified deployment of relevant national authorities, China's layout of large-scale scientific facilities has gradually improved, with more efficient operations and richer outputs, providing tremendous support for the development of China's scientific and technological endeavors and making outstanding contributions to solving critical bottlenecks encountered in national development.

2.2.2 Scientific Data Centers **(1) International Scientific Data Centers.** International scientific data centers are infrastructure deployed to serve national and global development, address major scientific questions, promote technological innovation, and facilitate sustainable development. For example, the International Research Center of Big Data for Sustainable Development Goals (CBAS) is a typical international scientific data center [40]. Its big data platform system (SDGs Big Data Platform) integrates various types of data, including basic geography, remote sensing, ground monitoring, and social statistics, covering the entire workflow of "big data storage—management—computational analysis—visualization." The center has developed an SDG data product production system enabling interactive online analysis of terabyte-scale data, online calculation of various indicators, and visualization displays. It has also developed core functions such as specialized SDG repositories to support continuous aggregation and open sharing of global SDG data resources. The center has built a proprietary environment supporting Earth big data management, processing, and analysis, with supercomputing capabilities of 10,000 trillion double-precision floating-point operations per second, 50 PB of data storage capacity, and 10,000 CPU core cloud computing capabilities. Currently, the platform has aggregated 16 PB of data, providing "one-stop" data computing, analysis, display, and sharing services for three typical user scenarios: the public, researchers, and decision-makers [24]. The platform has passed CODATA evaluation, and its scientific data services have reached 174 countries and regions.

(2) National Scientific Data Centers. In 2019, to further improve the science and technology resource sharing service system and promote the opening and sharing of scientific resources with society, the Ministry of Science and Technology and the Ministry of Finance jointly designated 20 national scientific data centers (Table 1), covering fields such as high-energy physics, space science and astronomy, biological genomics, environment and ecology, geology and seismology, agriculture and forestry, and meteorology. These centers are responsible for the deposition and sharing of scientific data, scientific computing, and data technology research in their respective domains. By the end of 2021, the national scientific data centers had aggregated over 100 PB of data, with hundreds of PB accessed annually, providing more than 100 million CPU hours of scientific computing services and offering important support for scientific discovery, technological innovation, and the national economy.

(3) CAS Scientific Data Center System. To implement the *Measures for Scientific Data Management*, CAS issued the *CAS Measures for Scientific Data Management and Open Sharing* in February 2019 and launched the construction of an integrated scientific data center network centered on three types of scientific data centers—general center, disciplinary center, and institute-level center—supported and driven by security, operation, and evaluation systems [41,42]. The CAS Scientific Data Center System has been preliminarily established (Table 2), achieving positive results in supporting China's scientific and technological innovation and playing an active role in underpinning major national strategies and major engineering projects.

Challenges Facing the Development of China's Open Data Infrastructure

China's early open data infrastructure suffered from poor reliability and relative isolation. Over the past two decades, the state has focused on encouraging the creation of data portals to address common foundational issues of making scientific data "searchable, browsable, and shareable," achieving considerable progress in building open data public platforms and supporting facilities. However, overall, significant challenges remain in promoting data reuse and fostering scientific and social innovation [43]. Specifically, four major problems exist:

(1) The current construction scale cannot meet the growing needs of data management and usage. In terms of construction scale and financial investment, current emphasis remains on large-scale scientific facilities or national-level data infrastructure, which is insufficient to meet the needs of the entire scientific community and society for open scientific data management and usage.

(2) Existing standard systems and technical capabilities still cannot meet construction requirements [36]. Although standards, algorithms, and tools for big data management and processing continue to emerge, building open data infrastructure that meets diverse needs remains highly challenging, including: (i) lack of effective standard systems and query methods; (ii) lack of standardized modeling for system architecture, making it difficult to reuse data across different disciplines and industries and hindering system integration and interoperability; (iii) insufficient emphasis on standardized data management processes; and (iv) lack of effective guidelines for standard usage and upgrading, resulting in compatibility difficulties between legacy systems and new systems.

(3) The construction model is singular, with weak support for data integration, data analysis, and decision-making. Currently, domestic open data infrastructure relies on single funding sources with limited cross-departmental and cross-domain cooperation, leading to isolated systems or redundant construction. Open data infrastructure primarily supports integration of data within the same field and type, but lacks robust capabilities for integrating heterogeneous datasets from different domains, collaborative data analysis, and supporting scientific and technological decision-making.

(4) Lack of overall planning for long-term preservation and reuse of scientific

data. China's large-scale scientific facilities generate massive amounts of scientific data with high long-term utilization value. Particularly after such facilities or major scientific projects cease operation, effective data preservation and management, software and computing support, and comprehensive technical documentation are required to ensure long-term data availability. Except for a few fields that have begun developing plans for long-term preservation and reuse, China still lacks relevant overall planning.

Considerations and Recommendations for Strengthening China's Open Data Infrastructure Construction

Open scientific data requires robust, sustainable infrastructure and sound policy support. The implementation path for open data infrastructure should aim to fully exploit data value, ensuring that data can be "preserved, flowed, and utilized effectively." The following four recommendations are proposed for strengthening China's open data infrastructure construction:

(1) Strengthen top-level design and unified planning for open data infrastructure construction, establishing comprehensive national and international data centers.

Open data infrastructure has extensive scope and rich connotations, requiring strengthened national-level top-level design and implementation path planning to ensure coherence and operability of open data infrastructure policies. Recommendations: (i) Open data infrastructure construction should focus on open data platform development. Centered on data, open data platforms integrate storage, computing, network, and software resources to maximize open data value. On one hand, coordinate open data platform construction by researching and establishing the overall framework, service system, certification standards, and evaluation mechanisms. On the other hand, emphasize sustainable development of open data platforms by optimizing investment mechanisms through three approaches: formulating differentiated data policies, providing evaluation for data collection and usage, and offering support services. Encourage and guide different innovation entities to participate in data development, forming a funding guarantee system for data resource construction and service operation that combines dominant investment from the state, ministries, and national data center operating agencies with diversified investment sources to ensure sustainable development of open data platforms. (ii) Open data infrastructure construction helps break down data barriers. China's existing open data infrastructure is mainly concentrated in various disciplinary data centers or sharing platforms, which cannot adapt to the vigorous development of new technological revolutions represented by big data, the Internet of Things, and artificial intelligence, nor accelerate cross-domain scientific data applications and transformation into real productivity. Building comprehensive foundational national and international data centers is an inevitable solution. The International Research Center of Big Data for Sustainable Development Goals has conducted pioneering explorations with remarkable results, providing accumulated experience to promote leapfrog development of China's open data infrastructure. Recommendation: Encourage the initiation and construction of scientific and social problem-oriented scientific data infrastructure to mobilize and revitalize multidisciplinary and cross-domain scientific data resources, providing support for solving large-scale, complex scientific problems and social challenges.

(2) Adhere to collaborative and open construction of scientific data infrastructure clouds.

Open data infrastructure can effectively enhance research efficiency, participation, and visibility, strengthen research quality and rigor, and promote interdisciplinary collaboration among research teams. At the 2019 CODATA Beijing meeting, Chinese scientists proposed the initiative of collaboratively building a "Global Open Science Cloud" (GOSC), which has since reached broad consensus and established regular dialogue mechanisms with major global information infrastructure, international organizations, and platforms, and developed the first China-Europe cross-continental cloud federation testbed. Recommendation: In the future, driven by international big science programs and large-scale scientific facilities, China should give full play to its dominant role in data resources with clear advantages, focus on promoting the research and development of relevant data analysis methods and toolsets, strengthen communication, exchange, and training with international organizations and other countries, and actively participate in and promote international-level data sharing and cooperative applications. Simultaneously, China should leverage the leading and driving role of CAS open infrastructure to cooperatively establish an internationally shared open data cloud service system.

(3) Develop an innovation-driven development paradigm based on open science and reshape international science and technology cooperation governance models.

Open science facilitates breakthroughs in the classical limits of technology for the new scientific and technological revolution and industrial transformation, forming new rules, policies, evaluation standards, and indicator systems. By embracing diverse scientific and technological approaches, open science couples research advantages from different regions, fields, and teams, enabling all of humanity to conduct research while "standing on the shoulders of giants" and creating cumulative effects. Open scientific data is one of the fundamental conditions for realizing open science. Embracing the open science concept can rapidly enhance China's research capabilities. As the carrier of open scientific data, open data infrastructure can provide policymakers with more comprehensive integrated data and information, offer solutions for comprehensive policymaking on global challenges, and provide new pathways for cross-domain and cross-regional collaboration among researchers. Recommendation: Follow the development trend from open data to open science, utilize advanced technologies and methods such as cloud computing, big data, and blockchain, empower big data with artificial intelligence, create a new data sharing model that integrates data-computing-services, promote multidisciplinary data correlation analysis and information fusion, deepen comprehensive applications of multi-domain data, and drive major scientific discoveries and decision support.

(4) Foster an integrated data ecosystem and develop an innovation-driven development paradigm based on open science.

Open science and citizen science, which have developed under the guidance of open data acquisition, are closely related to sustainability science. Oriented toward development and cooperation, scientific and technological innovation should primarily manifest in applying scientific and technological progress to create new demands, applications, business forms, and markets while innovating models of scientific and technological cooperation. In recent years, newly planned scientific data infrastructure in Europe and the United States has shown a trend toward problem-oriented approaches that break disciplinary boundaries. Examples include the European brain research infrastructure integrating neuroscience and distributed computing technology, and the European computing/communication experimental large-scale research infrastructure supporting energy consumption and green transactions. These facilities are expected to fully mobilize multi-domain data resources, integrate various fields of natural science as well as natural and social sciences, become pioneers in promoting cross-integration and cross-domain interoperability, and foster a scientific data sharing ecosystem that is interdisciplinary, cross-scale, and cross-temporal.

References

1. Sheng X P, Bi C C, Tang J J. A review of open science research at home and abroad. *Documentation, Information & Knowledge*, 2022, 39(4): 101-113. (in Chinese)
2. Ramachandran R, Bugbee K, Murphy K. From open data to open science. *Earth and Space Science*, 2021, 8(5): e2020EA001562.
3. Sheng X P, Yang Z Y. Analysis of the relationship among open science, open sharing and open data. *Library and Information Service*, 2019, 63(17): 15-22. (in Chinese)
4. Fan Y J, Fan C L. Seeking common ground to promote exchanges and cooperation—Interpretation, enlightenment and suggestions of major science and technology policy reports of American think tanks and scientific communities. *Bulletin of Chinese Academy of Sciences*, 2022, 37(2): 197-205. (in Chinese)
5. United Nations Educational, Scientific, and Cultural Organization. *UNESCO recommendation on open science.* (2021-11-23)[2022-12-20].

https://unesdoc.unesco.org/ark:/48223/pf0000379949.

6. Wang F, Ping A. Analysis of the foreign information resource development in the national library of China in the context of open access. *Journal of the National Library of China*, 2022, 31(4): 93-101. (in Chinese)
7. Guo H D, Wang L Z, Chen F, et al. Scientific big data and digital Earth. *Chinese Science Bulletin*, 2014, 59(12): 1047-1054. (in Chinese)
8. Borowitz M. *The Global Effort for Open Access to Environmental Satellite Data*. Cambridge: The MIT Press, 2017.
9. Liu F H, Zhang T. An emerging type of scholarly articles in the Open Science background: Research elements publishing. *Chinese Journal of Scientific and Technical Periodicals*, 2017, 28(2): 138-144. (in Chinese)
10. Scientific Data Sharing Research Group. Overall framework of scientific data sharing project. *China Basic Science*, 2003, 5(1): 63-68. (in Chinese)
11. Chen X L, Quan Z W, Mao G, et al. Review of scientific data research from the perspective of open sharing. *Journal of Library and Information Science*, 2022, 7(5): 72-79. (in Chinese)
12. Zhao R X, Zhao H, Zheng J H, et al. Scientific data management in scientific research institutions: Practice and prospects. *Journal of Agricultural Big Data*, 2019, 1(4): 65-75. (in Chinese)
13. Aruhan, Wu C, Li X X. International practice of open research data and its enlightenment to China. *Bulletin of Chinese Academy of Sciences*, 2020, 35(1): 11-18. (in Chinese)
14. Committee on Data for Science and Technology. *The Beijing Declaration on Research Data*. Paris: CODATA, 2019.
15. Wilkinson M D, Dumontier M, Aalbersberg I J, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 2016, 3: 160018.
16. Xing W M, Guo A Q, Qin S, et al. FAIR principle of scientific data management and sharing: Background, content and implementation. *Journal of Information Resources Management*, 2021, 11(2): 60-68. (in Chinese)
17. Carroll S R, Herczog E, Hudson M, et al. Operationalizing the CARE and FAIR Principles for Indigenous data futures. *Scientific Data*, 2021, 8: 108.
18. Sun J L, Huang D C, Li X B. The new progress of science data managing and sharing service in China. *World Sci-Tech R & D*, 2002, 24(5): 15-19. (in Chinese)
19. Zhang L L, Wen L M, Shi L, et al. Progress in scientific data management and sharing. *Bulletin of Chinese Academy of Sciences*, 2018, 33(8): 774-782. (in Chinese)

20. General Office of the State Council of the People's Republic of China. *Scientific Data Management Measures*. (2018-03-17)[2023-05-25]. https://most.gov.cn/xxgk/xinxifenlei/fdzdgknr/fgzc/gfxwj/gfxwj2018/201804/t20180404_{139023}.htm (in Chinese)

21. Guo H D. Big data, big science, big discovery—Review of CODATA workshop on big data for international scientific programmes. *Bulletin of Chinese Academy of Sciences*, 2014, 29(4): 500-506. (in Chinese)

22. Wang J L, Sun J L. Review, Reform and prospect analysis of world data center. *Advances in Earth Science*, 2009, 24(6): 612-620. (in Chinese)

23. Pan X D, Li X, Ran Y H, et al. Open access of scientific data in the context of open science: The practice of the National Tibetan Plateau Data Center. *Big Data Research*, 2022, 8(1): 113-120. (in Chinese)

24. Guo H D. A project on big earth data science engineering. *Bulletin of Chinese Academy of Sciences*, 2018, 33(8): 818-824. (in Chinese)

25. Guo H, Goodchild M F, Annoni A. *Manual of Digital Earth*. Singapore: Springer, 2020.

26. Guo H D, Liang D, Chen F, et al. Big earth data facilitates sustainable development goals. *Bulletin of Chinese Academy of Sciences*, 2021, 36(8): 874-884. (in Chinese)

27. Guo H D, Liang D, Sun Z C, et al. Measuring and evaluating SDG indicators with Big Earth Data. *Science Bulletin*, 2022, 67(17): 1792-1801.

28. Hou J G. Preface of special issue on big earth data for implementing the sustainable development goals. *Bulletin of Chinese Academy of Sciences*, 2021, 36(8): 873. (in Chinese)

29. Guan L L, Ma H Q, Wang C L. Practice and thoughts on establishing international data journals: A case study of Big Earth Data. *Chinese Journal of Scientific and Technical Periodicals*, 2020, 31(1): 56-62. (in Chinese)

30. Guo H D. Big data drives the development of Earth science. *Big Earth Data*, 2017, 1(1-2): 1-3.

31. Wang J L, Wang Y J, Zhang M, et al. Review of the progress of earth data science and sharing in 2020. *Science & Technology Review*, 2021, 39(1): 105-114. (in Chinese)

32. Guo H D, Chen R S, Xu Z W, et al. Big data for natural sciences, humanities and social sciences—Review of the 6th exploratory round table conference. *Bulletin of Chinese Academy of Sciences*, 2016, 31(6): 707-716. (in Chinese)

33. Wang Y, Zhou Y C, Wang Y G, et al. Promoting infrastructure construction in advance to support sci-tech self-reliance and self-strengthening at

higher level. *Bulletin of Chinese Academy of Sciences*, 2022, 37(5): 652-660. (in Chinese)

34. Teng J W, Liu Y S, Pi J L. Development of science and technology, times echo of big data—Acquisition, innovation and reconstruction of new information of earth sciences. *Progress in Geophysics*, 2016, 31(1): 1-22. (in Chinese)
35. Zhao Z M, Liao X F, Wang X L, et al. A Reference Model approach for developing agricultural big data infrastructures. *Journal of East China Normal University (Natural Science)*, 2019, (2): 77-96. (in Chinese)
36. Xi G Q, Fu H, Liu G Y. Current development of large scale scientific facilities in China and foreign experiences. *Science & Technology Review*, 2020, 38(11): 6-15. (in Chinese)
37. Wang Y F. The Current Situation and Future Development of Major Science and Technology Infrastructure in China. (2022-06-24)[2022-12-02]. http://www.npc.gov.cn/npc/c30834/202206/90acffdbc88045758409c17e7dd86971.shtml. (in Chinese)
38. Luo X A, Yang C X. Retrospect and consideration on construction of large research infrastructures of CAS. *Bulletin of Chinese Academy of Sciences*, 2012, 27(6): 710-716. (in Chinese)
39. Guo H D, Liang D, Chen F, et al. Innovative approaches to the Sustainable Development Goals using Big Earth Data. *Big Earth Data*, 2021, 5(3): 263-276.
40. Qiu C Y. Metadata research progress in scientific data-related practices at home and abroad. *Information and Documentation Services*, 2021, 42(5): 104-112. (in Chinese)
41. Li Y, Wen L M. Research on the implementation status, influencing factors and promotion strategies of the "Scientific Data Management Rule". *Library and Information Service*, 2021, 65(2): 65-74. (in Chinese)
42. Li X, Cheng G D, Wang L X, et al. Boosting geoscience data sharing in China. *Nature Geoscience*, 2021, 14(8): 541-542.

Author Biography

GUO Huadong is Director General of the International Research Center of Big Data for Sustainable Development Goals. He is an Academician of the Chinese Academy of Sciences, a Foreign Member of the Russian Academy of Sciences, a Foreign Member of the Finnish Society of Sciences and Letters, a Fellow of TWAS, and a Fellow of ISC. He specializes in remote sensing, radar for Earth observation, and Digital Earth science. He presently serves as Honorary President of the International Society for Digital Earth, Director of the International

Center on Space Technologies for Natural and Cultural Heritage under the Auspices of UNESCO, and Editor-in-Chief of two scientific journals *International Journal of Digital Earth* and *Big Earth Data*. E-mail: hdguo@radi.ac.cn

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